Новый алгоритм встраивания информации на основе переквантования, стойкий к статистической атаке восстановления ключа

В.А. Митекин[а,б], В.А. Федосеев[а,б]

[а] *Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Московское шоссе, 34, Самара, Россия*
[б] *Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, 443001, ул. Молодогвардейская, 151, Самара, Россия*

Аннотация

В предыдущей работе одним из авторов была представлена новая атака против известного алгоритма встраивания цифровых водяных знаков DM-QIM, позволяющая при наличии некоторого числа изображений со встроенной информацией восстановить и встроенный ЦВЗ, и ключ встраивания. В данной работе на основе анализа причин уязвимости алгоритма DM-QIM к данной статистической атаке предложена стойкая модификация данного алгоритма, получившая название IM-QIM. Данный алгоритм основан на использовании т.н. «корреляционно-стойкой» функции встраивания и позволяет обеспечить статистическую независимость модифицируемых компонент контейнера и отдельных битов встраиваемой последовательности. В рамках работы проведены эксперименты по исследованию стойкости алгоритма к аддитивному зашумлению и влияния параметров встраивания на качество результирующего изображения, которые показали превосходство IM-QIM над существующими алгоритмами QIM и DM-QIM на широком интервале значений параметров.

Ключевые слова: QIM; DM-QIM; IM-QIM; quantization index modulation; dither modulation; цифровой водяной знак

1. Введение

В сетях распространения мультимедийной информации одной из актуальных задач является защита авторских прав мультимедиа-продукции. За последние 25 лет было предложено огромное количество алгоритмов и программных продуктов, позволяющих решать данную задачу посредством встраивания в объект защиты цифровых водяных знаков (ЦВЗ) – малозаметной шумоподобной информации, содержащей закодированную информацию об авторе или владельце [1]. При оценке практической применимости подобных алгоритмов ключевой характеристикой является количественное соотношение между объемом шумовых искажений, вносимых при встраивании ЦВЗ, и стойкостью встроенного ЦВЗ к преднамеренным атакам и непреднамеренным искажениям мультимедиа. Таким образом, используемые алгоритмы и методы кодирования и встраивания водяного знака являются основными факторами, определяющими данное количественное соотношение в заданных условиях использования ЦВЗ (наличие и способы преднамеренных атак, возможность искажения и сжатия с потерями защищенного изображения и т.д.).

Одним из наиболее распространённых алгоритмов скрытого встраивания информации (в том числе ЦВЗ) в цифровые изображения является алгоритм Quantization Index Modulation (QIM). Его важным практическим преимуществом является обеспечиваемое им высокое соотношение «степень искажения – стойкость ЦВЗ» в условиях, когда основным типом искажений защищенного изображения является аддитивный белый гауссовский шум [2]. На основе исходного алгоритма (QIM) был разработан ряд модификаций [2-10], основанных на общем принципе – для встраивания различных битов информации используются различные шкалы переквантования. При этом, в зависимости от модификации, переквантованию могут подвергаться как непосредственно значения яркости отдельных пикселей, так и значения коэффициентов дискретного преобразования Фурье или дискретного косинусного преобразования исходного изображения. Учитывая тот факт, что базовый алгоритм QIM не обеспечивает помехоустойчивого кодирования встраиваемой информации, извлекаемый ЦВЗ оказывается уязвим к искажениям изображения-носителя.

Наиболее распространенная модификация алгоритма QIM, DM-QIM [2], основан на использовании дополнительного параметра - «массива подмешиваемых значений», который используется в качестве секретного ключа, необходимого и для встраивания, и для извлечения ЦВЗ. «Массив подмешиваемых значений» также позволяет скрыть специфичные искажения гистограммы изображения-носителя, возникающие при встраивании ЦВЗ методом QIM, и тем самым позволяет защитить ЦВЗ от преднамеренного обнаружения атакующим, не знающим ключа встраивания.

Несмотря на широкое использование алгоритма QIM и его модификаций, известны примеры атак, позволяющих обнаружить и/или извлечь встроенную информацию без знания секретного ключа. Так, в работах [11-13] были предложены методы, позволяющие по одному изображению с ЦВЗ с вероятностью не ниже 0.9 обнаружить и извлечь встроенный ЦВЗ без знания ключа. Кроме того, в работе [14] предложен другой подход, который на основе использования множества изображений, содержащих информацию, встроенную с использованием одного ключа, позволяет не только обнаружить встроенную информацию, но и восстановить секретный ключ системы. В настоящей работе предлагается новый алгоритм семейства QIM, названный IM-QIM (immune QIM), позволяющий защититься от данной атаки.

Работа организована следующим образом. Раздел 2 содержит краткое описание алгоритмов QIM и DM-QIM, а также метода атаки на эти алгоритмы, предложенной в работе [14]. В разделе 3 описывается предлагаемая модификация

алгоритма DM-QIM и обосновывается её потенциальная стойкость к атаке [14], а раздел 4 посвящён экспериментальным исследованиям предложенного алгоритма. Работу завершают заключение и благодарности.

2. Базовые методы и атака на них

2.1. Алгоритм QIM

Как отмечалось выше, алгоритм QIM применим для встраивания информации как в отдельные пиксели изображения, так и в спектральные отсчёты или в коэффициенты ДКП, квантуемые при сжатии в формате JPEG. Кроме того, он применим для защиты не только изображений, но и видео, а также звуковых сигналов. Однако для определённости рассмотрим данный алгоритм на примере модификации отдельных пикселей полутонового изображения.

Итак, пусть $I(n,m)$ – исходное полутоновое изображение-контейнер, где $n \in [1,N], m \in [1,M]$. Пусть $W(k)$ – двоичная последовательность, выступающая в качестве встраиваемого ЦВЗ, где $k \in [1, N \cdot M]$. Основными параметрами алгоритма являются:

- шаг переквантования $\Delta \in \mathbb{N}$, который определяет одновременно устойчивость встроенного ЦВЗ к аддитивному белому шуму и среднюю амплитуду так называемого “шума квантования” (искажений, вносимых при встраивании ЦВЗ);
- шкала переквантования, используемая для встраивания информации и задаваемая в виде функции $Q(x,\Delta)$, где x – квантуемое значение яркости. Простейшая шкала переквантования, которая повсеместно будет использоваться в данной работе, имеет вид:

$$Q(x,\Delta) = \Delta \cdot round\left(\frac{x}{\Delta}\right), \tag{1}$$

где $round(\cdot)$ - операция округления до ближайшего целого.

Формирование изображения $I^W(n,m)$ со встроенной информацией осуществляется по формуле

$$I^W(n,m) = \mathrm{E}_{QIM}\left(I,W,\Delta\right) = Q\left(I(n,m),\Delta\right) + \frac{\Delta}{2}W\left(k\right), \tag{2}$$

где k и (n,m) связаны между собой каким-либо биективным отображением, например: $k = nM + m$.

Таким образом, в результате встраивания информации $I^W(n,m)$ содержит значения, кратные $\Delta/2$, как показано на рисунке 1.

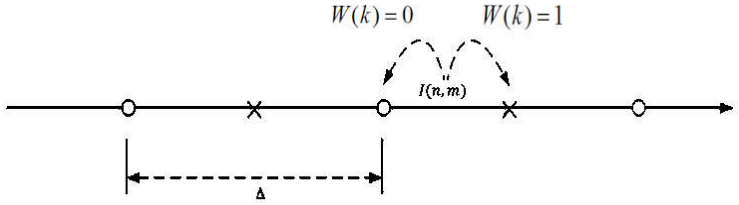


Рис. 1. Иллюстрация встраивания информации методом QIM.

Для извлечения информации воспользуемся не самым удобным, но наиболее универсальным способом, который можно применить и для модификаций QIM. Пусть $\widetilde{I^W}(n,m)$ – изображение, поступившее на принимающую сторону. В общем случае оно не совпадает с $I^W(n,m)$, поскольку в процессе передачи могло быть подвергнуто каким-либо искажениям или атакам. Тогда извлечение информации происходит по формулам:

$$\widetilde{I}_0(n,m) = \mathrm{E}_{QIM}\left(\widetilde{I^W},0,\Delta\right) = Q\left(\widetilde{I^W}(n,m),\Delta\right). \tag{3}$$

$$\widetilde{I}_1(n,m) = \mathrm{E}_{QIM}\left(\widetilde{I^W},W,\Delta\right) = Q\left(\widetilde{I^W}(n,m),\Delta\right) + \frac{\Delta}{2}. \tag{4}$$

$$\widetilde{W}\left(k\right) = \arg\min_{p\in\{0,1\}}\left|\widetilde{I^W}(n,m) - \widetilde{I}_p(n,m)\right|. \tag{5}$$

Иными словам, при извлечении ЦВЗ осуществляется подстановка битов 0 и 1 в формулу (2), причём в качестве контейнера используется $\widetilde{I^W}(n,m)$, а далее оцениваются отклонения полученных результатов.

Ввиду существенного сужения множества возможных значений пикселей, результат применения алгоритма QIM легко обнаруживается по гистограмме изображения [15], поэтому на практике чаще применяются его многочисленные модификации [2-10], среди которых одной из наиболее известных является DM-QIM.

2.2. Алгоритм DM-QIM

Алгоритм DM-QIM (Dither Modulation – QIM) предполагает использование двух дополнительных параметров – «массивов подмешиваемых значений», согласованных друг с другом и используемых при встраивании битов “0” и “1”:

$$d_0(k),\ d_1(k) \in [-\Delta/2;\Delta/2],\ k \in [1, N \cdot M]\ .$$

Пусть $d(k)$ – массив псевдослучайных целых чисел, равномерно распределённых на отрезке $[-\Delta/2;\Delta/2]$, который генерируется на основе секретного ключа. Тогда

$$\begin{aligned} &d_0(k) = d(k)\,, \\ &d_1(k) = d_0(k) - sign(d_0(k)) \cdot \Delta/2\ . \end{aligned} \tag{6}$$

Формула встраивания информации принимает вид

$$I^W(n,m) = \mathrm{E}_{DM-QIM}\left(I, W, d, \Delta\right) = Q(I(n,m) + d_{W(k)}(k), \Delta) - d_{W(k)}(k)\,, \tag{7}$$

то есть к значению яркости очередного пикселя перед переквантованием подмешивается соответствующее значение одного из массивов $d_0(k)$ или $d_1(k)$, соответствующее встраиваемому биту и его позиции в векторе. Вычитание шумоподобной добавки из переквантованных значений позволяет затруднить обнаружение встраивания DM-QIM по гистограмме результирующего изображения.

Извлечение информации происходит по формуле (5), где $\widetilde{I}_p(n,m),\ p = \{0,1\}$ формируются согласно изменённой формуле встраивания:

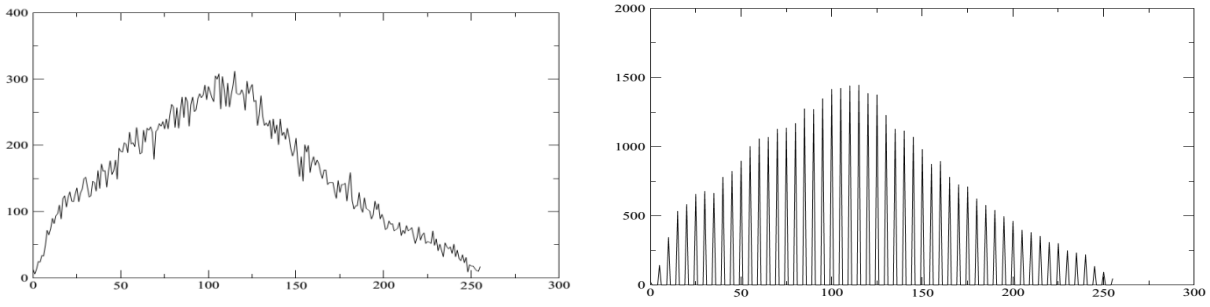
$$\widetilde{I}_0(n,m) = \mathrm{E}_{DM-QIM}\left(\widetilde{I^W}, 0, d, \Delta\right) = Q\left(\widetilde{I^W}(n,m) + d_0(k), \Delta\right) - d_0(k)\,. \tag{8}$$

$$\widetilde{I}_1(n,m) = \mathrm{E}_{DM-QIM}\left(\widetilde{I^W}, 1, d, \Delta\right) = Q\left(\widetilde{I^W}(n,m) + d_1(k), \Delta\right) - d_1(k)\,. \tag{9}$$

2.3. Описание атаки на DM-QIM

Принцип атаки на метод DM-QIM, предложенной в работе [14] и применимой также для других модификаций метода QIM [3-7], основан на том, что смещение яркости пикселей результирующего изображения относительно переквантованных значений определяется только значением $d_0(k)$ или зависящим от него $d_1(k)$. Таким образом, если атака позволяет восстановить шаг квантования, а также смещения яркости пикселей $d_0(k)$ и $d_1(k)$, то это позволит не только обнаружить факт встраивания, но и с точностью до одного бита (согласно формуле (6)) восстановить ключевую последовательность $d_0(k)$.

Предложенный в работе [14] метод атаки на алгоритм DM-QIM использует набор из T изображений $I_t^W(n,m)$, $t \in [1,T]$, в которые произведено встраивание информации при помощи одного ключа (встраиваемые последовательности при этом могут различаться). Далее, если взять какой-либо пиксель (n,m) и построить гистограмму его значений на всём наборе из T изображений, то подобная гистограмма будет иметь хорошо различимые пики, повторяющиеся с периодом $\Delta/2$. Для иллюстрации на рисунке 2 показан пример такой гистограммы и для сравнения пример аналогичной гистограммы, полученной по исходным контейнерам.



а) б)

Рис. 2. Гистограмма значений одного пикселя на множестве изображений: а) исходные контейнеры; б) результаты встраивания алгоритмом DM-QIM.

На основе анализа гистограммы восстанавливается как значение Δ, так и смещение в данном пикселе, по которому при наличии хотя бы одного изображения $I_t^W(n,m)$ с известной встроенной последовательностью $W_t(k)$ однозначно восстанавливаются $d_0(k)$ и $d_1(k)$, что позволяет полностью взломать систему и извлечь информацию из всех изображений-носителей, использующих тот же секретный ключ.

3. Предлагаемый метод IM-QIM

3.1. Анализ причин уязвимости DM-QIM и идея предлагаемой модификации

Источником уязвимости DM-QIM является тот факт, что в результате встраивания информации по формуле (7) неизбежно возникает корреляционная зависимость между яркостью отдельного пикселя изображения $I^W(n,m)$ (точнее, остатком от деления $I^W(n,m)$ $I'(n,m)$ на Δ) и отдельным битом ключа $d_{W(k)}(k)$. Атакующий в подобной ситуации может без какой либо информации о значении ключа сформировать выборку пикселей таким образом, чтобы по их статистическим характеристикам оценить неизвестный ему фрагмент ключа.

Можно предположить существование как минимум двух способов защиты от подобного типа атаки.

Во-первых, атаки можно избежать, используя различные значения ключа для каждого нового изображения-контейнера. Данный подход рассмотрен в ряде работ, где предложены либо способы генерации ключа на основе статистических характеристик изображения-контейнера, либо способы встраивания информации об используемом ключе в само изображение-контейнер. Как показано в [16, 17], в случае использования данного подхода возникает опасность применения атакующим более узкоспециализированных атак, основанных на т.н. «десинхронизации» ключа, т.е. невозможности восстановления исходного ключа из изображения-контейнера.

Во-вторых, при встраивании ЦВЗ возможно использовать более сложные алгоритмы встраивания, для которых корреляционная зависимость между отдельным битом ключа и отдельным пикселем (отсчетом) изображения-контейнера не возникает в результате встраивания ЦВЗ даже с использованием одного и того же ключа.

Так, рассмотрим широко используемые при разработке систем поточного шифрования так называемые корреляционно-стойкие функции [18].

Функция $F(X_0, X_1...X_q)$ является корреляционно-стойкой функцией $j-го$ порядка тогда и только тогда, когда для любых $u_1, u_2..u_j$, выбранных из диапазона $[1,q]$, случайная переменная $Z = F(X_0, X_1...X_q)$ и случайный вектор $(X_{u_1}, X_{u_2}...X_{u_j})$ являются статистически независимыми.

Наиболее известным на практике примером подобной функции является функция сложения по модулю 2 (функция XOR) $F_{XOR}(X_1, X_2, X_3...X_q) = X_1 \oplus X_2 \oplus X_3 \oplus ... \oplus X_q$, где $X_1...X_q \in \{0,1\}$. Действительно, несложно показать, что функция $F_{XOR}(X_1...X_q)$ является корреляционно-стойкой функцией *q-1* порядка, т.к.

$P\left(F_{XOR}(X_1, X_2, X_3...X_q) = z \mid X_1 = g_1, X_2 = g_2, ...X_{q-1} = g_{q-1}\right) = P\left(F_{XOR}(X_1, X_2, X_3...X_q) = z\right)$ для любых $z, g_1, g_2...g_{q-1}$

Фактически, данная функция не позволяет атакующему использовать априорную информацию об известных ему $(X_1, X_2, X_3...X_{q-1})$ для построения более точного, по сравнению с «подбрасыванием монеты», прогноза о значении $z = F_{XOR}(X_1, X_2, X_3...X_q)$.

Рассмотрим теперь следующую функцию (функция сложения по модулю Δ)

$$F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q)) = \Theta\left(I(n_1,m_1) + I(n_2,m_2) + ... + I(n_q,m_q), \Delta\right),$$

где $\left\{I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q)\right\}$ – некая секретным образом выбранная группа пикселей изображения I.

Покажем, что данная функция является корреляционно стойкой функцией *q-1* порядка при условии, что значения $\Theta(I(n_q,m_q),\Delta)$ (остатки от деления $I(n_q,m_q)$ на Δ) являются независимыми равномерно распределенными в диапазоне $[-\Delta/2, \Delta/2)$ случайными величинами. Действительно, если атакующему известно значение $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_{q-1},m_{q-1}))$ и отдельные значения из набора $\left\{I(n_1,m_1), I(n_2,m_2)...I(n_{q-1},m_{q-1})\right\}$, то величина $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q))$, с точки зрения оценки атакующим, будет также являться равномерно распределенной случайной величиной. Так, рассматривая возможные значения $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q))$, можно увидеть, что для любого (известного атакующему) значения $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_{q-1},m_{q-1})) \in [0, \Delta-1]$ будет существовать ровно одно значение $I(n_q,m_q)$, при котором результирующая функция $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q))$ будет равна нулю; аналогично, будет существовать ровно одно значение $I(n_q,m_q)$, при котором результирующая функция $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q))$ будет равна единице и т.д. вплоть до $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q)) = \Delta - 1$. Следовательно, для атакующего величины $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q))$ и $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_{q-1},m_{q-1}))$ являются статистически независимыми. Аналогично можно показать, что статистически независимыми являются $F_{IM}(I(n_1,m_1), I(n_2,m_2)...I(n_q,m_q))$ и любой вектор $(I(n_{u_1},m_{u_1}), I(n_{u_2},m_{u_2})...I(n_{u_r},m_{u_r}))$, где $r = q-1$, а $u_1, u_2...u_r \in [1,q]$.

Далее, возвращаясь к соотношениям (2)-(5) можно увидеть, что рассмотренное значение извлеченного бита $\widetilde{W}(k) = \arg\min_{p\in\{0,1\}} \left| \widetilde{I^W}(n,m) - \widetilde{I}_p(n,m) \right|$ коррелировано именно с величиной $\Theta(\widetilde{I^W}(n,m),\Delta)$, т.е. с остатком от деления $\widetilde{I^W}(n,m)$ на Δ, а не с самим значением яркости $\widetilde{I^W}(n,m)$. Действительно, если в выражениях (3)-(4) заменить конкретное значение $\widetilde{I^W}(n,m)$ на $\left(\widetilde{I^W}(n,m)+\Delta\right)$, то результат вычисления (5) извлеченного бита при этом не изменится при любых допустимых значениях $\widetilde{I^W}(n,m)$. Именно это корреляционная связь и используется для проведения атаки, рассмотренной в работе [14].

Идея предлагаемого алгоритма заключается в том, чтобы заменить исходную функцию переквантования яркости отдельного пикселя $Q\left(I(n,m),\Delta\right)$ в выражениях (3) и (4) на функцию переквантования суммы яркостей сразу нескольких пикселей. Тогда, согласно (2)- (5), значение встроенного бита ЦВЗ будет коррелированно уже не с яркостью отдельных пикселей изображения-контейнера, а с функцией $\Theta\left(I(n_1,m_1)+I(n_2,m_2)+...+I(n_q,m_q),\Delta\right)$ от значений яркости нескольких пикселей. В то же время корреляционной зависимости между битом ЦВЗ и яркостью отдельных пикселей группы $\left(I(n_1,m_1),I(n_2,m_2)...I(n_q,m_q)\right)$ наблюдаться не будет (т.к. указанная функция $\Theta\left(I(n_1,m_1)+I(n_2,m_2)+...+I(n_q,m_q),\Delta\right)$ является корреляционно устойчивой, как было показано ранее). Таким образом, описанная ранее атака становится реализуемой лишь в случае, когда атакующему известно точное разбиение пикселей изображения на группы для последующего переквантования.

Также отдельно отметим рассмотренное ранее требование к значениям $\Theta(I(n_q,m_q),\Delta)$, необходимое для возникновения корреляционной устойчивости функции F_{IM}. Для выполнения этого требования в предложенном алгоритме должно применяться аддитивное зашумление всех пикселей изображения равномерно распределенным белым шумом.

3.2. Формальное описание предлагаемого алгоритма IM-QIM

Принципиальным отличием алгоритма IM-QIM (statistically IMmune QIM) от исходного DM-QIM является способ использования секретного ключа – в предлагаемом алгоритме ключ используется для разбиения множества пикселей изображения на группы, к которым в дальнейшем применяется операция переквантования. Алгоритм IM-QIM использует также массивы $d_0^g(k)$, $d_1^g(k)$, $g = 2..G$ (об индексации речь пойдёт ниже) – аналоги массивов $d_0(k)$ и $d_1(k)$ в DM-QIM, но в алгоритме IM-QIM данные массивы не являются частью ключа и их знание не требуется при извлечении ЦВЗ. Это означает, что массивы $d_0^g(k)$, $d_1^g(k)$ могут генерироваться уникальным образом для каждого нового изображения и проблемы синхронизации ключа, в отличие от алгоритма DM-QIM, в данном случае не возникает.

Обобщённая формула встраивания IM-QIM может быть записана в виде

$$I^W(n,m) = Q(I(n,m)+\Xi\left(I,W,\kappa\right),\Delta) - \Xi\left(I,W,\kappa\right), \tag{10}$$

где $\Xi\left(\cdot\right)$ - некоторая функция, зависящая от изображения I, встраиваемого бита W и некоторого набора величин, определяемых секретным ключом κ (это может быть и одна величина - массив d), причём

$$Q(I^W(n,m)+\Xi\left(I^W,W,\kappa\right),\Delta) - \Xi\left(I^W,W,\kappa\right) = Q(I(n,m)+\Xi\left(I,W,\kappa\right),\Delta) - \Xi\left(I,W,\kappa\right). \tag{11}$$

Пусть $D = \{(n,m)\}_{n=1..N,\ m=1..M}$ – область определения изображения. Разделим это множество на G равных непересекающихся частей (групп пикселей) $D_1, D_2, ..., D_G$, объединение которых равно всему множеству D. Это разбиение может быть задано отображением вида

$$F : \mathbb{N}^2 \to \mathbb{N}, \quad F\left(n,m\right) = g,$$

которое зависит от секретного ключа. Далее пронумеруем пиксели, входящие в каждое из этих множеств: $\left(n_1^g, m_1^g\right), \left(n_2^g, m_2^g\right), ..., \left(n_{NM/G}^g, m_{NM/G}^g\right)$, где $g = 1..G$ – индекс, определяющий подмножество.

Тогда для встраивания можно использовать формулу (12), аналогичную (10), с той лишь разницей, что она используется только для пикселей множества D_1:

$$I^W\left(n_k^1, m_k^1\right) = \mathrm{E}_{IM-QIM}\left(I,W,d,F,\Delta\right) = Q\left(I\left(n_k^1, m_k^1\right) + \Xi\left(I,W,d,F\right),\Delta\right) - \Xi\left(I,W,d,F\right), \tag{12}$$

с функцией $\Xi(I,W,d,F)$ вида

$$\Xi(I,W,d,F)=\sum_{g=2}^{G}\left(I\left(n_k^g,m_k^g\right)+d_{W(k)}^g(k)\right), \tag{13}$$

где $d_0^g(k),\ d_1^g(k)$ – аналоги массивов $d_0(k)$ и $d_1(k)$, генерирующиеся по тому же принципу отдельно для каждого подмножества D_g.

Пиксели изображения, не относящиеся к D_1, не будут передавать скрытую информацию, а будут лишь искажаться в соответствии с процедурой переквантования и значениями массивов $d_0^g(k),\ d_1^g(k)$:

$$I^W\left(n_k^g,m_k^g\right)=Q\left(I\left(n_k^g,m_k^g\right)+d_{W(k)}^g(k),\Delta\right)-d_{W(k)}^g(k),\ \ g=2..G. \tag{14}$$

Для извлечения информации не требуются массивы $d_0^g(k),\ d_1^g(k)$. Вместо этого по принятому носителю встроенной информации рассчитываются оценки

$$\widetilde{d}_0(k)=\sum_{g=2}^{G}\widetilde{I^W}\left(n_k^g,m_k^g\right). \tag{15}$$

$$\widetilde{d}_1(k)=\widetilde{d}_0(k)-sign(d_0(k))\cdot\Delta/2\ , \tag{16}$$

после чего извлечение происходит по традиционной схеме, аналогичной (3)-(5) и (8)-(9):

$$\widetilde{I}_0\left(n_k^1,m_k^1\right)=\mathrm{E}_{IM-QIM}\left(\widetilde{I^W},0,d,\Delta\right)=Q\left(\widetilde{I^W}\left(n_k^1,m_k^1\right)+\widetilde{d}_0(k),\Delta\right)-\widetilde{d}_0(k). \tag{17}$$

$$\widetilde{I}_1\left(n_k^1,m_k^1\right)=\mathrm{E}_{IM-QIM}\left(\widetilde{I^W},1,d,\Delta\right)=Q\left(\widetilde{I^W}\left(n_k^1,m_k^1\right)+\widetilde{d}_1(k),\Delta\right)-\widetilde{d}_1(k). \tag{18}$$

$$\widetilde{W}\left(k\right)=\arg\min_{p\in\{0,1\}}\left|\widetilde{I^W}\left(n_k^1,m_k^1\right)-\widetilde{I}_p\left(n_k^1,m_k^1\right)\right|. \tag{19}$$

4. Экспериментальные исследования предложенного метода

4.1. Исследование возможности обнаружения встраивания на основе анализа гистограмм

Для исследования применимости атаки [14] для разработанного алгоритма IM-QIM был использованы изображения из набора [19]. В каждое изображение встраивался ЦВЗ алгоритмами QIM, DM-QIM, IM-QIM. Значение параметра встраивания Δ при этом было одинаково для всех изображения и равнялось 10. Далее для указанного набора была воспроизведена исходная атака [14], основанная на построении гистограммы для значений яркости всех пикселей с заданными координатами. На рисунке 3 показаны фрагменты подобной гистограммы для выборки из 200000 пикселей, построенной до и после встраивания (при этом для встраивания информации во все пиксели использовался один и тот же ключ). На первой-второй гистограммах, соответствующих алгоритмам встраивания QIM и DM-QIM, чётко просматриваются пики, следующие с частотой $\Delta/2$ и Δ соответственно. Как уже было отмечено ранее, это связано с наличием корреляционной связи между битом ЦВЗ (0 или 1) и величиной $\Theta(\widetilde{I^W}(n,m),\Delta)$. Иными словами, сам факт встраивания ЦВЗ алгоритмами DM-QIM или QIM формирует множество «предпочитаемых» значений яркости таким образом, что для всех значений в нем величина $\Theta(\widetilde{I^W}(n,m),\Delta)$ будет равна заданной константе (причем эта величина однозначно определяется величинами $\Delta, d_0(k)$ и $d_1(k)$, как было показано в [14]). В то же время на третьей гистограмме такой картины не наблюдается, т.е. корреляционная связь между яркостью отдельного пикселя и битом ЦВЗ не прослеживается. Таким образом, можно говорить о стойкости нового алгоритма к данной атаке.

4.2. Исследование искажений, вносимых при встраивании информации

В ходе исследований также анализировалось, как влияет разработанный метод встраивания на качество результирующих изображений при различных значениях параметра Δ в сравнении с двумя другими рассмотренными в работе методами. Для IM-QIM использовались значения G, равные 2, 4, 8 и 16. В роли показателя качества использовалось среднеквадратичное отклонение (MSE) носителя информации от контейнера. Следует заметить, что значение максимальной ошибки согласно формулам встраивания для всех алгоритмов составляет $\Delta/2$. Эксперимент проводился на множестве из 100 изображений тестового набора [19]. Результаты отражает диаграмма на рисунке 4.

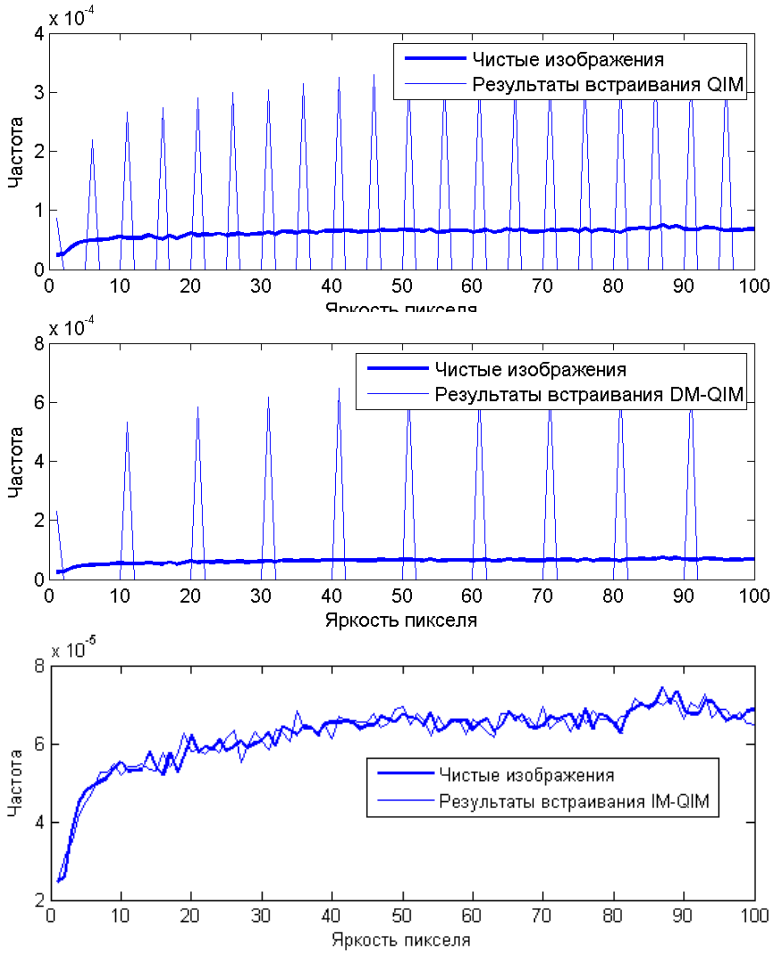


Рис. 3. Результаты атаки с анализом гистограмм на алгоритмы QIM, DM-QIM, IM-QIM (сверху вниз).

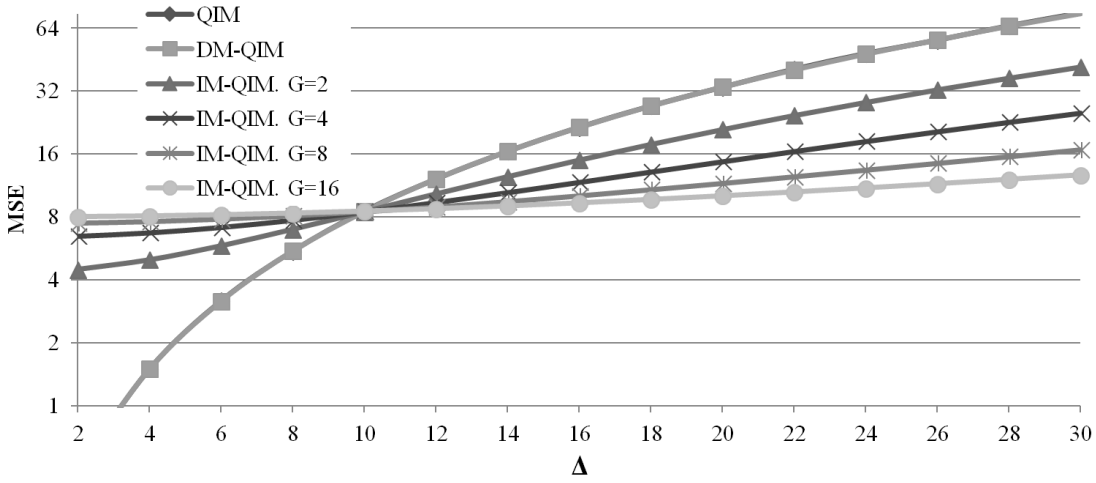


Рис. 4. Зависимость MSE при встраивании информации от параметра Δ для разных алгоритмов.

Из графиков видно, что при малых параметрах Δ IM-QIM приводит к большим искажениям, нежели базовые методы, однако начиная с $\Delta = 12$ он выигрывает их по качеству, причём выигрыш тем больше, чем больше значение G.

4.3. Исследование стойкости алгоритма к зашумлению

Очевидно, что ввиду использования только $1/G$ от всех пикселей для встраивания информации, предложенный алгоритм IM-QIM уступает аналогам по ёмкости контейнера. Однако немаловажным критерием является также стойкость встроенной информации к искажениям её носителя. При исследовании стойкости алгоритмов, основанных на QIM, часто рассматривают стойкость к зашумлению аддитивным белым гауссовским шумом (АБГШ) с нулевым средним и различными значениями дисперсии σ [20]. Более того, известны теоретические оценки стойкости некоторых модификаций QIM [2].

В рамках наших исследований была проанализирована стойкость к АБГШ трёх рассмотренных в данной работе алгоритмов (для IM-QIM, как и в предыдущем эксперименте, использовались значения G, равные 2, 4, 8 и 16). Эксперимент проводился на том же множестве из 100 изображений, встраивание осуществлялось при $\Delta = 10$. В качестве показателя стойкости бралась доля правильно извлечённых бит. Результаты исследования отражены на рис. 5.

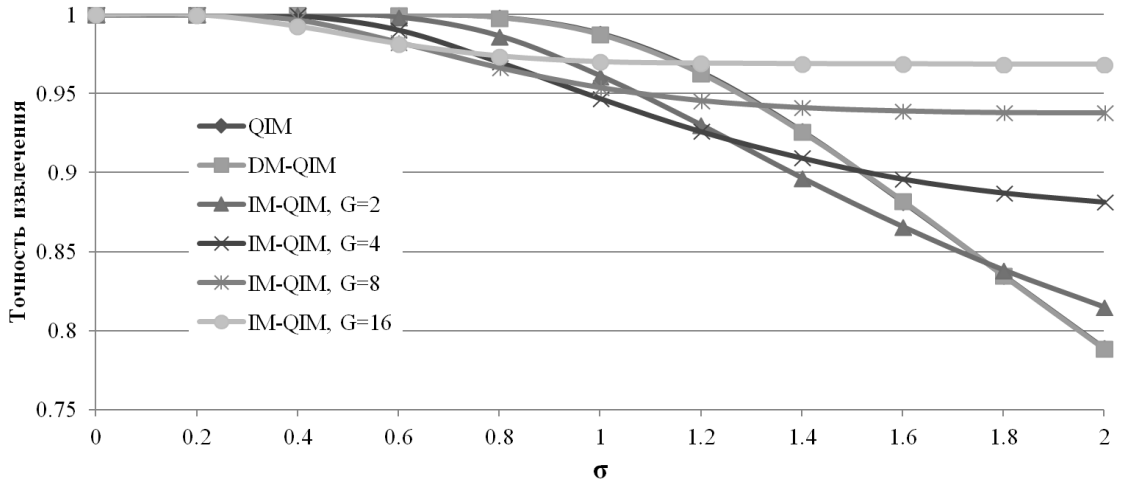


Рис. 5. Влияние АБГШ на точность извлечения информации для разных алгоритмов.

Графики показывают, что предложенный алгоритм IM-QIM превосходит по стойкости к АБГШ известные алгоритмы при больших значениях дисперсии шума, причём стойкость возрастает с увеличением числа G (которое обратно пропорционально ёмкости контейнера). Данный эффект можно объяснить тем, что при извлечении информации происходит суммирование отсчётов, подвергшихся зашумлению (формула (15)), что ввиду несмещённости шума снижает его влияние на точность извлечения информации.

5. Заключение

В работе предложен новый алгоритм встраивания ЦВЗ, относящийся к семейству алгоритмов QIM (алгоритмы управляемого переквантования). Основным преимуществом разработанного алгоритма является его стойкость к классу атак, основанных на выявлении статистической зависимости между яркостью отдельного бита ЦВЗ и отдельным битом ЦВЗ. Данное свойство алгоритма было достигнуто за счет использования так называемой «корреляционно-стойкой» функции сложения по модулю. Использование данной функции применительно к секретно выбираемым группам пикселей позволяет значительно затруднить атакующему поиск статистических зависимостей в анализируемых данных (изображениях и ЦВЗ). Так, при использовании значения $G = 4$ число попыток, необходимых для нахождения каких-либо 4-х пикселей, входящих в одну группу, может быть оценено как

$$\frac{C^4_{N\times M}}{(N\times M)/4} = \frac{4(N\times M-1)!}{4!(N\times M-4)!}. \qquad (20)$$

Данное соотношение означает, что, например, для изображения размером 1000×1000 пикселей атакующий вынужден будет проанализировать гистограммы примерно $1\cdot10^{17}$ возможных групп пикселей прежде чем найдет группу, необходимую для вычисления бита ЦВЗ.

Описанные в разделе 4 экспериментальные исследования показали, что предлагаемый алгоритм обладает следующими свойствами:

- обеспечивает уровень искажения изображения контейнера (MSE) при $\Delta > 10$ более низкий, чем базовые алгоритмы (причём он снижается при увеличении G);
- -обеспечивает информационную емкость контейнера $(N\cdot M)/G$ бит при использовании полутонового изображения-контейнера размером $N\times M$ пикселей;
- обеспечивает более высокую, по сравнению с базовыми алгоритмами, стойкость встроенного ЦВЗ к аддитивному белому гауссовскому шуму при больших значениях дисперсии шума σ.

Благодарности

Работа выполнена при поддержке РФФИ (гранты 15-07-05576, 16-37-00056 и 16-41-630676) и Минобрнауки РФ в рамках гранта президента РФ МК-1907.2017.9.

Литература

[1] Cox, I. Digital Watermarking and Steganography (2 ed.) / I. Cox, M. Miller, J. Bloom, J. Fridrich, T. Kalker. – San Francisco: Morgan Kaufmann Publishers Inc, 2009.

[2] Chen, B. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding / B. Chen and G.W. Wornell // IEEE Transactions on Information Theory. – 2001. – V. 47(4). – P. 1423–1443.

[3] Noda, H. High-performance JPEG steganography using quantization index modulation in DCT domain / H. Noda, M. Niimi, and E. Kawaguchi // Pattern Recognition Letters. – 2006. – V. 27(5). – P. 455–461.

[4] Jiang, Y. Adaptive spread transform QIM watermarking algorithm based on improved perceptual models / Y. Jiang, Y. Zhang, W. Pei, K. Wang // AEU - International Journal of Electronics and Communications. – 2013. – V. 67(8). – P. 690–696.

[5] Phadikar, A. Multibit quantization index modulation: A high-rate robust data-hiding method / A. Phadikar // Journal of King Saud University - Computer and Information Sciences. – 2013. – V. 25(2). – P. 163–171.

[6] Hakka, M. DCT-OFDM based watermarking scheme robust against clipping attack / M. Hakka, M. Kuribayashi, and M. Morii // IEICE Technical Report. – 2014. – V. 113(291). – P. 107–112.

[7] Fang, Y. CDMA-based watermarking resisting to cropping / Y. Fang, J. Huang, S. Wu // Proceedings of 2004 International Symposium on Circuits and Systems. – 2004. – V. 2. – P. 25–28.

[8] Huang, Y. B. A Dither Modulation Audio Watermarking Algorithm Based on HAS / Y. B. Huang, Q. Y. Zhang, Z. Liu, Y. J. Di, Z. T. Yuan // Research Journal of Applied Sciences, Engineering and Technology. – 2012. – V. 4(21). – P. 4206-4211.

[9] Khademi, N. Audio watermarking based on quantization index modulation in the frequency domain / N. Khademi, M. A. Akhaee, S. M. Ahadi, M. Moradi, A. Kashi // IEEE International Conference on Signal Processing and Communications. – 2007. – P. 1127-1130.

[10] Zolotavkin, Y. A new two-dimensional quantization method for digital image watermarking / Y. Zolotavkin, M. Juhola // 2015 17th International Conference on Advanced Communication Technology (ICACT). – 2015. – P. 155-160.

[11] Matam, B.R. Watermarking: How secure is the DM-QIM embedding technique? / B. R. Matam, D. Lowe // 16th International Conference on Digital Signal Processing. – 2009. – P. 1-8.

[12] Matam, B.R. Watermark-only security attack on DM-QIM watermarking: Vulnerability to guided key guessing / B. R. Matam, D. Lowe // Crime Prevention Technologies and Applications for Advancing Criminal Investigation. – 2012. – P. 85-99.

[13] Wang, Y. Steganalysis of block-structured stegotext / Y. Wang, P. Moulin. – Electronic Imaging. – 2012. – P. 477-488.

[14] Mitekin, V. A new key recovery attack against DM-QIM image watermarking algorithm / V. Mitekin // Proceedings of SPIE, International Conference on Machine Vision 2016. – 2017 (in press).

[15] Глумов, Н.И. Алгоритм встраивания полухрупких цифровых водяных знаков для задач аутентификации изображений и скрытой передачи информации / Глумов Н. И., Митекин В. А. // Компьютерная оптика. – 2011. – Т. 35(2). – С. 262-267.

[16] Митекин, В.А. Метод встраивания информации повышенной ёмкости в видео, стойкий к ошибкам потери синхронизации / В.А. Митекин, В.А. Федосеев // Компьютерная оптика. – 2014. – Т. 38(3). – С. 564-573.

[17] Mitekin, V. A new method for high-capacity information hiding in video robust against temporal desynchronization / V. Mitekin, V. Fedoseev // Proceedings of SPIE. – 2015. – V. 9445. – P. 94451A. – DOI: 10.1117/12.2180550.

[18] Siegenthaler, T. Correlation-immunity of nonlinear combining functions for cryptographic applications / T. Siegenthaler // IEEE Transactions on Information theory. – 1984. – Т. 30. – №. 5. – С. 776-780.

[19] BOWS-2 Contest - Image Dataset [Электронный ресурс]. – Режим доступа: http://bows2.ec-lille.fr (01.02.2017).

[20] Barni, M. Watermarking systems engineering: enabling digital assets security and other applications / M. Barni, F. Bartolini. – CRC Press, 2004.