IJCSE International Journal of Computer Science and Engineering Open Access
Technical Paper Volume-2, Issue-3 E-ISSN: 2347-2693

Data Leakage Detection Using Information Hiding Techniques

Ankit Tale[1], Mayuresh Gunjal[2*] and B.A. Ahire[3]

[1]*Departmet of Computer Engineering, Pune University, India, taleankit@yahoo.in*
[2] *Department of Computer Engineering, Pune University, India, mayur157@yahoo.co.in*
[3]*Assistant Professor, Department of Computer Engineering, Pune University, India,bhawanaahire@yahoo.com*

www.ijcseonline.org

Received: 5 March 2014 Revised: 14 March 2014 Accepted: 26 March 2014 Published: 31 March 2014

Abstract— This paper is detailed about the proliferation of media sharing through third party vendors and tracking of guilt agent using a key generation and encryption of key in media file using asymmetric key cryptography techniques. Main focus of this paper is on the security provided for proliferation of media which help the admin to track person for leakage. In the process of each key generation is based on timestamp of system, LSB replacement along with encryption of key using RSA algorithm. Stegnalysis is the art of detecting the message's existence and blockading the covert communication. The Least Significant Bit steganography is a technique in which least significant bit of the image is replaced with data bit.

Keywords/Index Term—Assymetric Key Cryptograhy,Steganalysis,Least Significant Bit,Watermarking.

I. Introduction

The advancement of Internet services and various storage technologies leads to digital revolution and this made significant increase in media piracy while sharing with third party or vendor. For example, a company outsources its data processing because of insufficient manpower, so the requirement data must pass to another company. Then the data is pass to distributor or the agents. Our goal is to detect when the distributor's sensitive data have been leaked by agents, and if possible to identify the guilt person that leaked the data.

Motivation:

- Data leakage is being a serious problem in this century it mainly affect the company's profiles and strategy.

Goals/Objective:

- To detect the guilt agent for leakage and decide whether the person should be trusted or not.

Contributions:

- In this work we replace Least Significant Bit of media without distortion of file.
- Encryption through asymmetric key algorithm provides confidentiality and integrity.

Outline of Paper:

Section 1, Introduction to Project Work.
Section 2, Mainly focused on related work.
Section 3, Need to Study Problem
Section 4, Methodology

II. Related Works

Watermark can be either inserted directly or integrated during process or implemented after data compression of video file is done. Now we shall briefly discuss some common video watermarking techniques.

Corresponding Author: *Mr. Mayuresh Gunjal*

2.1 Stegnography Using Least Signifcant Bit Algorithm

Stegnography Using Least Signifcant Bit Algorithm proposed by Mrs. Kavitha, Kavita Kadam, Ashwini Koshti, Priya Dunghav[9]. In this technique image using least significant steganographic algorithm is used for hiding data and to send the stego file to the user where the retrieving of the secret data is done. The hidden secret key/data into an image which acts as a file carrier are used to send without any modification of original data. If any changes happen in the image while inserting the secret message into the image, there are chances of data access by unauthorized person and try to modify the data. So, the data encryption into an image and decryption and steganography plays an important role for detection.

2.2 Watermarking using CDMA modulation

Watermarking using CDMA modulation was proposed by B. G. Mobasseri [7]. In this methodology one of the four LSB-planes are replaced by watermark planes. The bit-planes that were to be replaced are selected using a random periodic quaternary sequence. The watermark plane is generated using spread spectrum methodology. For detection of the watermark, the author has presented a two-level hierarchical correlation methodology. One of the prime motivations for watermark integrating into video coding structures such as MPEG-2, H.264 etc for reducing overall time complexity.

2.3 Watermarking based on region based energy modification.

Darmstaedter proposed a method for data hiding, where embedded data are manipulating for average energy or luminance intensities in sub-regions of each frame [4]. This method achieves a high embedding data capacity where one bit into every 8×8 block, and error control coding for ensuring robustness. Here the data sequence *U* is embed directly to cover data. The concept of block classification

© 2014, IJCSE All Rights Reserved

was introduced by author for eaiser water markings. With the classification of blocks, this scheme can take the advantage of local spatial characteristics and adjust its embedding strategy to improve imperceptibility and robustness criteria.

2.4 Least significant bit modification (LSB)

Shailender Gupta, Ankur Goyal and Bharat Bhushan proposed technique which is simple and straight-forward and uses the least significant bits to embed the watermark [2]. This method provides high capacity which can be used to embed the watermark frequently in a cover media. This technique is resistant against cropping. An approach to enhance the robustness is to applying a pseudo random generator to determine the LSB bits to modify. This technique can improve the security and prevent the third party from tracing the watermark.

Table 1 .Video Watermarking Classification

Host media		Text, Image, Audio, Video	
Visibility of Watermark		Visible, Invisible	
Robustness of water marking		Robust, Semi-fragile, Fragile	
Watermark data types		Noise, Authentication information, Image	
Embedding method	Spatial domain	LSB, Image checksum, Random function	
	Frequency domain	Look-up table	
	Compression domain	Spread Spectrum	DCT, Wavelet(DWT), Fourier(DFT)
Detection		Blind,Non-Blind, Semi-Blind	

III. Solution/Need/Importance of the study Problem Statement/Objectives

Media proliferation is a serious issue for companies as continuous leak of an important data affects the strategy and working mentality of company and man month. To stop such data leakage and catch the guilty agent we proposed this methodology for detection.

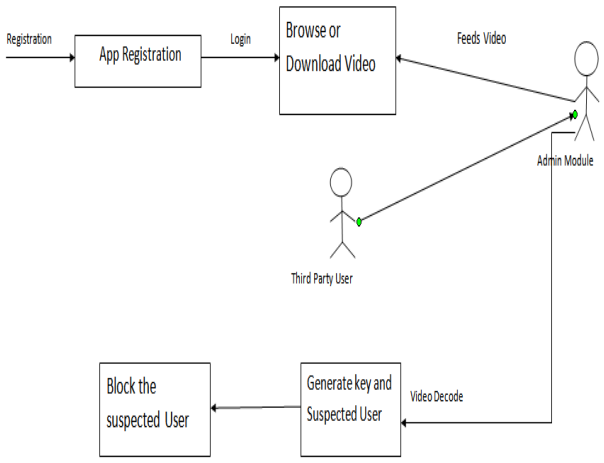
IV. Methodology

Traditionally, leakage detection is handled by an old technique called as water-marking, where a unique code is embedded in each copy that is distributed across. If that copy is later found in the hands of an unauthorized person, the leakage can be identified. Watermarks can be very useful in some cases, but again, needed modification of the data that document. Furthermore, watermarks can be destroyed if the data recipient being a malicious ones.

We proposed system works in same manner with existing system with an exception of encryption and embedding dynamically generated code in video. We had developed techniques for leakage detection for distributed video file. The system empowers admin to track the source user who has leaked the video and take appropriate action upon him. We used steganalysis, which is called as the art of is the discovering the existence of hidden information from data. Our proposed system will hide the data using LSB encoding which will only replaces the least significant bit of video file that are easy to replace without any effect. The purpose of proposed system is to detect when the video data have been leak by person and to identify the person that leads to leakage of important day.

We also develop a unique key generation algorithm based on system timestamp which invokes the appropriate set of consonant and vowels for generating key which will be embedded into the video file. Also we will encrypt the key using the well-known RSA algorithm before embedding it into the video file.

Fig a. System Architecture



4.1 Proposed steganography mechanism at transmission side.

At the transmission side the key generated using our key generation algorithm will be encrypted using RSA algorithm and the ASCII notation of the encrypted key will be converted to binary. Simultaneously the pixels values of the video file will be converted to binary. Both these binary values will be given as input to the LSB encoder which will replace the LSB values of the pixels with the binary value of the encrypted key. Finally the binary values of the pixels will then be converted to pixels and the stego video file will be generated which will be transmitted to the client for viewing. Here stego video file refers to the video file generated by the system after embedding the uniquely generated key.

4.2 Proposed stegnography mechanism at receiver side.

At the admin side, if the distributor finds a copy of the video file on the internet, or on somebody's laptop, the admin will give this file as an input to the system. The pixels of the video file will be converted to binary values. These

© 2014, IJCSE All Rights Reserved

binary values will be forwarded to the LSB decoder which will extract out the key embedded into the video file. The extracted key can then be matched with the keys stored in the database to find out guilty agent.

Fig b. Stegnography mechanism at transmission side

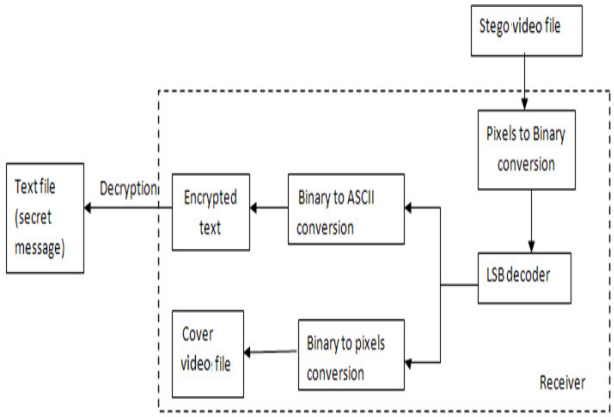
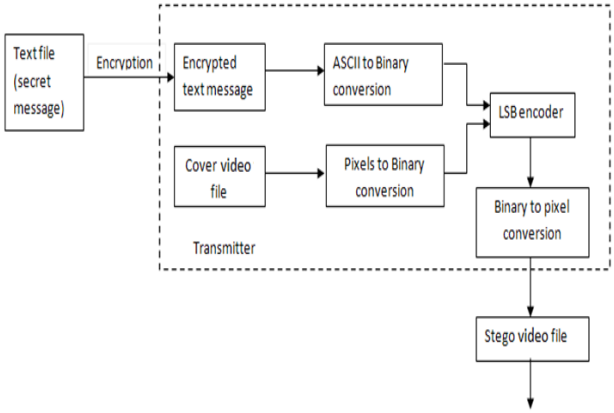


Fig c. Stegnography mechanism at receiver side



4.3 Key Generation Algorithm

The key generation algorithm works with the help of the time-stamp of the system. It works in the following steps.

1. Initialize the set of vowels and consonants.
2. Decide the length of key to be generated and strength of key.
3. Based on Strength alter set of vowels and consonants to include additional characters.
4. Retrieve current timestamp in seconds and check whether integer is even or odd.
5. If time is even generate key from vowels set else generate key from consonants.
6. In each process calculate random location of character from character set and retrieve each character.

4.4 Key Encryption Using RSA

The key generated using the key generation algorithm will be encrypted using RSA algorithm and single bit LSB replacement will be done in order to embed the key into the video file. The following points may be noted:

- RSA is a block cipher in which the plaintext and cipher text are co-prime integer between 0 to n-1 for some n.
- In the RSA algorithm each station independently and randomly choose two large co-prime number.
- As the number of pixels in image increases, the number of instruction (complexity) at the sender and receiver side increases.
- There is an increase of 10% to 66% in the number of instruction execution in comparison to one bit steganography for pure steganography combined with RSA algorithm.
- The complexity is higher for steganography combined with RSA algorithm in comparison to pure and steganography.
- The encryption will be 1024 or 2048 bit encryption.

V. Conclusion

We need to work with agents that may not be trusted, and we may not be certain if a leaked object came from an agent or from some other source. Watermarks usually affect the quality of the video files. In spite of these difficulties, we have shown that it is possible to assess the likelihood that an agent is responsible for a leak, and how single bit LSB replacement least affects the quality of the video file. Also we have shown that the key embedded into the video file when encrypted using RSA increases the security of the video file. Thus, we have developed the system which will help in reducing the piracy of copy righted video files. This system may also be further extended to prevent piracy of audio files.

VI. Scope for Further Research

We propose this work for video files this proposed system can further be used with some modification for picture, mp3 files, and documents. We can use these methods at various pay sites for user and also useful to track unwanted user who depends on picture from another profile for misuse.

References

[1] Papadimitriou P and Garcia-Molina, "Data Leakage Detection" Knowledge and Data Engineering, IEEE Transactions on Volume: 23, Issue: 1, Page No (51-63), Jan 2011.

[2] Kumar Ajay, Goyal Ankit, Kumar Ashwani, Chaudhary Navneet Kumar and Sowmya Kamath, "Comparative evaluation of algorithms for effective data leakage detection", Information & Communication Technologies (ICT), 2013 IEEE Conference, ISBN 978-1-4673-5759-3,Page No(177-182),April 11-12,2013.

[3] Shailender Gupta, Ankur Goyal and Bharat Bhushan "Information Hiding Using Least Significant Bit Stegnography and Cryptography" Vol: 4, No: 6, Page No (27-34), June 2012.

© 2014, IJCSE All Rights Reserved

[4] V. Darmstaedter, J. -F. Delaigle, D. Nicholson and B. Macq, "A block based watermarking technique for MPEG-2 signals: Optimization and validation on real digital TV distribution links", Proceedings 3rd European Conference on Multimedia Application, Services and Technique, ISBN 978-3-540-64594-8,Page No (190-206),May 1998.

[5] J. Lee and Sung-Hwan Jung, "A survey of watermarking techniques applied to multimedia" IEEE International Symposium on Industrial Electronics, Volume: 1, Page No (272-277), June 2001.

[6] Hamid Shojanazeri, Wan Azizun Wan Adnan and Sharifah Mumtadzah Syed Ahmad, "Video Watremarking Technique for Copyright Protection and Content Authentication", International Journal of Computer Information Systems and Industrial Management Application, ISSN 2150-7988, Volume: 5,Page No(652-660),2013.

[7] B. G. Mobasseri, "Exploring CDMA for watermarking of digital video" Proceedings of the SPIE, Volume: 3675, Page No (96-102), April 1999.

[8] F. Hartung, J. K. Su, and B. Girod, "Spread spectrum watermarking: malicious attacks and counter-attacks," from Conference Volume:3657 Proceeding SPIE Security and Watermarking of Multimedia Contents, San Jose, CA, Page No (147–158), Jan. 1999.

[9] Mrs. Kavitha, Kavita Kadam, Ashwini Koshti, Priya Dunghav "Stegnography Using Least Significant Bit Algorithm", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Volume:2, Issue:3, Page No (338-341),May-Jun 2012.

© 2014, IJCSE All Rights Reserved