International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 04 | Apr -2017 www.irjet.net p-ISSN: 2395-0072

Privacy-Preserving Technique for Insider Collusion Attack On Text And Image File Based On Fake Data

Ms. Nilima V. Kayarkar [1], Prof. Ms. Gangotri Nathaney[2]

[1] *M.Tech scholar, CSE Dept., WCEM, Nagpur, India*
[2] *Assistant Professor, CSE, Dept., WCEM, Nagpur, India*

Abstract - Data leakage happens whenever a system that is designed to be closed to an eavesdropper reveals some information to unauthorized parties. *We know that for business purpose, it is necessary to transfer important data among many business partner and between the numbers of employees. But during this transfer of data, information is reach to unauthorized person. So it is very challenging and necessary to find leakage and guilty person responsible for information leakage. In this system we find the person which is responsible for the leakage of text as well as image file. For this we used distributor and agent. Distributor means owner of data and agents means trusted parties to whom we send data. This system finds the insider collusion attack. An insider attack is a malicious threat to an organization that come from people within the organization such as employees, contractor or business associate, who have inside information concerning the organization's security practices, data and computer system. The main aim of this system is to find data of owner which is leaked and detect agent who leaked data. Here for text data we used kernel based algorithm and for image file we used steganography concept.*

Key Words: Data leakage, insider attack, distributor, agent, reveals, malicious threat, steganography.

1. INTRODUCTION

The research shows that 33 percent of information security attacks originate from internal employee means from the insider. Here insider means the person working in the organization. Now, problem of data breaching is one of the rapidly increase type of attack. Data breach is an incident in which sensitive, protected or confidential data has potentially been viewed, stolen or used by an individual unauthorized person. In proposed system we find the attacker who carried out attack on text data file as well as image file. Here we take excel and access file for text data and jpeg format for image file. To find guilty person here we used fake object, means distributor add fake data in the original data before distributed to agent. According to this fake data we find guilty person. In text file we generate fake data by using automatic fake data generator. To generate fake data we used kernel based algorithm and for image file we used steganography concept. Here we add fake data behind the image therefore agent cannot see the fake data they only see the image.

In this system we used two term

1. **Distributor:** It is owner of data who send secret information to agent.
2. **Agent:** It is the members within the organization, means it is insider and is semi-trusted. They may leak their own data to the outsiders.

In this system, we design the website where distributor and the authorized agent will log in. For new registration one has to fill the register form, and after successfully submitting the data he can logged into our system. When distributor wants to send file they add fake data according to type of file and then send file to agent.

2. LITERATURE SURVEY

In [1] authors propose an insider collusion attack that carried out on data mining systems. It explains how many insiders are sufficient to do this attack. In this system insiders means person within organizations collude with outsiders. This paper introduced many proposed privacy-preserving schemes to counter the attack.

In [2], this paper cloud computing services provide resource for organizations to improve business efficiency but also expose new possibilities for insider attacks.

In [3], it enhanced LSB technique which helps to successfully hide secret data into the cover image with minimum distortion. In this system there is no loss of original information. It is faster and reliable.

In [4], the main aim of the system is to develop a steganographic application that provides good security and reliability. The proposed system uses LSB steganography concept to provide higher security and also protects the message from stego attacks.

3. PROPOSED SYSTEM

In this system model is developed for finding the guilty agents. This system is used to find out attack carried out on text data as well as image file. In this system we used two functions.

a) **Distribution of file:** The distribution of file is shown in following flow diagram. While distributing file to agent

distributer can add fake data in the original file.

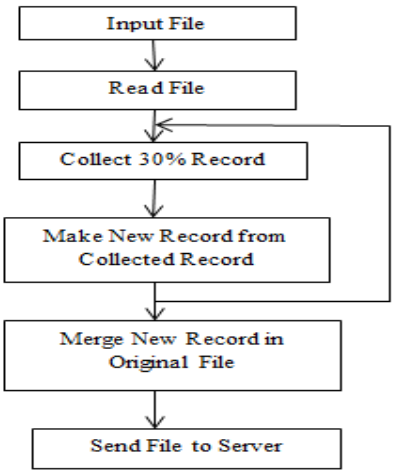

Fig 3.1: Flow diagram of distribution of file

b) Detection of leak file: The detection of leak file is shown by following flow diagram. Here we give leak file as input. The system read file and match fake record with file rows. It finds percentage of fake record and according to that it decides guilty person responsible for the leakage of text file.

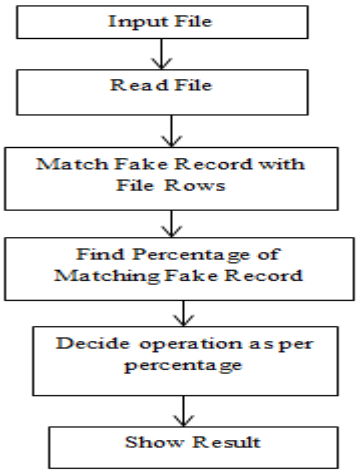

Fig 3.2: Flow diagram of detection of file

4. METHODOLOGY

We implement this system in two modules.

a) Kernel-based data mining method: We design model which find guilty person responsible for the leakage of text data. Means we used excel as well access file for demonstration. We add fake objects using kernel-based data mining method while distributing objects to agents. We inject realistic but fake data records to further improve our chances of detecting leakage and identifying the guilty person. Here guilty person is insider. While distributing file to agent distributer can add fake data in the file. We generate fake data by using automatic fake data generator. For the generation of fake data we used kernel-based data mining algorithm. According to this algorithm, it scans the original file and divides it according to how many percentages of fake data is added. Then it generates fake data, add it in the original file and send file to the agent. For example suppose we have excel file which contain name and phone number of 10 student .If we decide to add 30% fake data then according to this algorithm it divide original file into three part and generate fake data. Means it takes first name of one student and combine it with last name of another student and generate new fake name. Similarly it generates other entries.

b) LSB steganography algorithm: The term steganography is derived from Greek and its meaning is 'covered writing'. This algorithm is used to hide the fake data behind the image. Steganography is of the techniques through which existence of the message can be kept secret. We used image steganography that make use of Least Significant Bit (LSB) algorithm for hiding the data into image. Now a day image is most popular cover object used for steganography. Steganography system consists of three elements.

i) Cover image which hide secret message.
ii) Secret massage
iii) Stegano-image which is cover image with message embedded inside it.

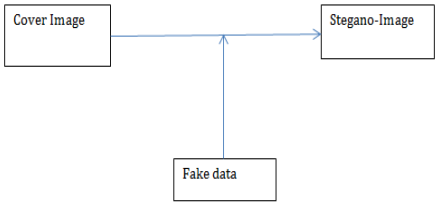

Fig 4b.1: Steganography concept

In this system secret message is our fake data which we add behind the image. For every transfer of file, new fake data is generated. Means fake data is unique for every transfer. Here we add date and time as a fake data. We take time in millisecond to maintain its uniqueness.

5. EXPERIMENTAL RESULT

a) Experimental result of excel and access file:

Here we used excel as well as access file for the demonstration. Because in many organizations excel or access file is used for storing the data.

Fig 5a.1: Home Page Window

In this window, we have admin login, distributer login, agent sign in and agent registration. According to requirement we make login. First we make distributor and the make login. Inside distributor we make number of agent.

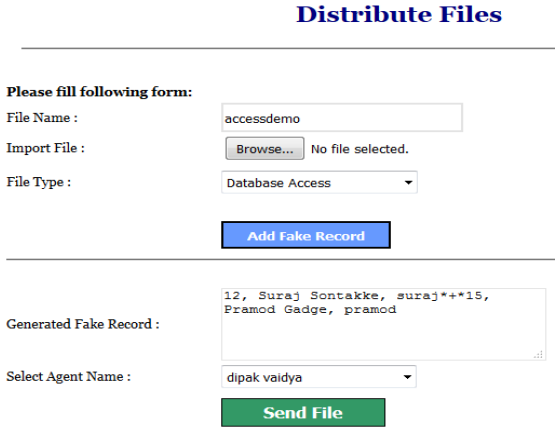

Fig 5b.2: Send Window

In this window distributor send file to agent. Here distributor select file name, type of file and then add fake data. Finally distributor select agent name to which file is to send and send file. Here to generate fake data we used automatic fake data generator.

Fig 5a.3: Record Detection Window

In this window we have to give file name which is leak and it gives us the name of agent who is responsible for leakage of file.

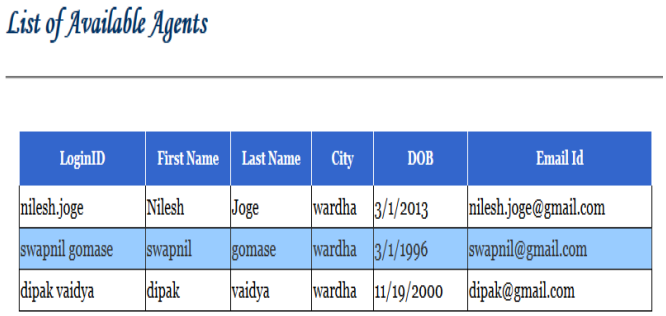

Fig 5a.4: window shows list of agent

This window shows list of all available agents.

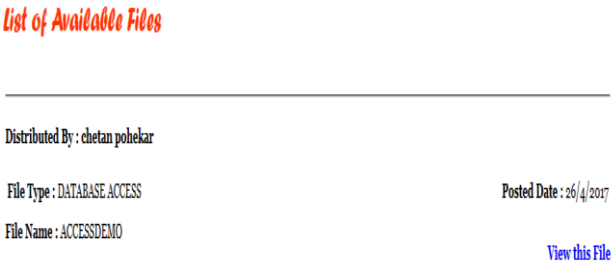

Fig 5a.5: Available file window

This window shows list of available file in agent. From this window agent can view the content of file. Here file contain fake record.

b) Experimental result of image file:

Here we used jpeg format image file for demonstration.

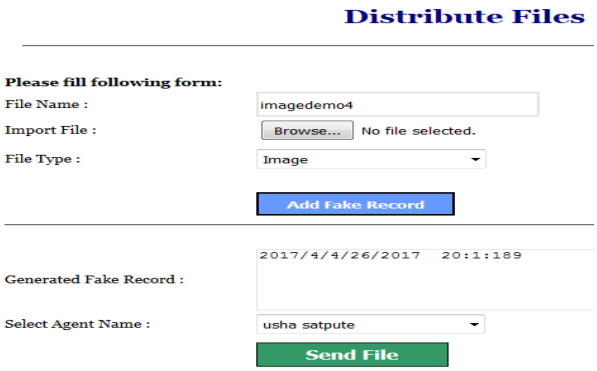

Fig 5b.1: send file window

Here distributor add fake data in the image.We used date and time as a fake data. Here time is in milisecond

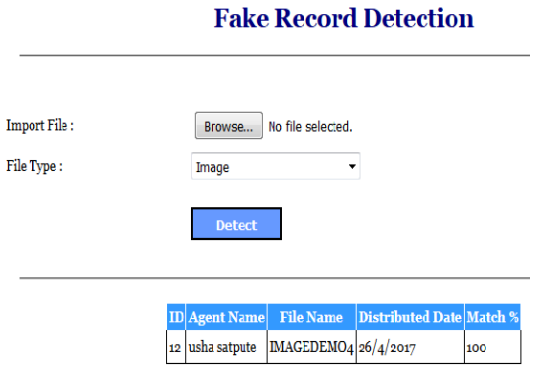

Fig 5b.2:Record detection window

Here it gives the name of agent which is responsible for leakage of image file.

Fig 5b.3: image before sending

This is the image which distributor want to send to agent.

Fig 5b.4: image after applying steganography

This image contain fake data. Here we used date and time as a fake data.

6. CONCLUSION

Everyday data leakage happens when confidential business information is leaked out. It is not certain that leaked data came from agent or any other means. The propose system can finds the guilty person responsible for the attack carried out on text data file as well as image file. Using this system we can increase the security and also find the guilty person who is responsible for attack. For image file we used LSB steganography method. The main aim of the project is to develop a steganographic application that provides good security. The proposed approach provides higher security and can protect the message from attacks. Here the image resolution doesn't change much and it is negligible when we embed the message into the image. We are using the Least Significant Bit algorithm in this project for developing the application which is faster, reliable and secured.

7. ACKNOWLEDGMENT

Initially, we would like to thank our almighty in the success of completing this work. We would extend our gratitude to all the experts for their critical comments.

8. REFERENCES

[1] Peter Shaojui Wang, Feipei Lai, Hsu-Chun Hsiao, And Ja-Ling Wu, "Insider Collusion Attack on Privacy-Preserving Kernel-Based Data Mining Systems", *IEEE Trans,* Apr.2016.

[2] W. R. Claycomb and A. Nicoll, ``Insider threats to cloud computing: Directions for new research challenges,'' in Proc. IEEE 36th Annu. Comput. Softw. Appl. Conf. (COMPSAC), Jul. 2012, pp. 387_394.

[3] A. E.Mustafa A.M.F.ElGamal M.E.ElAlmi Ahmed. BD"A Proposed Algorithm for Steganography in Digital Image Based on Least Significant Bit"

[4] Mrs. Kavitha, Kavita Kadam, Ashwini Koshti, Priya Dunghav, "Steganography Using Least Significant Bit Algorithm", International Journal of Engineering Research and applications, vol.2, issue 3, pp. 338-341May-June2012.

[5] A. Sarkar, S. Köhler, S. Riddle, B. Ludaescher, and M. Bishop, ``Insider attack identification and prevention using a declarative approach,'' in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2014, pp. 265_276.

[6] S. Goryczka, L. Xiong, and B. C. M. Fung, ``*m*-privacy for collaborative data publishing,'' *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 10, pp. 2520_2533, *Feb*. 2012.

[7] K. Chen and L. Liu, ``Geometric data perturbation for privacy preserving outsourced data mining,'' *Knowl. Inf. Syst.*, vol. 29, no. 3, pp. 657_695, Dec. 2011

[8] K. Liu, H. Kargupta, and J. Ryan, ``Random projection-based multiplicative data perturbation for privacy

preserving distributed data mining,'' *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 92_106, Jan. 2006.

[9] C. C. Aggarwal and P. S. Yu, ``A general survey of privacy-preserving data mining models and algorithms,'' *in Privacy-Preserving Data Mining Models and Algorithms*. USA: Springer, 2008, pp. 10_52.

[10] S. Hartley, *Over 20 Million Attempts to Hack into Health Database*. Auckland, New Zealand: The New Zealand Herald, 2014.

[11] K.-P. Lin and M.-S. Chen, ``Privacy-preserving outsourcing support vector machines with random transformation,''in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 363_372.

[12] M. Gönen and E. Alpayd_n, ``Multiple kernel learning algorithms,'' *J. Mach. Learn. Res.*, vol. 12, pp. 2211_2268, Jul. 2011.

[13] Amirthanjan,R. Akila, R & Deepika chowdavarapu, P., 2010. A Comparative Analysis of Image Steganography, International Journal of Computer Application, 2(3), pp.2010.

[14] Bandyopadhyay, S.K., 2010. An Alternative Approach of Steganography Using Reference Image. International Journal of Advancements in Technology, 1(1), pp.05-11.

[15]Chan, C.K. Cheng, L.M., 2004. Hiding data in images by simple lsb substitution: pattern recognition.vol 37. Pergamon.

[16] Atallah M. Al-Shatnawi, "A New Method in Image steganography with improved image quality", Applied mathematical science, Vol. 6, no79, 2012.

[17] Nagham Hamid, Abid Yahya, R. Badlishah Ahmad, Osamah M, "Image Steganography Techniques: An Over-view", International Journal of computer science and security, vol (6), Issue (3), 2012.

[18] Vijay kumar sharma, Vishal Shrivastava, "A Steganog-raphy algorithm for hiding image in image by improved LSB substitution by minimize technique", Journal of Theoretical and Applied Information Technology, Vol. 36 No.1, 15th February 2012.

BIOGRAPHIES

Nilima Kayarkar has received her B.E. degree in Information Technology in 2012. She is pursuing Master in Technology in Computer Science and Engineering from Wainganga College of Engineering and Management, Nagpur. Her areas of interest include security and image processing.

Gangotri Nathaney has received her B.E, degree in Information technology in 2011. She is a pursuing Masters in Technology in Computer Science and Engineering from Shri Ramdeobaba College of Engineering and Management, Nagpur-440013. Her areas of interest include Image Processing, Pattern Recognition and Artificial Intelligence