www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 4 April 2017, Page No. 21141-21144
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i4.59

Data Leakage Detection Using Cloud Computing

Abhijeet Singh, Abhineet Anand
Student, School of Computer Science and Engineering, Galgotias University, Greater Noida, India
avi3x4x@gmail.com
Professor, Department of Computer Science and Engineering, Galgotias University, Greater Noida, India
Abhineet.mnnit@gmail.com

Abstract

In the present scenario, the transfer of data is an important thing between one user to another. Data is mainly sent by the distributors which are generally the owner of data to the user which wants the information mainly the trusted third parties. The information sent by the distributor must be confidential and must be shared by a secure way. Some times during sharing of data, multiple copies of information is generated by different parties which cause a huge amount of loss, this is known as data leakage. To prevent this leakage of data one must put measures in order to detect the leakage in an early stage. The purpose of this paper to discuss data leakage and its prevention by various watermarking techniques.

Keywords- Watermarking, Data leakage, Discrete Wavelet Transform, Discrete Cosine Transform, robust watermarking.

Introduction

Cloud computing is one of the fastest emerging and bright technology in the current IT industry with almost every It company trying to get into it. Cloud computing is rising as the newest way to get delivery models for IT industry. It is a form of delivering IT services in the form SAAS, PAAS or IAAS.

Cloud is an alternative way for providing online services which are basically on demand such as networks, storage, server, software. It saves hardware cost and time by allowing pay by perusing i.e. pay according to user usage. In other words, Cloud computing is a mixture of new technology and platforms that provide storage and hosting services on the internet. Cloud computing provides less expensive on-demand infrastructures with good service.

The main advantages of using cloud computing are-

1. It reduces hardware and maintenance cost.
2. It is flexible
3. Can be accessed anywhere around the world
4. It is totally automated so need to worry about of software upgrades.
5. Off-site data storage.
6. Disaster assistance
7. Always up.

CLOUD BENEFITS
Easily upgraded
Productivity anywhere
Off site data storage
Lower cost of ownership
No IT maintenance costs
Always up
Disaster assistance

The cloud technology is totally dependent on the internet where the data is stored in its data centres of the service providers. The data security is one of the major challenges in the cloud computing technology.Less control over data may cause some serious security issues and threats which may lead to data leakage , data insecurity and attacks on data by an insider or an outsider. In order to save data from being leaked every IT company must focus on security issues of securing their data from different third parties.Sometimes the leakage is done by an insider mainly existing employees of the company, so the security must be beyond their employee's knowledge so that they could not have clue to crack it.

There is no specific time of data leakage it could happen at any time.The amount of damage done by a data leakage only depends on the quality of sensitive data leaked by the

DOI: 10.18535/ijecs/v6i4.59

person.If the data which is leaked is very much important to the institution.It may leave the institution in a helpless state.The leakage could lower the business and may result in the downfall of the company.

In order to prevent this problem different methods of data leakage prevention has been made such as Watermark method

Watermark Method

Watermark is a technique to prevent the copyright of the owner of the data.It is a technique in which a unique code is embedded in each distributors copy.It is basically encryption on a particular information which is to be distributed.The information can be in the form of image, video or any important file.The watermark helps the company to claim the ownership of particular information.

In this technique, a small pattern is added in the data mainly the tuples and the subset of data.The attributes of the tuple and subset are algorithmically coded so that they are controlled by a private key to be accessed only by the owner of data.This pattern represents the watermark.The data can be only accessed only when if a person has the key.

For detection of the watermark, access to the original data is not required.The watermark can be detected even in a small subset of data as long as the data contains some of the marks in it.The watermarking is done through a software which embeds watermarks by watermarking algorithms .the software introduces a few errors in the data.These errors are known as marks and all these marks together makes up a watermark.

Different Techniques of Watermark

-Watermarking by DCT (Discrete Cosine Transforms)

Discrete cosine transform (DCT) is a method for converting a signal into elementary frequency components.In this method, the image is first converted into 8x8 blocks of pixels.After DCT conversion, the mid-frequency range are selected which is based on Gaussian network classifier.Now the mid-frequency DCT coefficients are used for embedding. The DCT coefficients are modified by a linear DCT constraints.This will do not affect the visibility of the image and the watermark will not be removed by compression.

-Watermarking by DWT (Discrete wavelet transform)

This is a modern method which is widely used for watermarking, image compressions etc. This technique uses wavelet filters to transform the image. Wavelets are small waves of varying frequency and limited duration. The wavelet transform decomposes the image into three spatial directions horizontal, vertical and diagonal. The basic idea of DWT is to multi-differentiate decompose of the image into sub-image of the different spatial domain and independent frequencies.

-LSB (Least Significant bit)

In this method, the watermark is embedded in the LSB of pixels. This method is easy to implement but it is not very secure against attacks i.e. the watermark can be destroyed easily. The watermarking is done by choosing a subset of image pixels and then substituting the LSB of each of the chosen pixels with watermark bits.

-Watermarking by Embedding and Extraction

In this method, the insignificant part of the fractional part of the pixel intensity value of the main cover image is encoded to provide watermark.The watermark in the insignificant maintains the accuracy of the image.In this method the watermark is imperceptible.A large amount of watermark can be easily embedded and extracted using this method which will help companies and firms involved in digital information security products. It is an added advantage of this method. Various algorithms for embedding and extraction are used in this technique.

-Wavelet Based Watermarking

In this method, the multi-resolution data fusion is embedded where the image and watermark are both used and transformed into discrete wavelet form.The watermark is embedded into each wavelet level.The average of the estimates from each resolution level of wavelet decomposition is taken to detect the watermark.This algorithm works for JPEG compression, additive noise and filtering operations.

-Secure Spread Spectrum Watermarking

In order to save multimedia data such as audio, video and image, a watermark must be added in significant components of a signal if it is to be robust to common signal distortions and malicious attacks.But, the modifications of these components may lead to degradation of the data signal.So, watermark must be added to the spectral components of the data using techniques similar to spread spectrum communications, hiding a narrow band of the signal in a wide band signal. This watermark is very difficult for an outsider to remove, even many outsiders collude with different copies of the watermarked data.

-**Robust Watermarking Technique**

This watermarking technique can not only survive general operations such as compressions, adding noise, filtering and so forth but also geometric attacks such as rotation, scaling translation etc. It is often used for ownership protection. This method is used to encode ownership information directly into the data, so whenever the rightful ownership is in dispute, the data can be extracted and be used to find the rightful owner. The watermark should be stable during extraction and must not degrade the data.

-Watermarking of Digital Audio and Image using Mat lab

A watermark is encrypted using RSA algorithm and is embedded in the audio file using LSB technique. LSB is an old method which is not very vigorous against attacks. In this, the first watermark is encrypted and then embedded in the audio file, due to which removal of watermark becomes very difficult. This gives very high robustness. During retrieval of data, the embedded watermark is first retrieved and then decrypted. Similarly, for image watermarking, DWT technique is used. The watermark is embedded as a pseudo-noise sequence.This method makes image and audio very secure as the original watermark, must be known so that embedded watermark can be removed from the watermarked image or audio.

-Invisible Watermarking

This method provides an invisible robust watermarking scheme for embedding and extraction of a digital watermark in an image. One of the main features of this algorithm is that in this method a sub image used for watermarking. The watermark is embedded in the most significant region of the host image such that the modification of that region will corrupt the quality of the image. The watermark is created by two phases. The first phase involves synthesising of an image from the sub-image of the image. Then a compound watermark is created by embedding a logo (watermark) to the synthetic image by using a visible watermarking technique. This compound watermark is then invisibly embedded into the main block of the host image. This method proved its robustness and effectiveness under experimentation.

Need for securing the data

The data is mainly the information which might be sensitive and private. The information can be belonged to a firm, big IT Company or may be a simple user. If these data is leaked out it leaves the owner unprotected. The uncontrolled data can put a large organisation in a vulnerable position. When cyber criminals i.e. the one who stole the information sells it for money it costs the organisation money, reputation, brand value and its customer's trust. So, the data security is a necessity for the modern world where all the information is digital.

Conclusion

Data leakage is a silent but destructive type of threat. Sensitive information can be leaked without any knowledge. It may be an insider work or may be an outsider. So, sensitive data must be watermarked before its distribution so that it could be able to trace its origin with absolute certainty. When the information is watermarked it secures the data from being open to all and it becomes easier to find out the guilty by using the fake objects i.e. the watermark placed at different positions of information. This paper discussed various watermarking technique such as DCT, DWT, wavelet, invisible etc and their importance in data security.

References

[1] P. Papadimitriou and H. Garcia-Molina, “Data leakage detection,” IEEE Transactions on Knowledge and Data Engineering, pages 51- 63, volume 23, 2011.

[2] IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 3, March 2011 Data Leakage Detection Panagiotis Papadimitriou, Member, IEEE, Hector Garcia-Molina, Member, IEEE P.P

[3] An ISACA White Paper Data Leak Prevention P.P

[4] Ensaf Hussein, Mohamed A. Belal, “Digital Watermarking Techniques, Applications and Attacks Applied to Digital Media: A Survey”, IJERT, ISSN: 2278-0118, Vol. 1 Issue 7, September2012.

[5] Sandip A. Kale, Prof. S.V.Kulkarni, “Data Leakage Detection, International Journal of Advanced Research in Computer and Communication Engineering”, ISSN: 2278-1021, Vol. 1, Issue 9, November 2012.

[6] Upasana Yadav, J.P.Sharma, Dinesh Sharma, Purnima K Sharma, “Different Watermarking Techniques & its Applications: A Review”, IJSER, ISSN 2229-5518, Volume 5, Issue 4, April-2014.

[7] Cox, I.J.; Miller, M.L.; Bloom, J.A., “Digital Watermarking”, Morgan Kaufmann, 2001.

[8] Prabhishek Singh, R S Chadha, “A Survey of Digital Watermarking Techniques, Applications and Attacks”, IJEIT, ISSN: 2277-3754, Vol. 2 Issue 9, March-2013.

DOI: 10.18535/ijecs/v6i4.59

Authors Profile:-

ABHIJEET SINGH, pursuing B.TECH in computer science & engineering and specialization in cloud computing and virtualization in association with ”IBM” (2nd year) in “GALGOTIAS UNIVERSITY”.

DR. ABHINEET ANAND, Assistant Professor at “GALGOTIAS UNIVERSITY” and Program chair of “IBM” courses. (Aug 2016 present), Assistant Professor at ”UPES” (2012 to 2016), Director at “Rashcom Computer Education Pvt. Ltd.”. (Aug 1999 to 2012). Director at Arpan Assets and Finance Management Pvt. Ltd. Dates Employed Dec 2008 – Jul 2010. With his 15 years of academic and administrative experience, his research includes the following field of endeavour: Decision Tree, nearest neighbour method, Clustering, Rule induction, Optical Fibre Switching in Wavelength Multiplexing, Automata Theory.He has published more than 20 papers in International conference, 4 Intentional Journal, 3 National Journal and 3 National Conference. He has been part 6 special session at various conferences at international level as session chair/co-chair, contributed at 6 different conferences as Technical Program Committee member. His expertise also includes reviewer at more than 10 conferences and Publication group.

-