iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 8 **Issue:** V **Month of publication:** May 2020

**DOI:** http://doi.org/10.22214/ijraset.2020.5387

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 8 Issue V May 2020- Available at www.ijraset.com*

# Data Leakage Detection using Watermark Technique: A Survey

Shivam Agarwal[1], Shivaprabhu S K[2], Shraddha Banka[3], Poornima KS[4]
[1, 2, 3]*Students,* [4]*Assistant Professor, Computer science & Engineering, Dayananda Sagar College of Engineering, Bangalore, India*

***Abstract:*** ***Nowadays data leakage is a major issue which is causing a great loss in the IT world. That's why it should be protected thus decreasing the loss. A huge amount of data is being transferred on a daily basis, so it must be protected. To prevent these problems lot of surveys has been done and a lot of detection systems have been created. In this paper, we are discussing abo ut some of the techniques of data leakage detection.***
***Keywords: Watermarking, Guilty agent, Data leakage.***

## I. INTRODUCTION

As we know in today's world, the value of data is huge. So we have to protect it from any kind of leakages. Several encryption algorithms has been used for data security but still the leakers always find a way to disclose the valuable data which is causing a great harm to industry people. It's very difficult for anyone to identify who exactly the leaker is. Without any solid proof one cannot blame anyone for leakage. It creates ethical issues in the working environment.

A lot of technologies like Encryption, Identity management , firewalls, Access control, Machine learning content based detectors and others have already being imposed for the protection of the valuable data.

The effect of data leakage can be disastrous. it can vary from loss of valuable data, privacy, cyber-theft, to threat to economy and national security. In this paper we will be discussing about the major techniques such as h watermarking, fake data addition, guilty party for the detection of the data leaker with proof. The system should be able to perform in wide areas so that it doesn't restrict us and limit our scope.

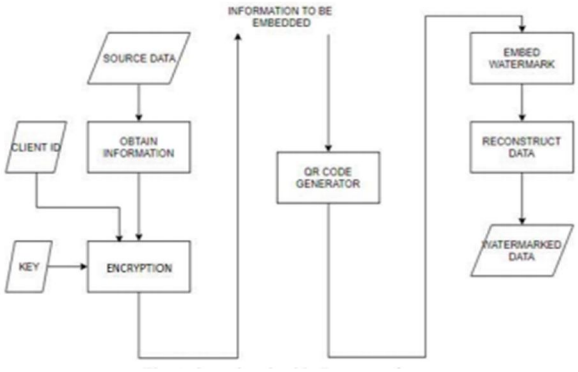
## II. WATERMARK ARCHITECTURE

In the data transfer architecture, we have two attributes. One is the distributor and the other is agent.

The distributor is from the company who is assigned to distribute the data to the third party.

Distributor are responsible for the data storage, distribution and the detection of tamper and finding the agent who is guilty.

The beginning phase is the retrieval of information and encryption of data. The message encryption is done using AES i.e. advanced encryption technique to restrict any tampering of data. We are using AES because it gives the flexibility of the key size. The DES algorithm has restrictions on the size of key. The key length of AES algorithm can be used with various combinations. 128 bit key length can use up-to 10 rounds, 192 bit to 12 rounds and 256bit to 14 rounds. The original AES key is different from our generated key [6].

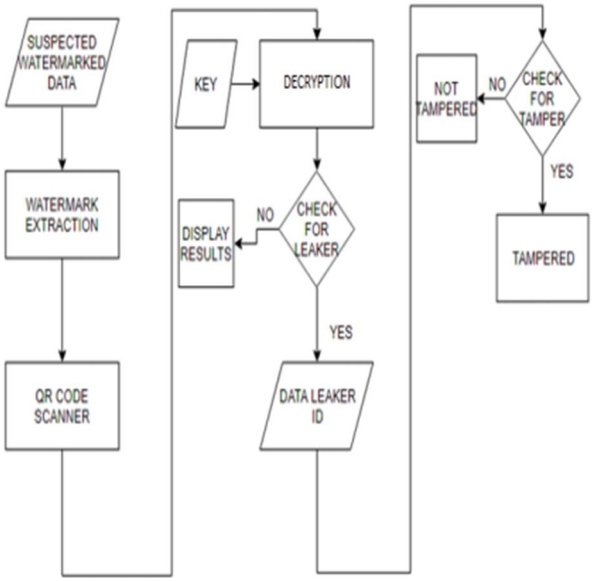



Steps involved in data transfer

The leaked data is kept under examination to extract the watermark. The QR code embedded is extracted and matched with the key given to the agents. With whoever the key matches is considered as the guilty agent. The key details are already present in the database. The extracted QR code from the data is then decoded using the AES decoding algorithm. The extracted data is then compared to our data with is present in the cloud database.

The security key given to every agent is stored in the cloud database which is then compared. The key is matched with the agent's key and hence the guilty agent is found.[1].

## III. LITERATURE SURVEY




Tamper detection and Data leaker detection.

A brief about the usage of Watermarking technique in different styles by different authors has been discussed in this section.

In this research paper[2], the authors are trying to convey a very interesting approach of encoding the insignificant portion of the fractional part of the pixel intensity value of the cover image to provide watermark. The fidelity of the cover image is maintained by the watermark in the insignificant part. The scaling factor value needs to be adjusted based on the visual property of the host image so that the proportion of the robustness and invisibility is perfectly maintained for optimal tradeoff. Selecting the part in which we should apply watermark is very critical. Here they propose an efficient buyer seller watermarking protocol based on homomorphism public-key crypto-system and composite signal representation in the encrypted domain to reduce both the computatio- nal overhead and the large communication bandwidth[1].

Here the approach used by the publisher is very smart and sensible, instead of perturbing of real data they are adding fake elements which does not harm the real data but looks very similar. In this way the data is modified but the meaning also remains same[5].

The most important part of catching the culprit is made possible by addition of forged data. Without these entities data administrator must not share the sensitive data. The technique used in here[7] for the generation of forged entities is very unique and special.

Forged entities are basically the data which more likely seems to be the original data but isn't. Once these are added to the original data in the unique manner it is distributed to the concerned parties.

Here they have used the original data for the generation of forged data, out of many algorithms, and it seems to be a better way of hiding the gibberish data. The reason behind the selection of this algorithm is simple, they wanted to ensure that each copy of the data or a file has unique data such that it does not raise any suspicion and helps in nabbing the data leaker.

For the generation of forged data for an agent, referred as UI, they have used a black-box function CREATEFORGEDDATA(Ri, Fi, Condi), that automatically catches input as set of all data Ri.Set Ri is needed as input so that the created forged entity is not only valid but also identical from other original entity[7].

In this paper[8], they have used watermarking technology on video file for which the methodologies used are Discrete wavelet transform, correlation based method, principle component analysis and discrete Fourier transform. It deals with the challenges and problems in content distribution. The digital watermark they have used in this peer-to-peer network is very robust fragile. Fragile so as to protect the integrity of the sensitive data and robust so as to protect the watermarking copyrights.

This paper discusses the requirement of digital watermarking which must have all the elements in correct proportion such as security, robustness, modification, imperceptibility and multiple watermarks inevitability.

Digital Watermarking is the technique of encrypting the digital data and insert that data into the digital signals. It is the best used for authenticity and integrity of the sensitive data. They have used spatial domain algorithms for digital watermarking to directly load the data in the original image. Color separation could also be used for the same.

Spatial domain method has high efficiency than transform domain method as it is easy and has high computing speed.

DCT is also one of the algorithm to apply digital watermarking, and is more robust than spatial domain technique. It represents data as frequency space, like Fourier Transform, rather than an amplitude space[9].

Image watermarking is a watermarking approach in which the technique is applied on an image. The watermark is embedded in image in some forms for example we can take color components in such a way the it can be detected when it is cross verified for leakage.

There is something called Capacity of watermarking which refers to the amount of information or secret data that can be hidden in the image without being noticed by the receiving party under normal condition.

Also some approaches use text based encryption in which they can insert key as text in the image with similar texture which is unnoticeable.

A co-relation detector is used to detect the watermark embedded in the transferred image[10].

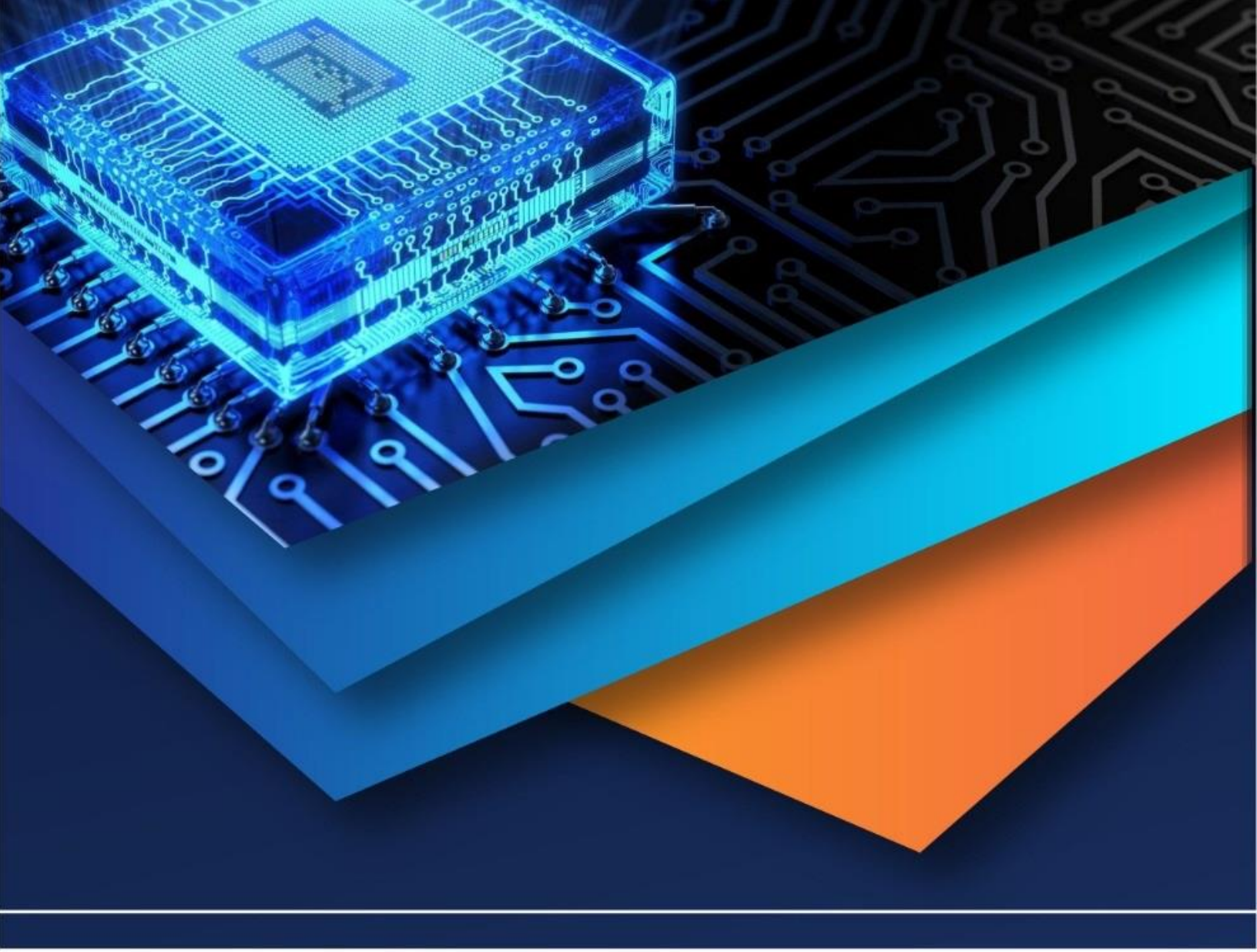
## IV. CONCLUSION

In a perfect world there shouldn't be any leakage in the data transfer. The agents who are responsible for data leakage has to be identified so that they should not continue doing it. It can be done by watermarking the data we are sending using some encryption decryption technique we have discussed earlier. We are looking forward for some better approach that helps in identifying the guilty agent. We are trying to make some modifications in the already existing approaches and present it in a new and better form which will be beneficial in future.

As of now we are trying to create a website using web technologies which will have a login portal for both distributors and agents from where they can send and receive data and the distributor can check for the data leaking agent.

## REFERENCES

[1] Mrs. Grinal Tuscano et al. Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 5, Issue 4, ( Part -6) April 2015, pp.153-158
[2] International Journal of Advanced Research in Computer and Communication EngineeringVol. 1, Issue 9, November 2012 Sandip A. Kale, Prof. S.V.Kulkarni
[3] International Journal of Advances in Engineering Research (IJAER) 2014, Vol. No. 8, Issue No. V, November K. Manoj Kumar, G. Shubhang, G. Rajesh Chandra
[4] Sandip A. Kale1, Prof. S.V.Kulkarni Department Of CSE, MIT College of Engg, Aurangabad, Dr.B.A.M.University, Aurangabad (M.S), India1,2
[5] Panagiotis Papadimitriou, Student Member, IEEE, and Hector Garcia-Molina, Member, IEEE
[6] Riya Naik, Manisha Naik Gaonkar Department of Computer Science & Engineering Goa College of Engineering Farmagudi, India
[7] K. Manoj Kumar+ G. Shubhang+ G. Rajesh Chandra Assistant professor, Department of Electronics and Computer Engineering, K L University, India. B Tech, Department of Electronics and Computer Engineering, K L University, India
[8] Sivasri S, Parthiban R Department of I.T, IFET college of Engineering, villupuram, India. Assistant Professor, Department of I.T, IFET college of Engineering, villupuram, India .
[9] Prof.A.S.Kapse, Sharayu Belokar, Yogita Gorde, Radha Rane, Shrutika Yewtkar Professor. A.S.Kapse, Dept. Computer Science & Engineering, P.R. Pote College of Engineering & Tech, Amravati, Maharashtra, India.
[10] Christine I. Podilchuk and Edward J. Delp IEEE SIGNAL PROCESSING MAGAZINE



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY