

The design of a steganographic system based on the internal MP4 file structures

M. Jókay

Abstract—We propose a new technique for inserting the information into the mp4 video files. This technique is based on a modification of the GOP (Group of Pictures) structure and alternation of the streams in the video file. The proposed steganographic system is implemented on the MP4 format. Although our method does not assume decoding of the video stream included in the MP4 file, its suitability depends on the used video encoder and more specially on the type of video scenes. A variable number of adjacent P and B frames in the individual GOPs (variable number of video frames in the MP4 chunk structures) is used to decode the inserted (hidden) information.

Keywords—GOP, I-frame, MP4 file, P-frame, steganography, video stream

I. INTRODUCTION

STEGANOGRAPHY deals with the design, analysis, and implementation of algorithms, methods, and systems, that enable:

- to transfer the information using an unsuspecting communication channel,
- to hide the existence of a message in a communication channel.

Steganography is one of the research areas that progresses rapidly. Before the expansion of digital communication, steganography was pursued only in the organizations with large budget and motivation for studying the technological aspects of hiding the transmitted messages. Therefore the usage of steganography in the near past was tightly connected with the usage and cryptanalysis of classical ciphers (see e.g. [1], [2], [3], and [4]). The distinction between cryptography and steganography can be seen on the schemas used for the transformation of the input data (see [5], [10], [11], [12], [14], [17], [18]).

The development of the Internet enabled a simple transmission of digital information such as music, pictures, texts, and various binary data. The capacities of the internet data storages enable to store and share large video sequences. Therefore the steganography allowing the insertion of a secret message into a video file became a focused research area nowadays.

A well known example of steganography is the insertion of the data into the least significant bits of the picture. Except using the redundant information for inserting a secret message to be hidden, it is possible to use also the techniques for generating the stego-medium, thus the structure itself contains the secret message. Although it seems that any structure can be used for steganography, not much attention was aimed at the connection of steganography with hard problems (e.g. [6], [7], [8], [9], [13], [15], [16]).

Nowadays, many devices, used for capturing video, can manipulate the video using the MPEG-4 standard (movie-cameras, cameras, mobile phones). Due to the high resolution in this standard, we decided to design and implement a steganographic system, that uses some features of the MPEG-4 encoder.

The structure of the paper is as follows: we describe several basic entities used by the MPEG-4 encoder in Section II. In the second section, we define a steganographic system and deal with the theoretical requirements that this system should meet. The design of the steganographic system that uses either the GOP structure or/and the length of the video chunks in the encoded video sequence is given in Section IV. In Section V we show the results of some measurements dedicated to suitability of the proposed steganographic system. Finally, possible problems and limitations of our steganographic system implementation are discussed.

II. THE MP4 STRUCTURES

Our steganographic system uses two structures that are present in the MP4 files, namely the periodically repeating groups of pictures (GOPs) and the interleaving of the video and audio bitstreams. The GOP structure is described in Section II-A and the interleaving of bitstreams in the MP4 files in Section II-B.

A. The GOP structure

GOP denotes the group of pictures. Since the MPEG-1 standard, several types of frames are used: I, B, and P frames. (Remark: I and P frames were already used in the H.261 standard, before the MPEG-1 standard.) I frames are transmitted encoded using the algorithms used for the JPEG files. It is the so-called intra-frame encoding, because an I frame contains the full picture information and does not need other frames in the GOP. Each GOP must start with an I frame. A GOP is defined to be a sequence of frames between two succeeding I frames. Fig. 1 shows the encapsulation of the GOP in the video stream. The video stream is composed of GOPs. A GOP

is formed by sequence of frames. Each frame consists from the slices of macroblocks, and each macroblock is a set of four 8x8 matrices.

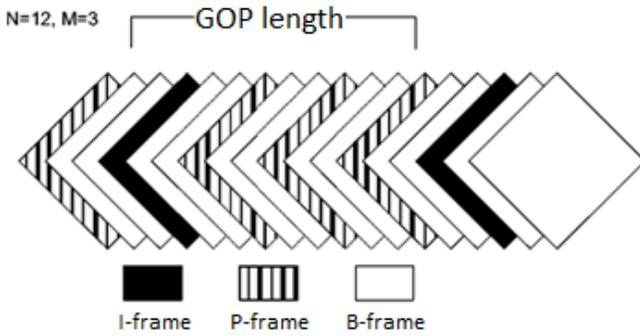


Fig. 2. Example of GOP with parameters $N = 12$ and $M = 3$.

P frames and B frames are outer-frame encoded. It means, they do not contain the full picture information, but only references to parts of the previous or succeeding frames. Each P frame is created using the previous I or P frame, the so-called forward prediction. Moreover, there is also the backward prediction used for the B frames. B frames are created using the bidirectional interpolation prediction. Firstly, the forward prediction is determined, then the backward. The encoder then makes a decision about the ratio for calculating the average from both of these predictions. The B frames are then not used for further predictions due to a large number of errors (the largest in comparison with I and P frames) they bring into the picture.

The number of used P and B frames in a GOP is not limited by the standard. The number of frames in a GOP determines its length, denoted by the parameter N . The distance between a P and an I frame is denoted by the parameter M . An example of a GOP with parameters N and M is shown in the Fig. 2.

B. Interleaving of bitstreams in the MP4 files

In fact, an MP4 file is a multimedia container for storing various data streams. The audio and video streams are the most commonly interleaved streams in it. The container/the MP4 file is specified by the international standard ISO 14496-14. The format is derived from the QuickTime container, developed by the Apple. The physical data organization in MP4 files is shown in the Fig. 3. All the data are organized into the so-called atoms. Each atom is identified by its type and length [21]. The atoms may contain other atoms. From the point of view of the our steganographic system, we are interested in the following atoms: moov, mdat, stbl, stts, stss, stsc, stsz, and stco. Moov encapsulates information on the structure of the data stored in the mdat atom. Except other important and control data, the moov atom contains also the stbl atom, that holds information tables about the individual stream patterns. The stts atom contains information on the time lengths of the individual streams, stss is formed by the list of frames, that can be accessed directly in the stream (i.e. without decoding other frames). The following table is stored in the stsc atom: each row holds information on the individual data streams. The stsz

atom contains the description of the physical placement of the streams, including their sizes. Finally, the stco atom keeps the start positions of the each group of data described by the stsc atom.

The audio and video data itself are mutually interleaved in the mdat atom. It is due to the necessity of the concurrent processing of more streams by the decoder. Therefore the time multiplex, that assembles the individual data from one type of a stream into the so-called chunks, was introduced. The structure of interleaved data in the mdat atom is shown in the Fig. 4. These chunks may have different sizes and different number of frames.

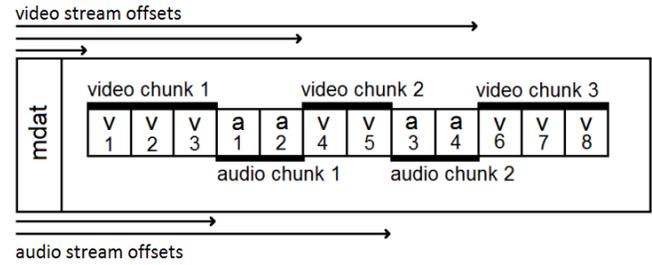


Fig. 4. Alternating of the streams in the MP4 file structure.

III. THE STEGANOGRAPHIC SYSTEM REQUIREMENTS

Definition 1: We define our steganographic system using the steganographic schema, determined by the two following transformations:

$$E : \mathcal{C} \times \mathcal{K} \times \mathcal{M} \rightarrow \mathcal{C} \quad (1)$$

$$D : \mathcal{C} \times \mathcal{K} \rightarrow \mathcal{M} \quad (2)$$

where \mathcal{C} is a set of cover-media, \mathcal{K} is a set of keys, and \mathcal{M} is the set of messages. $s = E(c, k, m)$ is a stego-medium with the following properties:

$$D(E(c, k, m), k) = m, \forall c \in \mathcal{C}, \forall k \in \mathcal{K}, \forall m \in \mathcal{M} \quad (3)$$

$$D(E(c, k_1, m), k_1) \neq D(E(c, k_2, m), k_2), \quad (4)$$

$$\forall k_1, k_2 \in \mathcal{K}, \forall m \in \mathcal{M}$$

$$D(E(c, k_1, m), k_2) \neq m, \quad (5)$$

$$\forall c \in \mathcal{C}, \forall k_1, k_2 \in \mathcal{K} : k_1 \neq k_2, \forall m \in \mathcal{M}$$

Definition 2: Next, we define the capacity of a cover, $cap = \log_2 |L|$, for the estimation of the efficiency of the steganographic system. L is the largest message, that can be inserted into the cover. The operator $|L|$ returns the length of the message, given by the number of symbols that represent it.

According to [19], steganography attempts to reach the three security goals: privacy, robustness, and undetectability. These properties are interconnected and it is very hard to achieve all

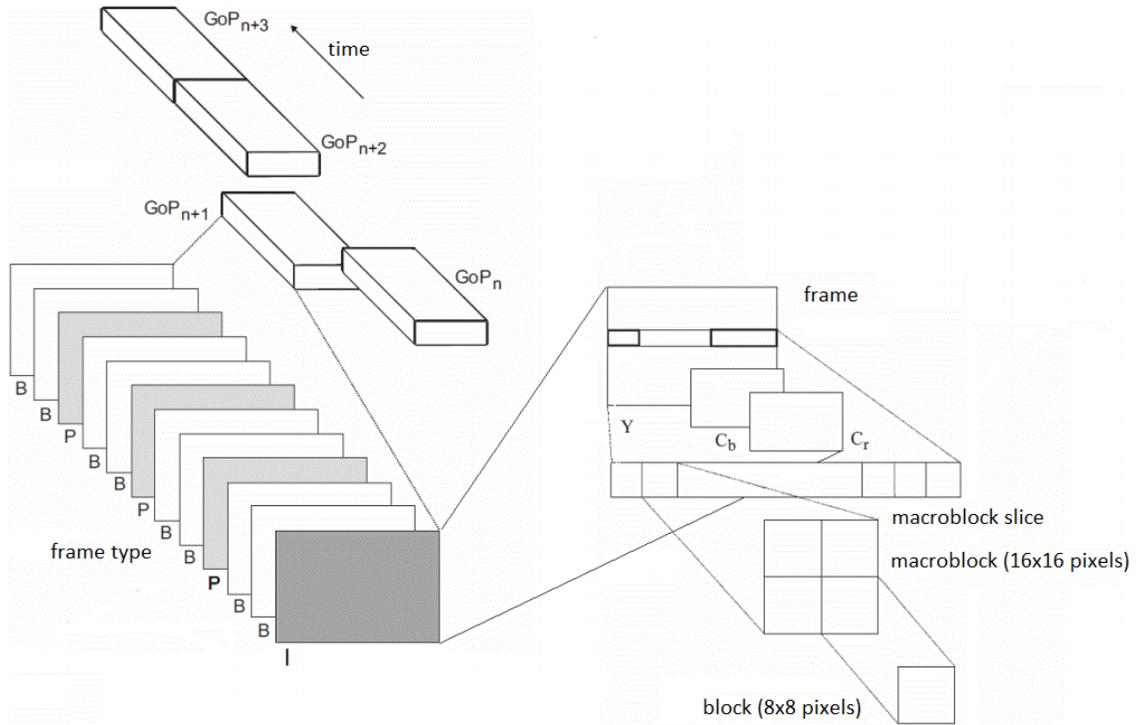


Fig. 1. The encapsulation of the GOP into video stream in MPEG-1 standard.

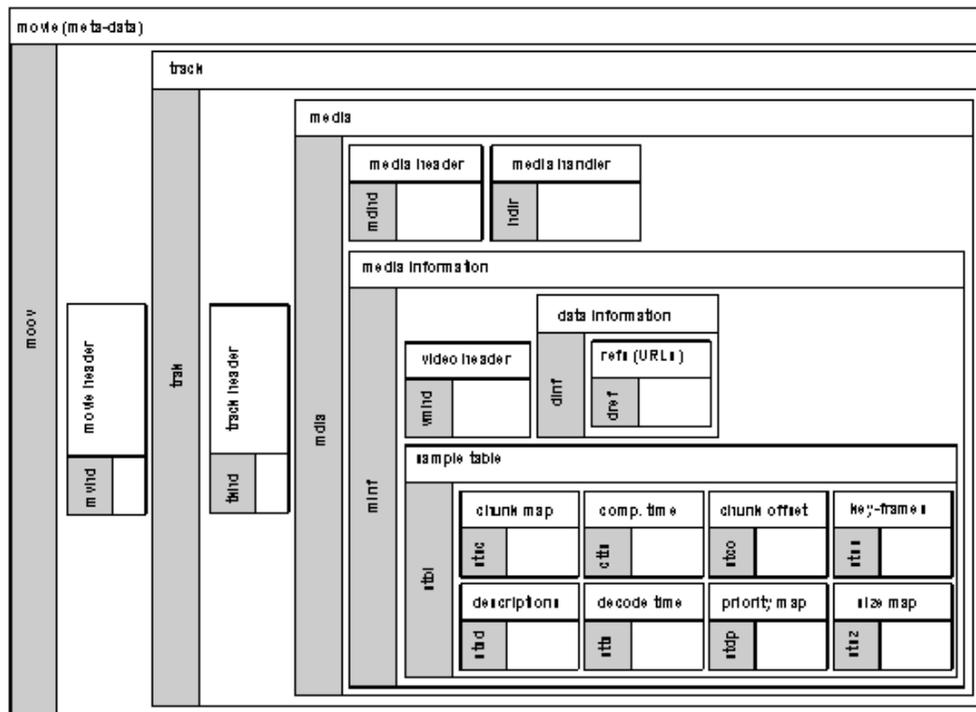


Fig. 3. The MP4 atoms structure.

of them together. As we can see (in the Fig. 5), to increase one of them it is necessary to decrease the others. We discuss the security level of each one for our proposed steganographic system in following sections.

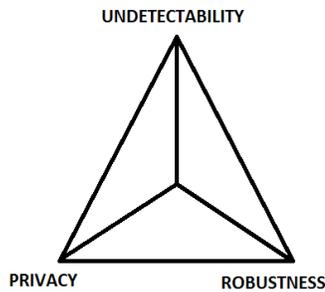


Fig. 5. The relations of the main security properties.

A. Privacy

Privacy means the ability to keep the secret, in such a sense, that an unauthorized person is not able to get the secret information.

The design of our steganographic system is built on the Kerckhoffs' principle: we assume that the used steganographic schema is known to the attacker. The privacy in our system is achieved using the encryption of the secret message using a properly chosen encryption algorithm parametrized by a secret key. The encrypted secret message is then inserted into the cover. Thus an attacker with the knowledge of the used steganographic system is not able to obtain the secret message without the secret encryption key. The randomness checking of the encrypted message can be done using various statistical tests, see e.g. [20] for a new one.

B. Robustness

Robustness is connected with the transmission of the stego-medium through a communication channel. The transmission itself can be subdivided into sending the message, the transmission through the communication channel, and finally, the receipt of the message by a legitimate user. Robustness characterizes the successfulness of the message transmission without damaging the inserted secret information. Therefore one of the research targets in steganography is to develop such algorithms for hiding messages into covers, that will ensure the integrity of the secret information even if the stego-medium is modified during its transmission through a communication channel.

Our proposed steganographic system uses the internal structure of the MPEG-4 standard GOP encoder for hiding a secret message. This structure is independent on the file type used for storing the video stream. Therefore our proposed steganographic system is resistant against a change of a file type, that preserves the video data. However, it is not resistant against reencoding the video using a different encoder. Technically, the system can not be resistant against this attack, because each encoder creates the GOP structures anew using the user specifications and input limitations.

C. Undetectability

The undetectability itself covers two requirements. The first one is the impossibility of detecting whether a medium con-

tains a hidden secret message only using the medium characteristics. The second one is the impossibility of confirming (or disproving) the existence of the hidden secret message in the stego-medium providing the attacker knows the technique used for the insertion of the secret message into the cover. Therefore the steganographic algorithm must be designed in such way that it does not change significantly the characteristics of the medium, that can be used for distinguishing a stego-medium from a medium with no hidden message inserted. Moreover, it should be impossible to determine whether a medium contains a secret (hidden) message or not, even if the steganographic algorithm is known.

In order to achieve the undetectability, an encoding of the inserted message that preserves the original statistical properties of the cover has to be chosen. The original statistical properties are those that are modified by inserting a secret message into the cover.

Our proposed steganographic system inserts the information using the GOP structure modification, as will be shown further in Section IV. Due to the privacy requirement it is not possible to insert a message into an arbitrary video encoded using the MPEG-4 standard. If the GOP structure is fixed, i.e. it does not change during the whole video sequence, then any manipulation with GOP is easy to be detected. Therefore it is necessary to check to what extent the GOP structure can be manipulated before inserting the secret message.

The second requirement concerning undetectability is to make it impossible to decide whether a medium (a video in our system) contains or does not contain a secret (hidden) information. This requirement is met using a proper encoding of the information. This encoding must preserve the characteristics of the randomness distribution of change dynamics of the GOP structure during the whole sequence of the used video. Details concerning the encoding are given in the next Section.

IV. DESIGN OF THE STEGANOGRAPHIC SYSTEM

A. Using of the GOP structure

The design is based on the fact, that the MPEG standard does not specify the number nor the order of P and B frames in the GOP structure. It does not specifies the number of frames in the individual chunks of the video and audio streams, as well. The main idea of our algorithm is to modify their number and/or their order in GOPs. The steganographic method successfulness assumption is the GOP variability introduced by the video encoder itself. If the video encoder uses a fixed GOP structure, it is not possible (using our method) to insert a secret information without making a suspicion.

The steganographic system, proposed by us, uses the both mentioned structures in a similar way. In the first case (using of the GOP structure) we encode a single bit of secret information using a change of the number of P/B frames in a GOP in our steganographic system. The change is done at the end of a GOP in order to avoid (minimize) the picture degradation introduced by inserting/deleting a P/B frame.

15	15	15	15	15	16	15	15	15	15	15	16	15	15	15	15	15	16	15	15	15	15	16	15	15	15	
0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0

Fig. 6. Example of the video chunks in the MP4 file.

B. Using of the chunk structure

In the second case (using of the number of frames in a video chunk) is the situation as follows. The data streams in the MP4 file are alternating. The number of frames in one chunk depends on the framerate of the encoder. For illustration (see Fig. 6) we mention the video stream with the chunks of size 15 or 16, respectively. If the framerate is static, then in the video stream there are only two possible numbers representing the count of the frames in chunks (in our example these are the numbers 15 and 16). This property is usable for encoding the original sequence into the related binary sequence: we change the first number into zero and the second into one. The first line in the Fig. 6 represents the count of the frames in the chunks and the second line is the related binary sequence.

The steganographic system based on this structure can alternate the number of frames in the individual chunks and thus encode the binary information. If the number of frames in some chunk of the video streams is changed, it is necessary to upgrade the relevant related control information in the stts, stss, stsc, stsz, and stco atoms of the MP4 file. For example (see Fig. 7), in the case of inserting a bit 0, we check the number of the frames in the corresponding chunk. If this number is even, we leave this chunk without a change. But if it is odd, we need to change the number of frames in this chunk to an even number. If we move one frame from next chunk in the stream, we need to update all the next chunks to avoid synchronization problems. This situation is shown in Fig. 7.

We can use this technique not only for the video stream, but also for the audio stream.

C. Encoding algorithm

The process of encoding the information is the same for both methods, therefore we explain the algorithm only for the GOP structure.

If the encoding, in which an even GOP length encodes a 0 bit and an odd GOP length encodes a 1 bit, is used, the following situations may occur:

- 1) If we want to insert a 0 bit:
 - a) the length of the GOP is even: no change is done,
 - b) the length of the GOP is odd: the last frame in the GOP is removed and the first frame following the next I frame is doubled to keep the synchronization.
- 2) If we want to insert a 1 bit:
 - a) the length of the GOP is odd: no change is done,
 - b) the length of the GOP is even: the last frame in the GOP is removed and the first frame following the next I frame is doubled to keep the synchronization.

Doubling the first frame following the I frame in the next GOP may introduce an error into the video. This error could be compensated by inserting a so-called empty B/P frame. This frame references the unmodified first frame in the GOP.

Algorithm for inserting the information into a video sequence:

- 1) Check the suitability of the video file (cover), i.e. the variability of GOPs in the cover is determined. The number of GOPs that can be used for encoding is determined.
- 2) If the variability is less than a decided threshold then quit.
- 3) Encrypt the information that should be inserted.
- 4) Determine the capacity *cap* of the cover. The capacity of the cover is determined by the number of GOPs that can be used for encoding the secret information. The capacity can only be estimated and not determined exactly for an arbitrary message. Encoding a single bit using a GOP can affect the length of the next GOP, thus the number of suitable GOPs can be increased/decreased.
- 5) If the capacity of the cover is insufficient then quit.
- 6) Encode the information into the video using the above described scheme based on the modification of GOP lengths.

Algorithm for decoding the information:

- 1) For all the GOPs in the video file decode the bits of the inserted information according to their lengths:
 - a) if the GOP length is even, then decode a 0 bit,
 - b) if the GOP length is odd, then decode a 1 bit.
- 2) Decrypt the obtained sequence.

V. MEASUREMENTS

The one of the main goals of the steganography is undetectability. The proposed steganographic system uses the chunk structures in the MP4 files, and these structures are correlated with the framerate. We have been interested in this property and we have examined 1000 videos from Youtube. All of them have the static framerate, thus in each video file from this internet service has only two values for length of the video chunks. We have analyzed this lengths, and in Fig. 8 are shown the results. 613 video files have the main length of the video chunks equal to 15, 232 have length 13, and 111 have the length 16. 95% of all the video files have the lengths 13, 15, or 16.

Next, we examined the characteristics of the binary sequences related to video chunks. The main result is that there is no video with uniform distribution. The histogram of the ratio (count of the zeros versus ones) for all the examined files is shown in the Fig. 9.

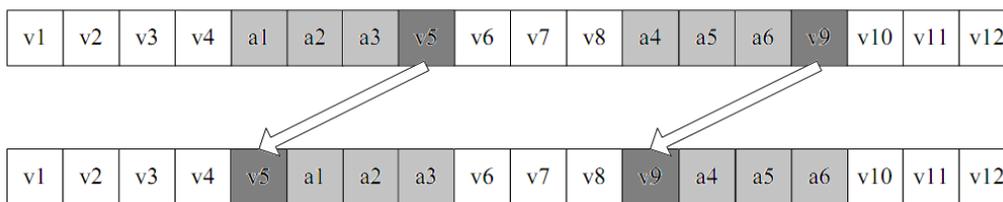


Fig. 7. Example of encoding one bit of the information into chunks.

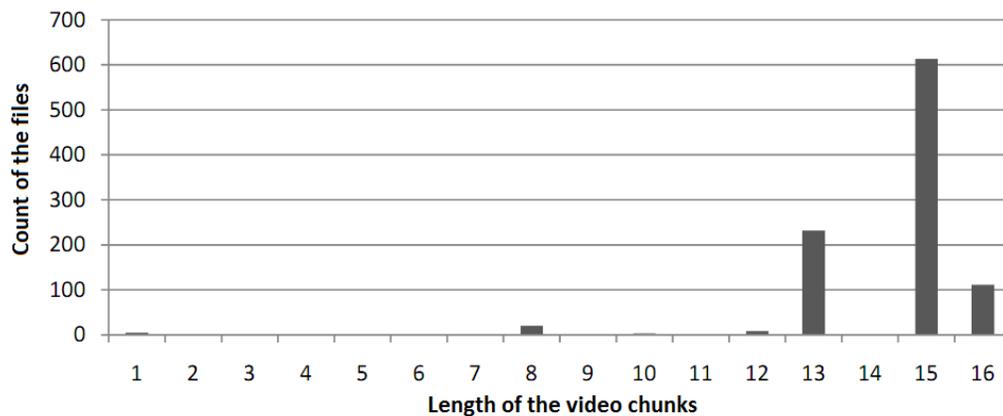


Fig. 8. The video files classified by the length of the video chunks.

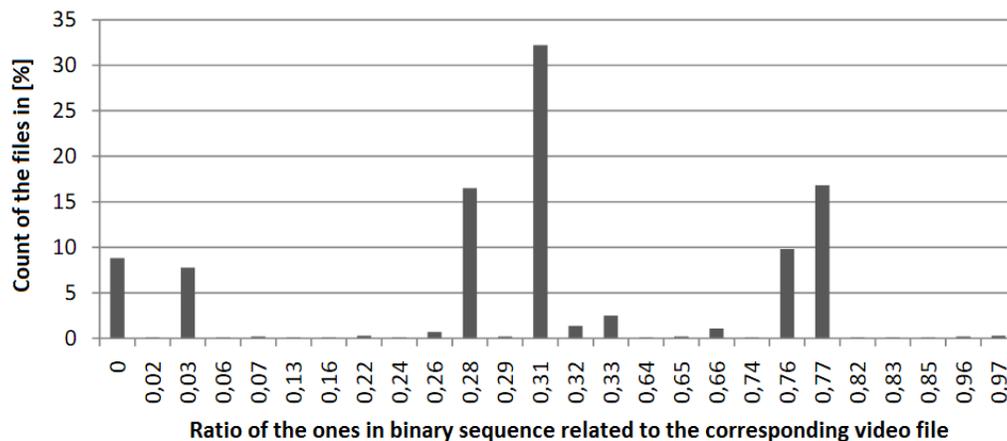


Fig. 9. Distribution of the video files according to ratio of zeroes in related binary sequence.

The binary sequence related to each of the examined video files can be mapped onto one of 33 different patterns (see Fig. 10). This pattern is repeated periodically in the whole video file. The reason is the static framerate used for encoding the corresponding video. As we can see, these patterns are not random. There is no possibility to use the structure of the video chunks for inserting the secret message without suspicion. If some message is inserted, the structure of the binary sequence is modified, and (with a high probability) does not correspond to any of the pattern classes (show in the Fig. 10). The attacker can examine only this characteristic to reveal the existence of the secret information.

As the last research in this area we have focused on the public (multimedia) Internet storage services. We examined

all of the available providers of this service. The goal of this research was to summarize the characteristics about supported video formats. If there are some available multimedia storages with property of dynamic framerate, we can use them to transport the medium (video file with a secret message encoded by our steganographic system) without any suspicion. The results can be seen in Fig. 11.

VI. CONCLUSION

The efficiency of the proposed steganographic system tightly depends on the content type of the video. If an almost static video is used (e.g. only a single object in the front changes and the background remains unchanged), the capacity of the cover is very small. For example, using a 14 minutes



Fig. 10. Patterns from all the examined video files.

service	FLV container				MP4 container			
	Format	FPS	Bitrate (Kbps)	Resolution	Format	FPS	Bitrate (Kbps)	Resolution
Blip.tv	VP6	24	400	1280x720				
Clipshack	H.263	30,3	458	320x240				
Dailymotion	H.263	29		320x180	AVC	Variable	Variable	512x288
Dotsub.com					AVC	Variable	Variable	420x316
Flickr					AVC	29,97	Variable	640x480
Glumbert	VP6	25	131	320x240				
Metacafe	AVC	30	664	640x360				
Openfilm					AVC	25	Variable	920x518
Sevenload	H.263	25	486	300x240				
Viddix	VP6	25	749	432x324				
Video.aol.com	H.263	25	464	468x352				
Vimeo					AVC	25	Variable	1280x720
Yahoo Video	VP6	30	636	400x222				
YouTube	AVC	24	1036	854x480	AVC	Variable	Variable	1920x1080

Fig. 11. List of the available multimedia storages on the Internet.

long static video, approximately 400 bits of a secret message (i.e. only a short 50 bytes long message) can be inserted. On the contrary, a dynamic video of the same length allows to insert many times more bits, because a dynamic video results in a more frequent reference I frame occurrences. It is due to the anti image degradation practice of the video encoder. The capacity is influenced by the dynamics of the video and also by the framerate of the encoder. If the framerate value is static, the steganographic capacity of the media is dramatically decreased, because it is impossible to modify the sizes of the individual chunks without making a suspicion. On the other hand, we can use available Internet providers of the multimedia storage services to the transport of the secret message in the video file encoded with a variable framerate.

REFERENCES

- [1] O. Grošek, M. Vojvoda and P. Zajac, *Klasické šifry (Classical ciphers)*, Bratislava, 2007, ISBN 978-80-227-2653-5, pp. 214 (in Slovak).
- [2] O. Grošek and P. Zajac, *Automated cryptanalysis of classical ciphers*, Encyclopedia of Artificial Intelligence, 2009, ISBN 978-1-59904-849-9, pp. 186–191.
- [3] O. Grošek and P. Zajac, *Automated cryptanalysis*, Encyclopedia of Artificial Intelligence, 2009, ISBN 978-1-59904-849-9, pp. 179–185.
- [4] O. Grošek, M. Vojvoda, M. Zanechal and P. Zajac, *Základy kryptografie (Elementary cryptography)*, Bratislava, 2006, ISBN 80-227-2415-7, pp. 184 (in Slovak).
- [5] O. Grošek and P. Zajac, *Efficient Selection of the AES-Class MixColumns Parameters*, WSEAS Transactions on Information Science and Applications, Vol. 4, Iss. 4, 2007, ISSN 1790-0832, pp. 663–668.
- [6] M. Sýs, *Isotopy Classes of Latin Squares*, Journal of Electrical Engineering, Vol. 58, No. 7/s, 2007, ISSN 1335-3632, pp. 97–99.
- [7] P. Zajac, *Discrete Logarithms and Degree Six Number Field Sieve : A practical Approach*, Saarbrücken : VDM Verlag Dr. Müller, 2009, ISBN 978-3-639-12331-9, pp. 91.
- [8] P. Zajac, *Remarks on the NFS complexity*, Tatra Mountains Mathematical Publications, Vol. 41, Bratislava : Slovak Academy of Sciences, 2008, ISSN 1210-3195, pp. 79–91.
- [9] P. Vagánek and P. Zajac, *Implementácia ECC v C# (ECC implementation in C#)*.NET, Časopis EE, Vol 14, No 5/s, 2008, pp. 197–201 (in Slovak).
- [10] O. Grošek and P. Zajac, *Graphs Connected with Block Ciphers*, WSEAS Transactions on Information Science and Applications, Vol. 3, Iss. 2, 2006, ISSN 1790-0832, pp. 439–443.
- [11] M. Vojvoda, *On the Hash Functions*, Begabtenförderung im MINT-Bereich. Mathematik, Informatik, Naturwissenschaften, Technik : Band 14, Ulm : Aegis, 2006, pp. 15–26.
- [12] M. Sýs and P. Zajac, *Discrete Logarithm Problem and Its Applications in Cryptography*, Begabtenförderung im MINT-Bereich 12, Ulm : Aegis, 2005, pp. 129–146.
- [13] O. Grošek and M. Sýs, *Isotopy of Latin Squares in Cryptography*, Tatra Mountains Mathematical Publications, Vol. 45 : NILCRYPT '10, 2010, ISSN 1210-3195, pp. 27–35.
- [14] I. Semaev and M. Mikuš, *Methods to Solve Algebraic Equations in Cryptanalysis*, Tatra Mountains Mathematical Publications, Vol. 45 : NILCRYPT '10, 2010, ISSN 1210-3195, pp. 107–136.
- [15] I. Slaminková and M. Vojvoda, *Cryptanalysis of a Hash Function Based on Isotopy of Quasigroups*, Tatra Mountains Mathematical Publications, Vol. 45 : NILCRYPT '10, 2010, ISSN 1210-3195, pp. 137–149.
- [16] P. Zajac, *Generalized Line Sieve Algorithm*, ELITECH 07, The 9th Conference for PhD Students Electrical Engineering and Information Technology : Bratislava, Slovak Republic, Bratislava, 2007, ISBN 978-80-227-2655-9.
- [17] O. Grošek and P. Zajac, *Searching for a Different AES-Class Mix-Columns Operation*, Proceedings of the WSEAS Conference : 6th International Conference on Applied Computer Science, WSEAS Press, 2006, ISBN 960-8457-57-2, pp. 307–310.
- [18] O. Grošek and P. Zajac, *A remark to minimal graphs connected with block ciphers*, Proceedings of 4th WSEAS International Conference on Information Security, Communications and Computers (ISCOCO 2005), December 16-18, 2005, Tenerife, pp. 78–82.
- [19] E. Cole, *Hiding in Plain Sight — Steganography and the Art of Covert Communication*, New York : Wiley John & Sons, 2003, ISBN 0471444499.
- [20] O. Grošek, M. Vojvoda and R. Krchnavý, *A New Matrix Test for Randomness*, Computing, Vol. 85, 2009, ISSN 0010-485X, pp. 21–36.
- [21] P. Fernando and T. Ebrahimi, *The MPEG-4 book*, Upper Saddle River, NJ : Prentice Hall PTR, 2002. ISBN 0-13-061621-4.