ИССЛЕДОВАНИЕ МЕТОДА СТЕГОАНАЛИЗА ЦИФРОВЫХ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ, ОСНОВАННОГО НА ФЕНОМЕНЕ ЛИНЕЙНОЙ КОЛЛИЗИИ

К.А. Ахрамеева[1*], **Л.Г. Попов**[1]

[1]Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация
*Адрес для переписки: oklaba@mail.ru

Информация о статье
УДК 004.021
Язык статьи – русский

Ссылка для цитирования: Ахрамеева К.А., Попов Л.Г. Исследование метода стегоанализа цифровых видеопоследовательностей, основанного на феномене линейной коллизии // Труды учебных заведений связи. 2017. Т. 3. № 2. С. 28–36.

Аннотация: *В статье представлено подробное рассмотрение метода стегоанализа цифровых видеопоследовательностей, основанного на феномене линейной коллизии. Описывается условие статистической невидимости, служащее основой для разработки устойчивого к атаке линейной коллизии цифрового стеговложения. Продемонстрированы результаты экспериментальной проверки (в виде таблиц) эффективности работы, написанной на языке С++ программы, реализующей атаку (2-го типа) по удалению стеговложения. В исследовании используются различные форматы и типы видеопоследовательностей, отличающиеся количеством кадров в секунду, скоростью смены сцен, степенью взаимной корреляции между кадрами, битрейтом и т. д.*

Ключевые слова: *стегоанализ, цифровая видеопоследовательность, цифровое стеговложение, феномен линейной коллизии, статистическая невидимость, временные корреляции, коллизии 1-го и 2-го типов, разложение на кадры, линейное сложение, контейнер, битрейт, уровни избыточности и шума, сжатие, фильтр.*

RESEARCH OF DIGITAL VIDEO STEGANALYSIS METHOD FOUNDED ON THE PHENOMENON OF LINEAR COLLUSION

K. Ahrameeva[1], **L. Popov**[1]

[1]The Bonch-Bruevich Saint-Petersburg State University of Telecommunication, St. Petersburg, 193232, Russian Federation

Article info
Article in Russian

For citation: Ahrameeva K., Popov L. Research of Digital Video Steganalysis Method Founded on the Phenomenon of Linear Collusion // Proceedings of Educational Institutions of Communication. 2017. Vol 3. Iss. 2. PP. 28–36.

Abstract: *A detailed consideration of the method of digital video steganalysis founded on the phenomenon of linear collusion is represented in this paper. A statistical invisibility condition, which*

is a foundation of developing collusion-resistant embedding of the secret binary message in a digital video, is described. Tables demonstrate the results of experimental verification of the efficiency of the program written by C++. This program implements Type II collusion to delete a covert data from a digital video. Different formats and types of video, which differs from each other by number of frames per second, the rate of motion, correlation coefficients between frames, bitrate and so on, are used in this research.

Keywords: *steganalysis, digital video sequence, digital covert data, the phenomenon of linear collusion, statistical invisibility, temporal correlations, Type I and Type II collusions, decomposition into frames, linear addition, container, bitrate, redundancy and noise levels, compression, filter.*

Известно, что в исходном (необработанном) виде видеопотоки – это последовательности кадров изображения, поэтому сложность и гибкость пространства для вложения дополнительной информации в видео значительно больше, чем для изображений из-за присутствия пространства времени. Этот дополнительный объём позволяет нагрузке быть более избыточной и надёжно вложенной, что возможно за счёт использования более сложных временных маскирующих характеристик человеческого восприятия. Кроме того, атакующий имеет свободу шире использовать корреляции в объёме сигнала, чтобы разработать более эффективную оценку цифрового стеговложения или атаку удаления. Один важный класс таких атак известен как линейная коллизия множества кадров [1, 2]. В общем случае коллизия может быть линейной или нелинейной, используя в своих интересах сходства и различия среди кадров, чтобы уменьшить энергию стеговложения по отношению к тому, что является покрывающей информацией.

Два типа линейной коллизии

Коллизия появляется, когда коллекция видеокадров комбинируется с конечной целью получить свободную от вложения копию оригинала. Фреймы могут формировать продолжительную во времени подпоследовательность или происходить из сильно различающихся частей видео. Ключевая идея – это использование временной избыточности исходного видео либо стеговложения, чтобы оценить избыточный компонент. В случае линейной коллизии над множеством кадров отдельные видеокадры с вложением масштабируются и суммируются, чтобы сформировать результирующий фрейм, который представляет оценку оригинального кадра или стеговложения. Если масштабирование для всех кадров одинаково, тогда общая атака состоит из усреднения набора видеокадров с вложением. Интуитивно, эта операция имеет эффект усиления компонента стеговложения или исходного видео, который повторяется от кадра к кадру и ослабления того, который отличается.

Линейная коллизия первого типа возникает, когда огромное количество визуально непохожих видеокадров помечены линейной комбинацией цифровых стеганографических вложений с фиксированным шаблоном. Это обычно встречается во многих существующих вложениях для видео [3–5].

Коллизия второго типа возникает, когда огромное количество визуально похожих кадров помечены линейной комбинацией цифровых стеговложений с независимыми шаблонами. Примером такой атаки может служить усреднение кадров. Этот случай соответствует, например, вложениям, которые используют разные двумерные псевдошумовые последовательности, чтобы пометить каждый фрейм [6]. Рассмотрим определение линейной коллизии [7].

Определение 1. Даны две случайные величины A и B с определёнными математическим ожиданием и дисперсией; считается, что A – это с ϵ – оптимальной среднеквадратической ошибкой (MSE – Mean Square Error) оценка B, если и только если:

$$E\left[\left(\hat{A}-B\right)^2\right] < \epsilon, \tag{1}$$

где E – оператор математического ожидания; $\hat{A} = \sqrt{\mathrm{var}(B)/\mathrm{var}(A)}(A - EA) + EB$ и $\mathrm{var}(\cdot)$ – дисперсия аргумента (случайной величины).

В определении 1 случайная переменная A нормализуется, чтобы отразить те же самые математическое ожидание и дисперсию как у B. Также требуется:

$$0 < \mathrm{var}(A), \mathrm{var}(B), EA, |EB| < \infty. \tag{2}$$

Следующее определение из [7] определяет класс рассматриваемых атак.

Определение 2. Допустим, дан ряд видеокадров с вложением $X_k = U_k + \alpha_k W_k,\ k = 1, \ldots, n$, где U_k – исходный k-ый фрейм, α_k – глубина вложения для k-го фрейма, W_k – k-ый компонент стеговложения. Тогда линейная коллизия – это процесс формирования линейной комбинации фреймов:

$$\bar{X} = \sum_{k=1}^{n} \beta_k X_k = \sum_{k=1}^{n} \beta_k U_k + \sum_{k=1}^{n} \beta_k \alpha_k W_k, \tag{3}$$

где β_k – некоторый коэффициент.

При этом $\bar{X}$ представляет собой оценку с ϵ-оптимальной среднеквадратической ошибкой:

1) компонента стеговложения $\bar{W} = \sum_{k=1}^{n} \beta_k \alpha_k W_k$;

2) исходного компонента $\bar{U} = \sum_{k=1}^{n} \beta_k U_k$.

В случае 1) возникает атака коллизии 1-го типа; в случае 2) – атака коллизии 2-го типа.

Определение 2 выражает в краткой форме атаку усреднения кадров, принимая $\beta_k = 1/n$.

Суждение 1. Предполагая, что масштабированные элементы стеганографического вложения $\alpha_k W_k$ не зависят от оригинальных кадров U_k, необходимое условие для каждой из двух форм линейной коллизии, описанных выше, даётся в следующем виде [7]:

$$\rho(\bar{X}, \bar{U}) < \sqrt{\tilde{\epsilon}(2-\tilde{\epsilon})}\ (\text{Тип 1}); \tag{4}$$

$$\rho(\bar{X}, \bar{U}) > 1 - \tilde{\epsilon} \text{ (Тип 2)}, \quad (5)$$

где $\bar{X} = \sum_{k=1}^{n} \beta_k X_k$, $\bar{U} = \sum_{k=1}^{n} \beta_k U_k$, $\tilde{\epsilon} = \epsilon/\big(2\mathrm{var}(B)\big)$, то есть:

$\rho(\bar{X}, \bar{U}) > \sqrt{\tilde{\epsilon}(2 - \tilde{\epsilon})}$ – линейная коллизия 1-го типа невозможна;

$\rho(\bar{X}, \bar{U}) < 1 - \tilde{\epsilon}$ – линейная коллизия 2-го типа невозможна.

Как в случае коллизии 1-го типа, суждение 1 даёт только необходимое условие для коллизии 2-го типа. Условие недостаточно, поскольку надлежащая разработка стеговложения может (как показано далее) предоставить защиту против этих атак.

В нашем случае при создании программного фильтра оценка последовательности кадров осуществляется при помощи частного случая линейной коллизии, который называется простая линейная коллизия, так как коллизионные веса (collusion weights) применяются для всех фреймов одинаковые. Имеется скользящее окно, чтобы обозначить временное соседство, используемое для усреднения кадров. Предполагается, что это окно содержит визуально похожие фреймы. Берётся окно размером $2L + 1$ фреймов сосредоточенное на фрейме k, чтобы усреднить видеопоследовательность. Оценка k-го фрейма определяется следующим выражением:

$$\widehat{U}_k = \mathfrak{C}_P(Y_k) = \begin{cases} \dfrac{1}{2L+1} \displaystyle\sum_{i=1}^{2L+1} Y_i, & 1 \le k \le L; \\ \dfrac{1}{2L+1} \displaystyle\sum_{i=k-L}^{k+L} Y_i, & L < k \le N - L; \\ \dfrac{1}{2L+1} \displaystyle\sum_{i=N-2L}^{N} Y_i, & N - L < k \le N, \end{cases} \quad (6)$$

где $\mathfrak{C}_P$ – оператор коллизии с параметром P, являющимся длиной окна коллизии; Y_i – кадры, в которых предположительно могут быть вложения; $\widehat{U}_k$ – оценка исходного кадра U_k; N – общее количество кадров в видеопоследовательности; L – некоторое число.

Стоит отметить, что общие формы оператора коллизии $\mathfrak{C}$ представляют из себя попиксельное получение максимума, минимума, математического ожидания или среднего значения над рядом кадров видеоизображения [7–9].

Эффективность оценки $\widehat{U}_k$ как аппроксимация U_k зависит от величины L в связи со скоростью движения в видеопоследовательности. Поэтому, если коллизия применяется к данной видеопоследовательности Y_k, которая может содержать, а может и не содержать стеговложение, то полагается, что в обоих случаях для медленно меняющегося контента и правильно выбранного значения L, результат будет аппроксимацией (приближением) U_k. Таким образом, если секретное сообщение погружено в видео, то вычитание $\widehat{U}_k$ из Y_k даёт $Y_k - \widehat{U}_k \approx Y_k - U_k = \alpha W_k$ – оценку масштабированного Гауссовского стеговложения с нулевым математическим ожиданием. Если в Y_k не присутствует дополни-

тельная информация, то результат будет независим от любой характеристики такой, например, как гауссовость, которая предполагается по отношению к стеговложению. Это различие используется классификатором образцов, подробное рассмотрение которого выходит за рамки данной работы.

Из всего выше сказанного видно, как статистическая избыточность в покрывающем видео может помочь стегоаналитику в обнаружении скрытых данных. Увеличенная межкадровая корреляция улучшает производительность коллизии.

Статистическая невидимость

Дана последовательность оригинальных видеокадров U_k , $k = 1, \dots, n$ и последовательность видеокадров с вложением $X_k = U_k + \alpha_k W_k$, предполагается, что компонент стеговложения W_k, погружённый в видео статистически невидим, если и только если коэффициент корреляции между любыми двумя оригинальными кадрами a и b равен коэффициенту корреляции между двумя соответствующими кадрами с вложением, то есть $\rho(U_a, U_b) = \rho(X_a, X_b) \forall a, b \in \{1, \dots, n\}$.

Это свойство называется статистической невидимостью, поскольку атакующий, анализирующий видеопоследовательность на покадровой основе, не замечает какой-либо статистической разницы между оригинальной последовательностью и видео с вложением.

В [7] предлагается теорема о взаимосвязи между статистической невидимостью и многокадровой коллизией. Однако перед знакомством с ней необходимо сделать следующие предположения:

1) Видеокадры имеют общее конечное матожидание и дисперсию (среднюю энергию), то есть $E(U_k) = \mu_U$ и $\mathrm{var}(U_k) = \sigma_U^2$.

2) Элементы стеговложения W_k имеют нулевое матожидание, то есть $E(W_k) = 0$ и имеют общую ненулевую конечную дисперсию $\sigma_W^2 > 0$. Следовательно, $E(X_k) = E(U_k)$.

3) Глубины вложения α_k имеют конечный второй момент $E(\alpha^2)$.

4) Элементы стеговложения W_k, глубины вложения α_k и исходные кадры U_k независимы друг от друга.

Теорема 1. По предположениям (1)–(4) следующие утверждения эквивалентны:

$$\rho(X_a, X_b) = \rho(U_a, U_b) \forall a, b \in \{1,2, \dots, n\}; \quad (7)$$

$$\rho(U_a, U_b) = (E\alpha_a\alpha_b / E\alpha^2)\rho(W_a, W_b) \forall a, b \in \{1,2, \dots, n\}; \quad (8)$$

$$\rho(\bar{X}, \bar{U}) = \rho(X_a, U_a) \forall a \in \{1,2, \dots, n\}, \quad (9)$$

где ρ – коэффициент корреляции.

Свойство (7) описывает условие статистической невидимости; свойство (8) определяет зависимый от оригинального видео критерий создания сте-

говложения; свойство (9) гарантирует, что стеговложение, удовлетворяющее данным критериям, обладает статистической устойчивостью к атакам линейной коллизии 1-го и 2-го типов.

Таким образом, при разработке устойчивого к атаке линейной коллизии цифрового стеганографического вложения необходимо учитывать рассмотренные выше условия.

Фильтр *Collusion*

Для практического исследования эффективности линейной коллизии по удалению цифрового стеговложения (атака 2-го типа) погружённого в видеопоследовательность, был создан программный фильтр под названием Collusion (см. рис. 1).

Рисунок 1. Общая схема реализуемого алгоритма фильтрации видеопоследовательностей

В качестве объектов обработки используются видеофайлы с расширением AVI (Audio Video Interleave). Для погружения используется текстовый файл с расширением .txt. С помощью свободно распространяемой утилиты MSU StegoVideo [10] создаётся стегосистема со «слепым» декодером, так как на приёмной стороне при извлечении скрытой информации покрывающее сообщение неизвестно. Алгоритм вложения также не известен («черный ящик»).

В результате экспериментов по вложению текста с последующим сжатием кодеком H.264 выяснилось, что информация теряется полностью вне зависимости от используемого битрейта и уровня избыточности. Поэтому было решено использовать кодек Xvid с различными профилями, а также кодек Lagarith Lossless, осуществляющий сжатие без потерь.

При моделированиях использовалась длина окна коллизии $Z = 2L + 1 = 5$, которая является оптимальной, поскольку полагается, что глубина вложения секретного сообщения небольшая. В результате осуществляемой обработки видеопоследовательностей извлечение смысловой информации с помощью программы MSU становится затруднительным или полностью невозможным, что демонстрирует рис. 2.

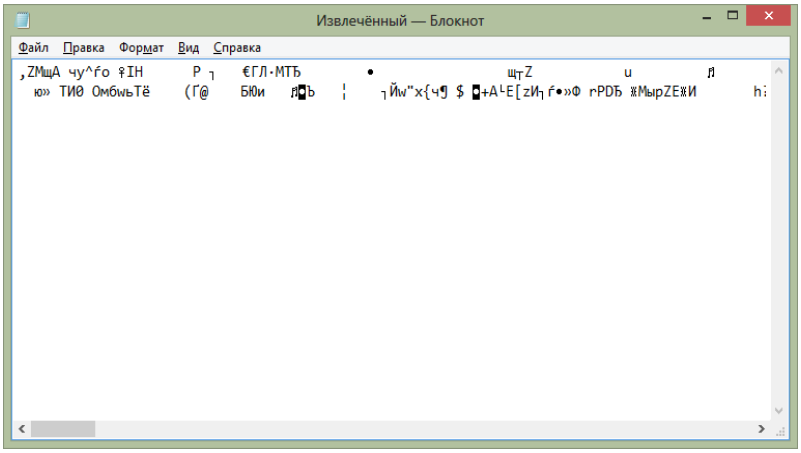

Рисунок 2. Результат извлечения

Ниже в таблицах 1–5 приведены результаты моделирования.

Таблица 1. Сравнение кодеков

№ п/п (название видеоролика)	Название видеокодека					
	Xvid 1.3.4				Lagarith Lossless	
	XvidHome		DivX 720HD			
	Ошибки до атаки, %	Ошибки после атаки, %	Ошибки до атаки, %	Ошибки после атаки, %	Ошибки до атаки, %	Ошибки после атаки, %
1. (busta–kamennye cvety.avi)	17,64	100	11,76	100	0	100
2. (chas pik.avi)	5,88	100	11,76	100	0	100
3. (simoni muha.avi)	94,08	пустой файл	100	пустой файл	0	100
4. (kolyaska.avi)	52,92	100	82,32	100	0	100
5. (ozvuchka.avi)	17,64	100	23,52	100	0	100

Таблица 2. Битрейт (низкая скорость смены сцен)

Битрейт, kbit/s	Название видеокодека					
	Xvid 1.3.4					
	XvidHome		XvidHD 720		DivX 720HD	
	Ошибки до атаки, %	Ошибки после атаки, %	Ошибки до атаки, %	Ошибки после атаки, %	Ошибки до атаки, %	Ошибки после атаки, %
250	23,52	100	23,52	100	23,52	100
500	23,52	100	23,52	100	23,52	100
1000	23,52	100	23,52	100	23,52	100
1500	23,52	100	23,52	100	23,52	100
2000	23,52	100	23,52	100	23,52	100
2500	23.52	100	23,52	100	23,52	100
3000	23,52	100	23,52	100	23,52	100

Таблица 3. Битрейт (высокая скорость смены сцен, пониженный коэффициент корреляции *R* между кадрами)

Битрейт, kbit/s	Название видеокодека	
	Xvid 1.3.4	
	XvidHome	
	Ошибки до атаки, %	Ошибки после атаки, %
250	0	100
500	17,64	100
1000	11,76	100
1500	0	100
2000	0	100
2500	0	100
3000	0	100

Таблица 4. Количество кадров в секунду (fps – frames per second)

Frames per second (fps)	**Название видеокодека**	
	Xvid 1.3.4	
	XvidHome	
	Ошибки до атаки, %	Ошибки после атаки, %
24	23,52	100
25	23,52	100
30	0	100
60	0	100
120	0	100
200	0	100

Таблица 5. Возможный объём вложения в зависимости от битрейта

Битрейт, kbit/s	**Название видеокодека**	
	Xvid 1.3.4	
	XvidHome	
	Количествово символов без пробелов, зн.	Количествово символов с пробелами, зн.
250	782	1694
500	807	1694
1000	974	1695
1500	974	1696
2000	1070	1696
2500	1229	1697
3000	1303	1702
5000	1359	1705
10000	1359	1705

На рис. 3 представлено сравнение качества изображения до (слева) и после (справа) атаки. Наблюдаются незначительные искажения исходного изображения, приемлемые с точки зрения стороннего наблюдателя.

Рисунок 3. Сравнение качества изображения: слева – до атаки, справа – после атаки

Выводы

1) Рассмотренный метод простой линейной коллизии подходит для практической реализации в приложениях реального времени и показывает хорошие результаты по удалению (для типовых видеофайлов).

2) Чем больше избыточности в видео, тем больший объём информации можно вложить при меньших потерях.

3) Предпочтительнее использовать lossless кодеки.

4) С увеличением битрейта при извлечении наблюдается уменьшение количества нераспознанных символов (при быстрой скорости смены сцен).

5) Ниже 5000 kbit/s увеличивается количество ошибок.

6) Выше 5000 kbit/s ошибок нет, но отсутствует увеличение количества вложенных бит.

7) Для большего объёма вложения разумнее увеличивать количество кадров при сохранении приемлемого значения битрейта.

Список используемых источников

1. Budhia U., Kundur D., Zourntos T. Digital Video Steganalysis Exploiting Statistical Visibility in the Temporal Domain // IEEE Transactions on Information Forensics and Security. Dec. 2006. Vol. 1. Iss. 4. PP. 502–516.

2. Небаева К.А., Попов Л.Г. Исследование методов стегоанализа цифровых видеопоследовательностей // Актуальные проблемы инфотелекоммуникаций в науке и образовании. V Международная научно-техническая и научно-методическая конференция: сб. научных статей. СПб.: СПбГУТ, 2016. С. 489–493.

3. Hartung F., Eisert P., Girod B. Digital Watermarking of MPEG-4 Facial Animation Parameters // Comput. Graph. Jul.–Aug. 1998. Vol. 22. Iss. 4. PP. 425–435.

4. Kalker T., Depovere G., Haitsma J., Maes M. A Video Watermarking System for Broadcast Monitoring // Proc. SPIE. Jan. 1999. Vol. 3657. PP. 103–112.

5. Cox I.J., Kilian J., Leighton F.T., Shamoon T. Secure Spread Spectrum Watermarking for Multimedia // IEEE Trans. Image Process. Dec. 1997. Vol. 6. PP. 1673–1687.

6. Mobasseri B.G. Exploring CDMA for Watermarking of Digital Video // Proc. SPIE. Jan. 1999. Vol. 3657. PP. 96–102.

7. Su K., Kundur D., Hatzinakos D. Statistical Invisibility for Collusion-Resistant Digital Video Watermarking // IEEE Trans. Multimedia. Feb. 2005. Vol. 7. Iss. 1. PP. 43–51.

8. Trappe W., Wu M., Wang Z., Liu K. Anti-collusion Fingerprinting for Multimedia // IEEE Trans. Signal Process. Apr. 2003. Vol. 51. Iss. 4. PP. 1069–1087.

9. Wang Z., Wu M., Zhao H., Trappe W., Liu K. Collusion Resistance of Multimedia Fingerprinting Using Orthogonal Modulation // IEEE Trans. Image Process. Jun. 2005. Vol. 14. Iss. 6. PP. 804–821.

10. Всё о сжатии данных, изображений и видео / Проект, идеи: Ватолин Д., реализация: Петров О. MSU StegoVideo – уникальная утилита для встраивания информации в видео (фильтр для VirtualDub/отдельная программа) // URL: http://www.compression.ru/video/stego_video/filter_settings.html.