The architecture of the management system of complex steganographic information

O O Evsutin[1], **R V Meshcheryakov**[1], **A S Kozlova**[2], **T M Solovyev**[1]

[1] Tomsk State University of Control Systems and Radioelectronics, 40, Lenina Ave., Tomsk, 634050, Russia
[2] Tomsk State University, 36, Lenina Ave., Tomsk, 634050, Russia

E-mail: eoo@keva.tusur.ru

Abstract. The aim of the study is to create a wide area information system that allows one to control processes of generation, embedding, extraction, and detection of steganographic information. In this paper, the following problems are considered: the definition of the system scope and the development of its architecture. For creation of algorithmic maintenance of the system, classic methods of steganography are used to embed information. Methods of mathematical statistics and computational intelligence are used to identify the embedded information. The main result of the paper is the development of the architecture of the management system of complex steganographic information. The suggested architecture utilizes cloud technology in order to provide service using the web-service via the Internet. It is meant to provide streams of multimedia data processing that are streams with many sources of different types. The information system, built in accordance with the proposed architecture, will be used in the following areas: hidden transfer of documents protected by medical secrecy in telemedicine systems; copyright protection of online content in public networks; prevention of information leakage caused by insiders.

1. Introduction

At present time, the rapid increase of multimedia traffic proportion in data networks is caused by the high popularity of different kinds of audio and video interactive streaming applications that are used by the Internet users for video conferences, training, remote medical services, television broadcasting, etc.

So, there is a need for solving problems such as the prevention of confidential information leakage, assurance of impossibility of unauthorized multimedia content changes, copyright protection for multimedia content, medical confidentiality protection (for medical applications).

One of the most effective approaches for solving such problems is based on the steganography methods (which are the practice of transferring files, messages, images, or video within another file along with the concealing of the transmission fact) by inserting the hidden embeddings to the protected content for different purposes [1].

Steganography methods are classified as follows:

- Methods of hidden information embedding into digital objects to ensure its confidentiality during transmission and/or storage [2–5].

- Methods of digital watermarks embedded into the multimedia content to protect information from unauthorized copying and use [6–8]. Here, by the digital watermark we mean an invisible tag embedded in digital objects for confirming the authenticity of the protected objects as well as a conventional watermark.
- Methods of identification numbers embedded into the multimedia content to protect it from unauthorized copying and replication [9, 10]. Identification numbers can be considered as a kind of digital watermarks. The difference is that digital watermarks are not unique while an identification number is unique for each copy of the software or multimedia content. These methods often do not separate from each other.
- Methods of headers embedding. These methods are similar to the methods of the first type; they are not intended to provide privacy, but designed for hidden addition of auxiliary public information into the multimedia content.

Thus, a wide range of problems related to information security and digital content processing can be effectively solved by using steganographic methods.

The analysis of the US Patent Office Database [11] showed that the number of patented technical solutions relating to steganographic devices and methods is annually increasing. In particular, about 150 patents relating to the subject area were issued in 2014. World's leading IT companies such as Microsoft Corporation, Google Inc., Sony Corporation, etc. are engaged in such developments.

However, the analysis of the collected patent documents published between 2000 and 2015 and the literature review allowed us to establish the following fact: the existing solutions for multimedia data protection using digital steganographic methods are intended for private use and for the solution of individual information security problems and often designed for work with the content of a specific type (i.e. they do not have the complexity and flexibility).

2. Application of steganographic methods for scientific and practical problems

2.1. The main applications of steganography methods

The following principal scientific-practical problems, for which a management system of complex integrated steganographic information can be applied, are considered:

- copyright protection for video content, including materials of video conferencing and video-reportage recordings;
- video authenticity protection in surveillance systems;
- medical data protection;
- information leakage detection and prevention.

The ability of watching the same video on any number of mobile or stationary devices leads to the emergence of its multi-format multiple copies with various levels of quality. This different video content representation allows unauthorized copying of video, which leads to copyright infringement and inability to register a significant number of views legally. One way to confront the unauthorized video replication and transcoding in the film industry is the use of formats such as “Ultra Panavisio”, “IMAX” and others. The materials of video conferencing and video-reportage recordings are special cases of video content.

At present time, the most common applications that use video conferencing for business purposes are the following: online meetings and meetings between employees, webinars, online interviews in recruitment, marketing activities and online press conferences, presentation of services and products, video consultations, telemedicine. One of the major trends of modern market development is a consequence of the cloud computing spread, transferring the entire infrastructure to service providers and the use of videoconferencing as a service. The users of videoconferencing software application, including a large number of corporations and government organizations, are increasingly demanding copyright protection and confidentiality.

Steganography technology provides the ability of embedding the additional information about video, facts about its views and transcoding facts for manufacturers and rightholders of video conferencing recordings.

The processing and protection of the surveillance system data have its own specificity. Traditionally, video surveillance systems are used for the security of the protected site perimeter, but now there is a tendency to expand the functionality of these systems.

The intelligent video surveillance system allows accumulating data for making decisions, for example, the data about customers: their numbers, shop routs, etc. The video combined with the technological process data increases the quality control level at an assembly line.

Thereby, video surveillance data become a source of additional information for the effective production and business organization.

In this case, steganography can provide authenticity protection in the video surveillance system with automatic embedding of digital watermarks (DVRs, cameras, external and internal monitoring, etc.).

Telemedicine is another specific and rapidly developing area where the multimedia content presence is caused by the use of different systems of monitoring of a personal health status, registration and transfer of the data produced by embedded and carried medical devices, by the use of medical video conferencing and local or cloud medical information storage.

Such kind of multimedia content a priori has to be protected in respect of medical confidentiality and protection of personal information in the long run, often for life.

Today, the video recording of clinical and outpatient study and medical operations is introduced in many medical organizations. Normalization, unification of data, using specially designed formats, recording of medical attributes, data authentication (including the place and time) – these are the factors relevant today for the entire world medical community. So, in this case the problems of information protection and video content protection against modification occupy leading positions. Steganography applications are an effective instrument providing personalization and a possibility to proof tampering of original data. It should be noted that the steganographic methods are used not only by legitimate users and information owners to ensure its protection or to solve some auxiliary problems. Hackers also can use them for the organization of confidential information leakage through insider channels.

A significant part of the insider incidents is carried out with the aid of steganographic techniques when confidential information is taken away from company by means of the hidden embedding to multimedia with innocuous content. Steganalysis of such content type in the presence of the hidden embedded data embedding allows one to identify and prevent such incidents.

2.2. Existing systems for steganographic data protection and steganalysis

There are many systems for embedding hidden information into digital images, audio and video data. Here are the most well-known: Mp3Stego (for embedding information into MPEG-1(2) Layer 3 audio); ImageSpyer (for embedding data into images); S-Tools (allows one to place small files into the GIF and BMP images or WAV audio); MSU Stego Video (for hiding data in AVI files).

Complex systems of network traffic steganalysing in real-time are MBEGA – an optimized complex steganalysis method for searching stego images in the flow of network traffic, which was proposed by S. Geetha and N. Kamaraj (India) [12], and the StegAlyzerRTS real-time steganography scanner developed by SARC (Steganography Analysis and Research Center, Backbone Security.Com, Fairmont, USA) [13].

The algorithm implemented in the MBEGA system includes many different techniques for automated result obtaining. Each of these techniques is looking for the set of statistical image characteristics (about 140 characteristics: the study of the LSB correlation of statistical properties, cross-correlation of structure properties, Fridrich’s features, higher-order statistical features, etc.) and compares them to average network traffic patterns or statistics in an image database. The accelerated modification of the genetic algorithm is used for the search optimization.

StegAlyzerRTS is the world's first commercially available network traffic analysis tool for detecting of hidden embeddings in real time. The latest version of this application is able to operate in networks with the capacity of 1Gbit/s.

The StegAlyzerRTS algorithm is based on the comparison of potential stego container features with the file fingerprints from the database which contains multiple file profiles associated with 1225 steganografic and watermark applications. In addition, the algorithm allows retrieving hidden information from suspicious files and does not affect network performance.

Thus, all existing solutions are aimed either at embedding hidden data into files or at network traffic steganalysis.

According to the results of the literature review and a patent-governmental research, it can be stated that at the moment, there is no complex software solution for steganographic information management, which includes both the embedding/extracting of hidden information and the traffic steganalysis in local and global networks.

3. Description of the architecture of the distributed steganography management system

To solve the problems that we examined above, we offer to use a distributed management system of steganography information.

The system has a modular architecture and is oriented to solve a wide range of steganography problems including real-time processing. Apart from the features of existing embedding/extracting and analysing algorithms, during the system design, we paid much attention to scalability and the possibility of using it in distributed computing. It is connected to the huge multimedia data volumes to be processed and the high level of computing complexity of the existing steganography algorithms.

The scheme of modules interaction of the management system is presented in figure 1.

The media content transmit/receive module was designed for exchanging data between different multimedia sources. In particular, the ability to work in local networks, in the Internet, and also using different analog and digital television receivers is assumed. The components of the module which are serving for data transmitting/receiving could be divided logically or physically.

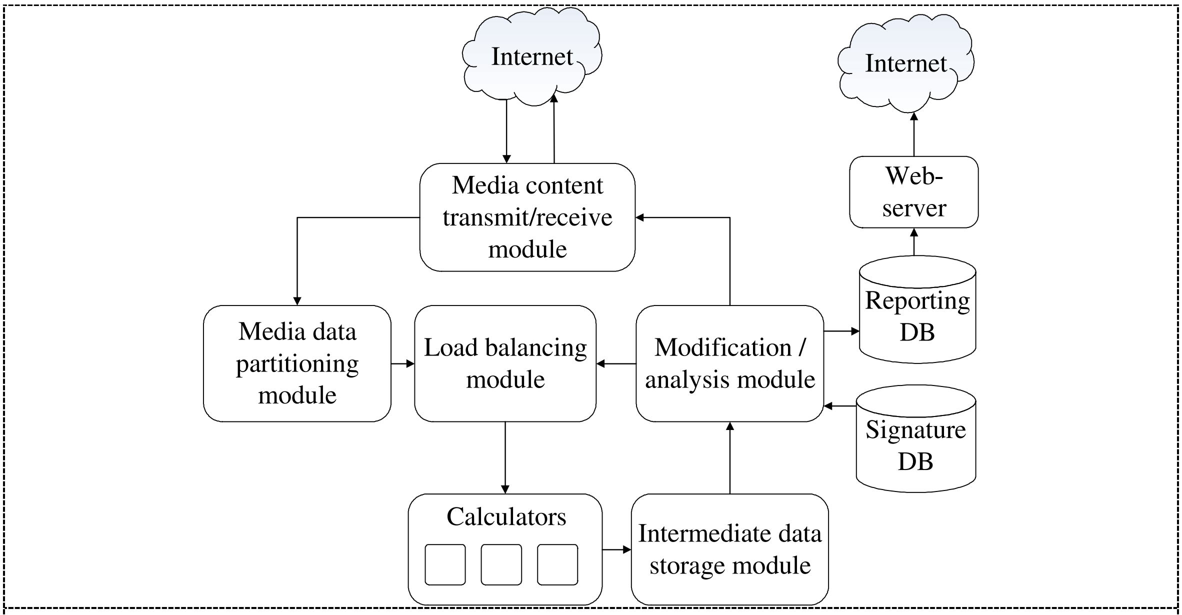
Internet
Internet
Web-server
Media content transmit/receive module
Reporting DB
Media data partitioning module
Load balancing module
Modification / analysis module
Signature DB
Calculators
Intermediate data storage module

Figure 1. The architecture of the proposed system.

The processing of the resulting content consists of two main steps: stream fragmentation into structural elements and modified data analysing. Structural elements imply the logically divided parts of a data stream which are specific for the specific media content type and its representative format.

Usually, media data are sent in the compressed mode. Consequently, the choice of structural elements is defined by the syntax feature of the compression standard. The structural elements for digital images could represent a palette, a block of pixels of a specific size, and some of specified pixels. In case of video content, it could be frames, slices, macroblocks, motion vectors, etc. Audio content could be fragmented into channels, frames, samples [14]. Apart from the media data format, the type of fragmentation into structure elements is defined by algorithms of modification/analysis implemented in the system. Within the developed system, the problem of the structural element selection of the analysing content is solved by the module of media data fragmentation.

The module that implements the management of stenographical information is the module of modification/analysis. The module implements a set of algorithms of embedding/extraction of stenographical data to work with trusted containers, and also detection algorithms of hidden attachments in untrusted containers. It should be noted that the described module works not directly with the streams of media data, but with its structural elements that were selected at the stage of fragmentation.

The algorithm of steganographic embedding detection may use previously formed special steganographic signatures during operation. The steganographic signature means the complex of characteristics (properties) or patterns (templates) associated with the media containers containing hidden embeddings. It is proposed to store these signatures in a specialized database for the efficient operation. Such kinds of databases are filled during the system creation and its operation in an automatic or semi-automatic way.

The result of the modification/analysis module (depending on the objectives) is either a multimedia container with/without extracted hidden embeddings or a report about presence/absence of hidden embeddings in the suspicious container.

In the first case, the data are sent to the transmit/receive module of media content. In other cases, the results are written in the dedicated reporting database. The report access can be obtained either from a local network or from the Internet through a web server interacting with the reporting database after the authentication and authorization procedures.

The algorithms implemented by the modification/analysis module and the media data partitioning module can be characterized by great computational complexity, and as a consequence, the proposed system architecture is focused on the use of distributed computing. Organization and control of computing resources are performed by the load balancing module. This module receives the tasks from the data partitioning module and the modification/analysis module in the form of four objects (S, A, D, M). S – the set of the input data identifiers, by which the input data can be received requesting either the intermediate data storage module or the media content transmit/receive module; A – the algorithm identifier to be executed; D – a set of output data identifiers, M – metadata including information about the algorithm complexity and the requirements to the type and speed of calculation.

Then, the load balancing module decides to delegate the task to the specific calculator depending on the system load, types of available calculators and the task. After the task is completed, the output data are sent either to the intermediate data storage module or directly to the module that set this task.

The intermediate data storage module provides optimal information distribution over the available devices in terms of access speed. At the same time, particular qualities of algorithms are taken into consideration for using this information in subsequent stages of system operation.

The proposed system can be deployed on a single server or on a cluster. Moreover, the system architecture is focused on cloud technologies as one of the most effective ways of scaling. The distributed character and the possibility of using heterogeneous calculators allow solving a wide class of steganography problems in real time.

4. Conclusion

Within the limits of the given paper, the literature review and patent research were conducted in the area of steganography methods and information protection. It was established that the currently existing scientific and technical decisions are intended for private use, aimed at solving only a limited range of information security problems and do not possess flexibility and scalability.

Therefore, the development and implementation of the information system which allows embedding hidden information into multimedia for different purposes without reference to the data type and data stream analysing for the presence of unauthorized embeddings is relevant. In this paper, the architecture of such system is introduced for the use in real time by means of the distributed computing. System algorithmization as regards to embedding of hidden information into multimedia content and extracting information has to be built on the basis of classical methods, their combinations and newly proposed approaches to stego container construction, which is based on multimedia data information sequences. In turn, the algorithms of determining the embeddings in multimedia content must be hybrid and combine mathematical statistics and computational intelligence.

5. Acknowledgments

This work was supported by the Ministry of Education and Science of the Russian Federation within the 1.3 federal program «Research and development in priority areas of scientific-technological complex of Russia for 2014-2020» (grant agreement № 14.577.21.0153 of November 28, 2014; identifier RFMEFI57714X0153).

References

[1] Fridrich J 2010 *Steganography in digital media: principles, algorithms, and applications* (Cambridge: Cambridge University Press)

[2] Shanableh T 2012 Data hiding in MPEG video files using multivariate regression and flexible macroblock ordering *IEEE T. Inf. Foren. Sec.* **7** 455–464

[3] Kanan H R and Nazeri B 2014 A novel image steganography scheme with high embedding capacity and tunable visual image quality based on a genetic algorithm *Expert Syst. Appl.* **41** 6123–6130

[4] Evsutin O O, Kokurina A S, Shelupanov A A and Shepelev I I 2015 An improved algorithm for data hiding in compressed digital images based on PM1 method *Computer Optics* **39** 572–581

[5] Evsutin O O, Kokurina A S, Mescheryakov R V and Shumskaya O O 2016 An adaptive algorithm for the steganographic embedding information into the discrete Fourier transform phase spectrum *Advances in Intelligent Systems and Computing* **451** ed A Abraham, S Kovalev, V Tarassov and V Snášel (Springer International Publishing) 47–56

[6] Glumov N I and Mitekin V A 2011 A new semi-fragile watermarking algorithm for image authentication and information hiding *Computer Optics* **35** 262–267

[7] Lusson F, Bailey K, Leeney M and Curran K 2013 A novel approach to digital watermarking, exploiting colour spaces *Signal Process.* **93** 1268–1294

[8] Ali M, Ahn C W and Siarry P 2014 Differential evolution algorithm for the selection of optimal scaling factors in image watermarking *Eng. Appl. Artif. Intel.* **31** 15–26

[9] Li Y N 2015 Robust content fingerprinting algorithm based on sparse coding *IEEE Signal Proc. Let.* **22** 1254–1258

[10] Thanh T M and Iwakiri M 2016 Fragile watermarking with permutation code for content-leakage in digital rights management system *Multimedia Systems* **22** 603–615

[11] United States patent and trademark office. URL: http://www.uspto.gov/

[12] Geetha S and Kamaraj N 2010 Optimized image steganalysis through feature selection using MBEGA *Int. J. of Computer Networks & Communications* **2** 161–175

[13] SARC – Steganography Analysis and Research Center: Home. URL: http://www.sarc-wv.com/

[14] Salomon D 2007 *Data compression: the complete reference* (London: Springer–Verlag)