ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015.

Data Leakage Detection Using LSB

Sanchit S. Mhatre[1], Vaibhav V. Kakhandaki [2], Bhagyashri P. Yeola[3], Rakesh Badodekar[4]

U.G Student, Information Technology, Sinhgad Institute of Technology, Lonavala, India [1, 2, 3]

Professor, Information Technology, Sinhgad Institute of Technology, Lonavala, India[4]

Abstract: An administrator distributes sensitive data to some trusted agents (third parties) and if one or more of them will leak that sensitive data and found in an unauthorized place (e.g. on the web or somebody's laptop),then we can catch & prosecute them as per Law. The question may arise in mind, how to deal with such problem? There is a unique key embedded in each copy of the data by distributor before distributing that data to agents and this key remains within data. In some cases, we can also inject "realistic but fake" data records to further improve our chances of detecting leakage and identifying the guilty party. We are using List Significant bit (LSB) [1] steganography concept here to embed the key in the data. So when the distributor finds his secret data at unauthorized place, he will simply take that data and check for the unique key. If the key in that data matches with the key present in his database then the system will display the name of that guilty agent and then he can mark it as invalid user [2].

Keywords: Steganography, LSB [Least Significant Bit], FPS, Watermarks, Message key, Serialized.

INTRODUCTION

I. This paper is based on detecting the guilty agents who leaked the data. If we take an example of any business enterprise, sometimes sensitive data must be handed over

II. Supposedly trusted third parties. Our goal is to detect the sensitive data that has been leaked by guilty agent. Here we are using Steganography for Video category. Hiding information in text is historically the most important method of steganography. There are several steganography techniques used for hiding data such as batch steganography, permutation steganography, least significant bits (LSB), bit-plane complexity segmentation (BPCS) and chaos based spread spectrum image steganography (CSSIS) [3].

III. The most common and popular method of modern day steganography is to make use of the LSB of a picture's pixel information [3]. Thus the overall image distortion is kept to a minimum while the message is spaced out over the pixels in the images. This technique works best when the image file is larger than the message file and if the image is in grayscale [4].

In our system, administrator will keep record of agents along with the key sent with them and decodes the data found on other site and get key. He compares that key with his own records stored in database, to match the key sent along with the agents.

System can detect the leakage of data thereby keeping the trusty workers with them. We are using MySQL server for storing purpose and apache server for connectivity which are in built servers in XAMPP software [4].

The operation of the data leakage detection is shown below in the following figure,

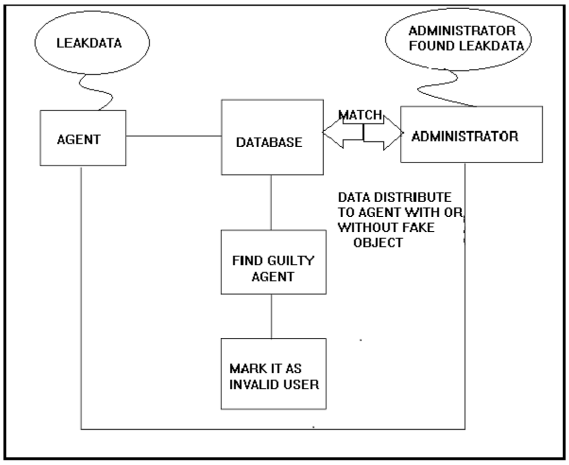
LEAKDATA
ADMINISTRATOR FOUND LEAKDATA
AGENT
DATABASE
MATCH
ADMINISTRATOR
DATA DISTRIBUTE TO AGENT WITH OR WITHOUT FAKE OBJECT
FIND GUILTY AGENT
MARK IT AS INVALID USER

FIG. 1 OPERATION OF DATA LEAKAGE DETECTION

An administrator distributes sensitive data to their users. If user leaked some of the data and administrator found that specific data at unauthorized place, then he retrieves that data from the unauthorized place and matches data with his hidden key which is stored in the administrator's database "If the key matches", then we can say that the person who leaked the data is guilty else the person is trustworthy. We also add some fake data (Fake Records) and admin distributes to the trusted party circle to find the loyalty with company or system.

STEGANOGRAPHY

Steganography is the art and science of invisible communication. This is accomplished through hiding information in other information, thus hiding the existence of the communicated data. The word steganography is derived from the Ancient Greek language "*stegos*" meaning "cover" and "*grafia*" meaning "writing" defining it as covered writing" [3]. It is in contrast to cryptography, where the survival of the message itself is not masked, but the actual content is hidden. Steganography is implemented in different fields such as "Military" and

DOI 10.17148/IJARCCE.2015.43104

"Industrial applications". By using lossless steganographic techniques valuable data can be sent and received securely. Traditionally, steganography was based on hiding secret information in text & image files .Lately, there has been growing interest in implementing steganographic techniques to video files as well as audio files. The advantage of using video files in hiding information is to be adding security against hacks due to the relative complexity of video because of frames per second; compared to image files and audio files. There are several steganography techniques used for hiding data such as batch steganography, permutation steganography, least significant bits (LSB), bit-plane complexity segmentation (BPCS) and chaos based spread spectrum image steganography (CSSIS) [1].
Figure 2 shows the four main categories of file formats that can be used for steganography.

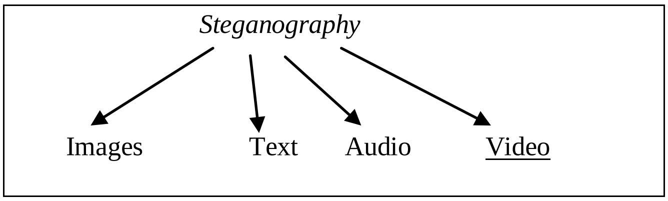

Fig.2 Categories of Steganography

LEAST SIGNIFICANT BIT METHOD

This approach is very simple. In this method the least significant bits of seventy percent of the bytes inside an image or a frame considering video can replaced with a bits of the secret message (key) [3]. The proposed algorithm is replacing one LSB of selected pixel bytes in video frame it becomes very difficult for intruder to guess that a key is hidden in the video as individual frame are difficult to analyze in a video running at 30 frames per second and we substitute one LSB key [4].

EXISTING SYSTEMS

- Watermarks involve some modification of the original data and also watermarks can destroy if the data recipient is malicious. We use watermarks mostly in paper currencies of various countries [4].
- Present day transactions are considered to be "un-trusted" in terms of security, i.e. they are relatively easy to be hacked and also we have to consider the transfer of large amount of data through the network will give errors while transferring. Only single level of security is present in the existing systems [5].
- The other problem of existing system is now a days hacking activities are growing day by day & hackers can easily hack important information and security is not sufficient to stop them.

Though security status increased at a high level but the major drawback of new status of security is "cost". Hence we need better solutions which have good security level with lower cost that is what market actually needs a suitable solution with less strain for pockets.

PROPOSED SYSTEM

- The used security techniques are not appropriate to prevent hacking and the new security technique is high of cost. Then we need a different technique which is more efficient and provides a better security level.
- The administrator is serialized with each and every database or temporary file created.
- In the proposed paper, for each video file multiple users can view or download that file. While downloading that file from the system, admin generates a key for that user who is downloading that file using that user's metadata. After that a random number of frames are picked up for ciphering.
- Only 70% of the video file can be used for encrypting because 25% to 30% of the file can get changed or altered during transfer process.
- Coming back to message key, we can use one to ten bytes of the key for encoding. Care is taken that the message key bits are directly proportional to the number of bytes selected for ciphering from the video file.
- The padding process is the next one. One bit of key for 1 byte of actual data is used with padding bits for the "key" and then the "EX-OR" operation is performed and by this process each bit of "key" is encrypted with selected number of bytes.
- During the detection process, the video that has been leaked by any one of the users is downloaded by the administrator and the de-ciphering process is carried out using exact opposite operation during ciphering. The each user's metadata is gathered in a temporary file and cross-referenced by deciphered key which gives information about the particular user and thus the guilty one is found and blocked by the administrator.

WORKING

System Flow

1. Login by distributor
2. Sign up for new agent
3. Details of agents are stored in database
4. Distributor distributes data
5. Unique "code" + "username" in database
6. Agent transfer data to unauthorized person
7. Finds sensitive data at unauthorized place
8. Check the unique code
9. Find guilty user.

DESIGN ARCHITECTURE

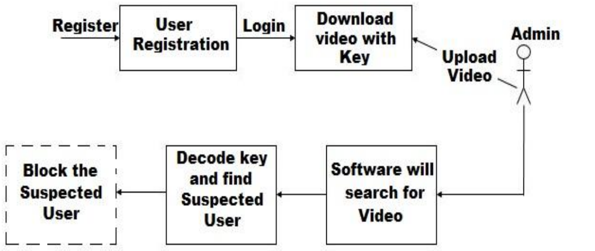

Fig.3 Design architecture

Algorithm used

Least significant bit (LSB) is the best method for data protection [5]. LSB method is very simple and a

DOI 10.17148/IJARCCE.2015.43104

commonly used approach for developing Steganography system because the amount of space that an image can provide for hiding data will be more comparing with other method [6]. This method is probably the easiest way of hiding information in an image and yet it is surprisingly effective.

Algorithm used for hiding

Each pixel of size 8 bits is hidden in video frame of 8 pixels. If image size is a_1*b_1 and frame size if a_2*b_2.Then number of pixels in one row of 1 frame that can be hidden are given by Q=n*2/8 pixels, Number of frames that can be hidden in a video are given by [5] [8]

```
Step 1. P= ($b_1/b_2$)*8
Step 2. For i=1 to p
Step 3. For j=1 to a1
Step 4. For k=1 to q
Step 5. Write bits in LSB if frame pixel (8 pixel required)
Step 6. End for.
Step 7. End for.
Step 8. End for.
```

Algorithm used for Un-hiding

To unhide the image, LSB of each pixel in the frame is fetched and a bit stream is constructed to construct the image [5] [8].

```
Step 1. For i=1 to p
Step 2. For j=1to a1
Step 3. For k=1 to q.
Step 4. Read pixel.
Step 5. Find LSB.
Step 6. End For.
Step 7. Construct bit stream to be written in recovered image.
Step 8. End For.
```

WORK CARRIED OUT

All software are installed and kept updated. The important thing regarding the system is that there will be only administrator and many users of the system. The registration for the admin is the first priority of the system inscribed as per the paper. The registration process must be carried out with due care. And each of the fields for registration is validated. Now comes the user in play.

The registration part is considered first. The validation process for each field of the user is also considered during the process. The registered user and registered admin's information is kept or saved in the database in their respective fields.

Administrator has the responsibility for uploading the video file, monitoring the system, track of records, key generation, padding bits generation, hiding and un-hiding procedure at backend & status of the users such as "block" or "unblock".

User has the active fields such as viewing his profile, changing his credentials, downloading files that has been uploaded by the admin and his transactions are saved & kept track-off by the admin.

Validation process and the remaining things about the validation process is carried out by the administrator himself. The key hiding procedure which has been explained in the proposed system of this paper is carried quite efficiently. And the "Key hiding process" is carried out before the file is downloaded by the particular user using his respective metadata.

The last procedure of the unhiding process is still undergoing the construction phase of the system. But it has been also explained in the proposed system terms.

CONCLUSION

Our work is in progress towards making "Data Leakage Detection" application. So that we can recognize the person who has leaked the sensitive information of the particular firm.

In the perfect world, no need to hand over sensitive data to agents that may unknowingly or maliciously leak it.

If we had to hand over sensitive data, we need to protect that data.

The goal of the paper is being achieved as per our efforts regarding the system requirements and its execution flow. The last step of our paper would be achieved in short term of period.

Our future idea includes the investigation of user guilt models that capture leakage scenarios.

REFERENCES

[1] Rosziati Ibrahim and Teoh Suk Kuan: Steganography Algorithm to Hide Secret Message inside an Image Computer Technology and Application 2 (2011) 102-108

[2] e.en.softonic.com/app/xampp

[3] Moerland, T., "Steganography and Steganalysis", Leiden Institute of Advanced Computing Science,www.liacs.nl/home/ tmoerl/privtech.pdf

[4] Rosziati Ibrahim and Teoh Suk Kuan: "Steganography Algorithm to Hide Secret Message inside an Image", 102-108, February 25, 2011.

[5] Shailender Gupta, Ankur Goyal, Bharat Bhushan, "Information Hiding Using Least Significant Bit Steganography and Cryptography", I.J.Modern Education and Computer Science, 2012, 6, 27-34.

[6] Hemant Gupta, Dr. Setu Chaturvedi 1,2,Technocrats Institute of technology Bhopal, "Video Data Hiding Through LSB Substitution Technique", International Journal Of Engineering And Science Vol.2, Issue 10 (April 2013), Pp 32-39.

[7] Hemant Gupta,Dr. Setu Chaturvedi, Technocrats Institute of Technology Bhopal(M.P.), "Video Steganography through LSB Based Hybrid Approach", Volume 6, Issue 12 (May 2013), PP. 32-42

[8] Mohammed A.F. Al Husainy, "Image Steganography by mapping pixel to letters" in Journal of computer science 5(1),33 -38, 2009

[9] S. Suma Christal Mary M.E (Ph.D) Lecturer Department of CSE PSN College of Engg & Technology, Tamilnadu, India, " IMPROVED PROTECTION IN VIDEO STEGANOGRAPHY USED COMPRESSED VIDEO BITSTREAMS", Vol. 02, No. 03, 2010, 764-766