iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.41731

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: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Research Paper on Detection and Prevention of Data Leakage

Dhiraj Gupta[1], Dr. Umarani Chellapandy[2]
[1]*School of CS and IT, Jain University, Bangalore*
[2]*Professor, School of CS and IT, Jain University, Bangalore*

Abstract: ***A data distributor has given sensitive data to a group of supposedly trusted agents (third parties). Data are leaked and found in an unauthorized place or in the hands of an un authorized person. There must be an acknowledgement by the distributor that the information is compromised from one or more agents instead of being gathered independently. We suggest data allocation strategies that upgrade the likely hood of identifying leakages. These methods don't believe alterations of the released data.***
Keywords: ***data distributor, leaked data, data allocation, the guilty party***

I. INTRODUCTION

When there is an unauthorized transferal of data from an organization to an external source then it is known as data leakage. An organization keeps data as its most important property. The data in an organization can include a customer's information including his financial details. Leakage of this type of information can have a big impact on the customer and the organization.

The data that is transferred can be electronical or physical. Usually, the threats for data leakage comes from email, mobile phone, USB drives, laptops and web. Every single day, there is some confidential breach of data because of low data security.

Leakage of data moves it to the hands of an unauthorized individual which can cause a loss to the customer or the organization directly or indirectly. To solve this problem, there is a need of a well-organized system that will protect the data and report any threats on its leakage. This paper tries to analyze this problem and built and effective system to solve this problem.

II. ANALYSIS AND INTERPRETATION

Data Leakage detection can be handled by watermarking like embedding a singular code in distributed copy. If that replicate is later discovered within the hands of an unauthorized party, the leaker is often identified. Watermarks are often very efficient in some cases, but again, involve some alteration of the first data. Again, watermarks can be destroyed if the information recipient is malicious. If a hospital gives records of patients to researchers, they will devise new treatments.

A corporation may have partnerships with many companies that need sharing customer data. Another enterprise may outsource its processing, so data must tend to varied other companies. We call information owner, the distributor and therefore the supposedly trusted third parties the agents. The existing methodology also includes a performance metric for distance of behavior and its performance is analyzed with the use of Dynamic Time Warping. This system tries to give consistent character of the behavior of the host.

The goal of our system is to disclose when a distributor's crucial information has been leaked by agents. Then we try to identify the agent that has leaked the information.

Data is altered and made less crucial before being handed to agents. We develop demure techniques for leakage detection of a set of records or objects. In this section we attempt to develop a model to assess the guilt of some agents.

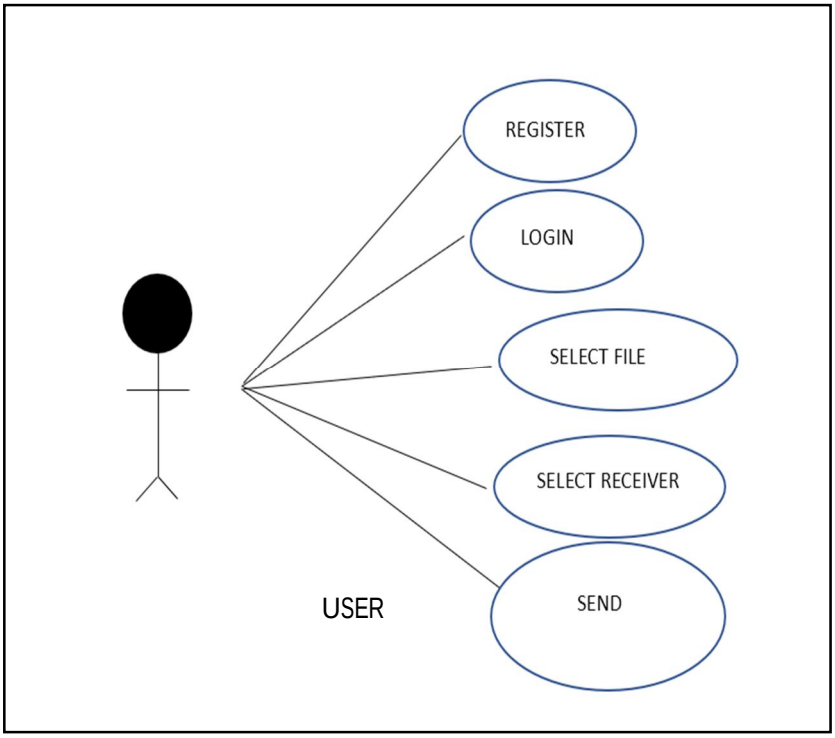
We use algorithms to distribute objects among agents, in such a way that improves the possibility of identifying a leaker. We also add fake objects to the distributed set by keeping it as an option. Such objects are not real entities but seem realistic to the agents. The fake objects act as a type of watermark, without altering any individual members for the entire set. If the agent gives more than one or one fake objects that were leaked, then the distributor can be sure that agent was guilty.

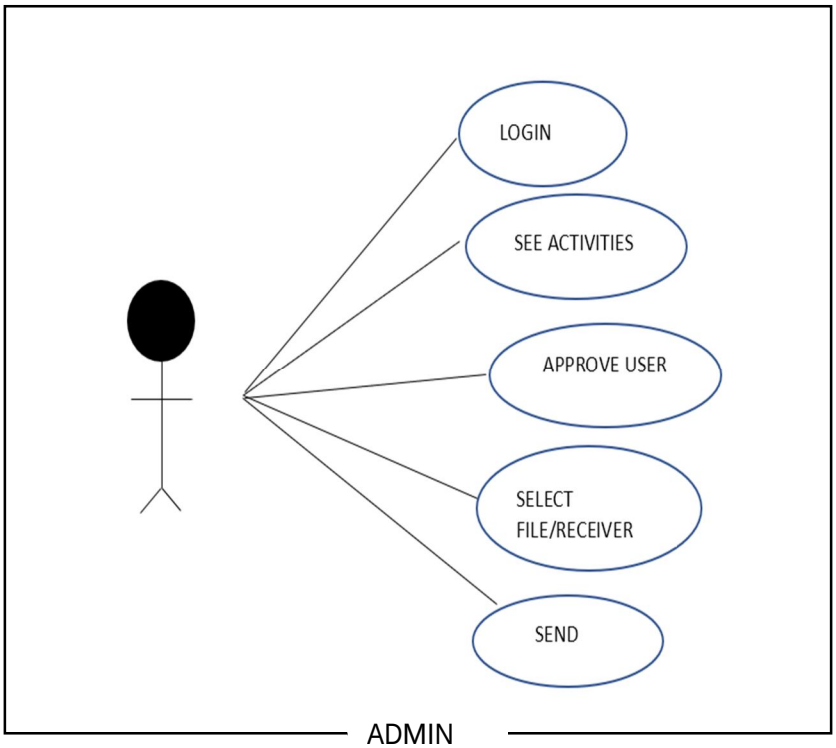
©IJRASET: All Rights are Reserved

III. SYSTEM DESIGN

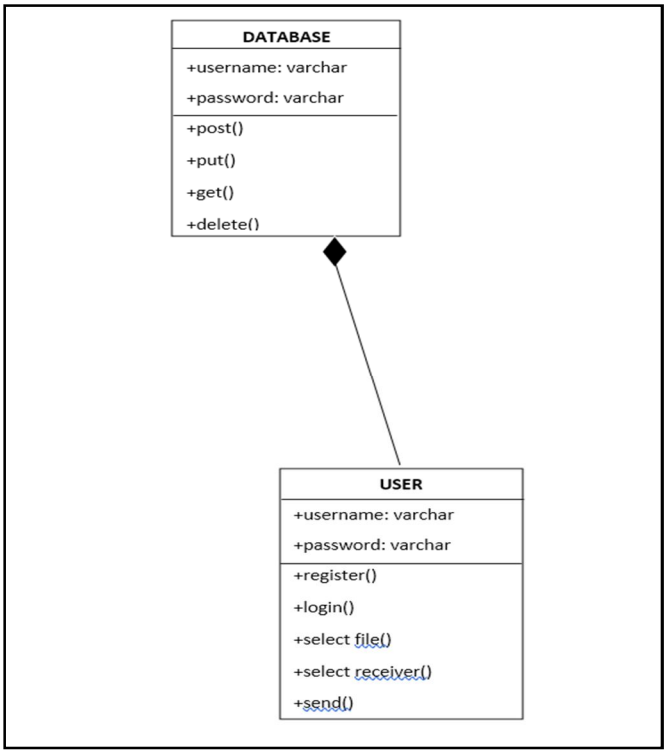
A. *UML Diagrams*

1) *Use Case Diagram:* A use case diagram in the Unified Modeling Language (UML) is a shape of behavioral diagram defined thru the usage of and crafted from a Use-case evaluation. Its reason is to provide a graphical evaluate of the functionality provided with the useful beneficial useful resource of a tool in phrases of actors, their dreams (represented as use times), and any dependencies a number of the ones use times. The primary purpose of a use case diagram is to expose what tool capabilities are finished for which actor. Roles of the actors in the tool may be depicted.

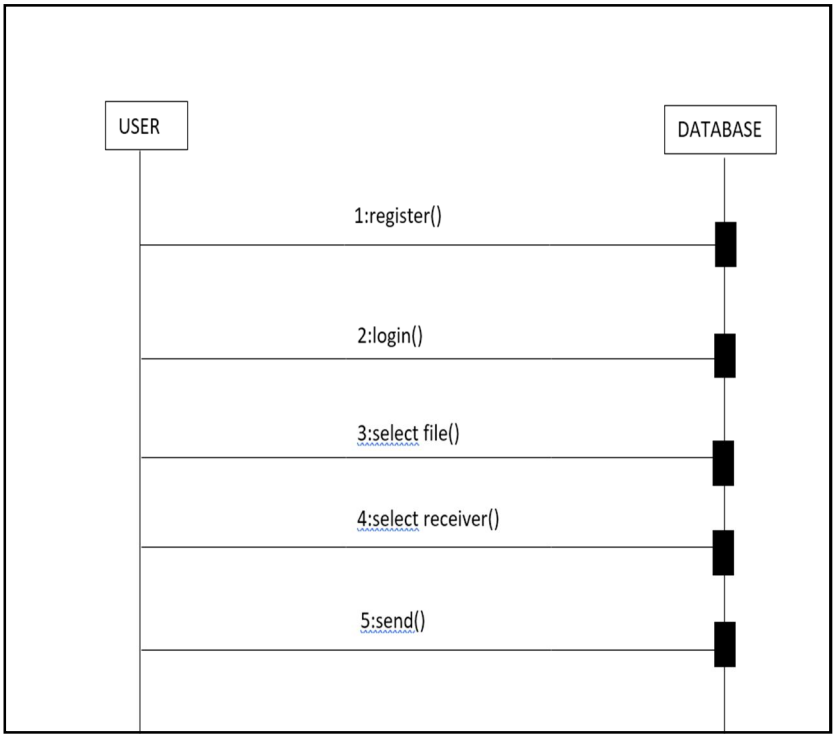




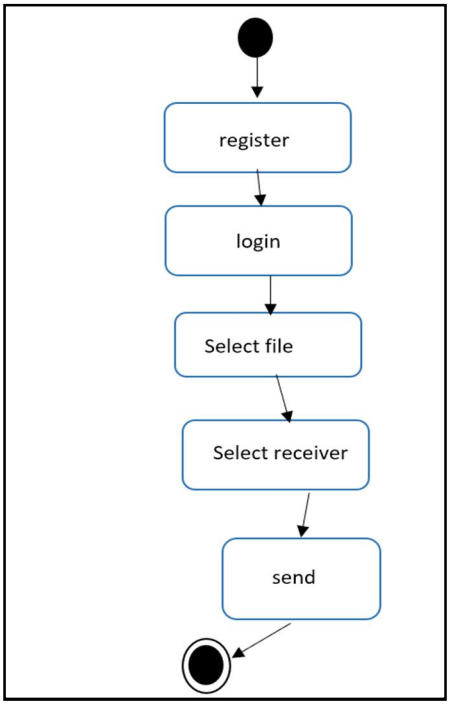
2) *Class Diagram:* In software engineering, a category diagram within the Unified Modeling Language (UML) is a form of static shape diagram that describes the form of a device with the beneficial aid of displaying the tool's training, their attributes, operations (or techniques), and the relationships most of the instructions. It explains which splendor includes statistics.



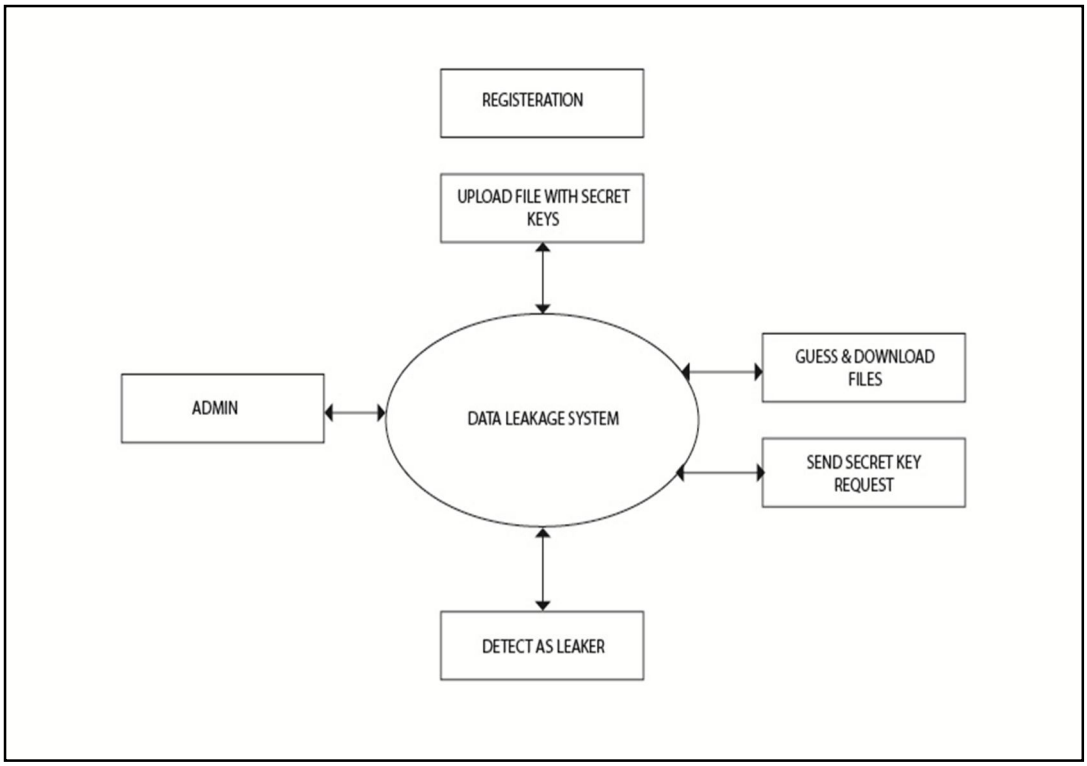
3) *Sequence Diagram:* A series diagram in Unified Modeling Language (UML) is a shape of interplay diagram that suggests how techniques perform with one another and in what order. It is a acquire of a Message Sequence Chart. Sequence diagrams are occasionally called event diagrams, occasion situations, and timing diagrams



4) *Activity Diagram:* Activity diagrams are graphical representations of workflows of stepwise sports activities sports activities and movements with manual for desire, generation and concurrency. In the Unified Modeling Language, interest diagrams can be used to give an cause at the back of the economic company and operational step-via using-step workflows of components in a gadget. An hobby diagram suggests the overall go with the flow of manage.

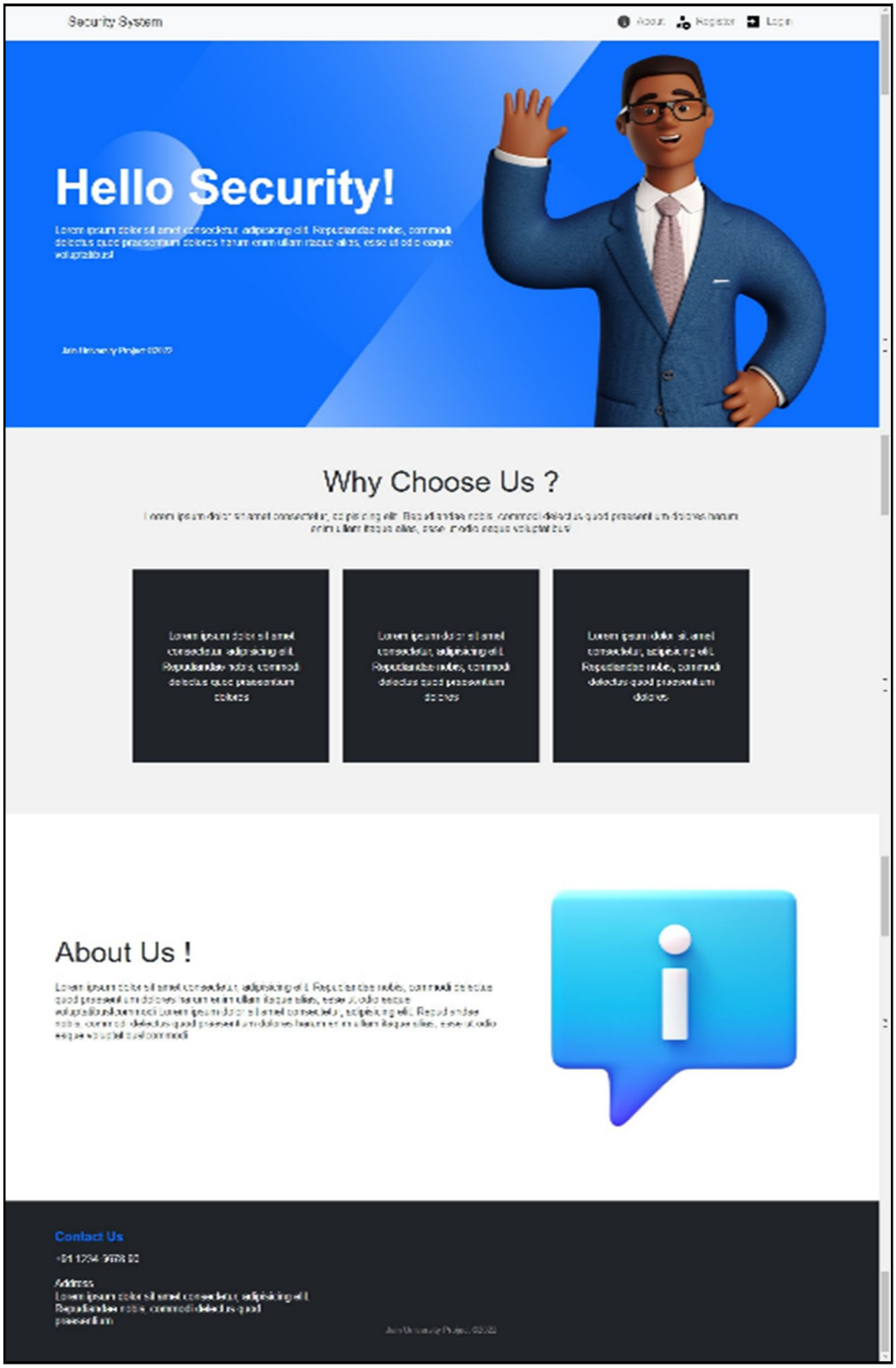


5) *Data flow Diagram:* DFD is a simple graphical representation of flow of data of a process or a system. A DFD uses very limited number of signs to represent the functions performed by the system.

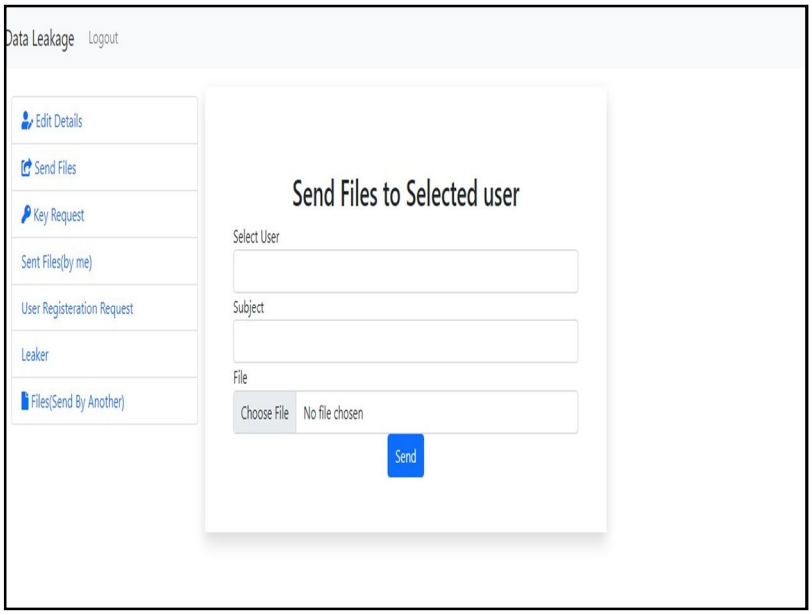


IV. IMPLEMENTATION

1) *Home Page:* This is the home page of our web application. A user can navigate through the home page and can register with the application. A regular user can also login using his credentials.



2) *Admin Page:* The admin of the web application can log in to the admin page. He can monitor all the activities done by the users and also remove suspicious users.



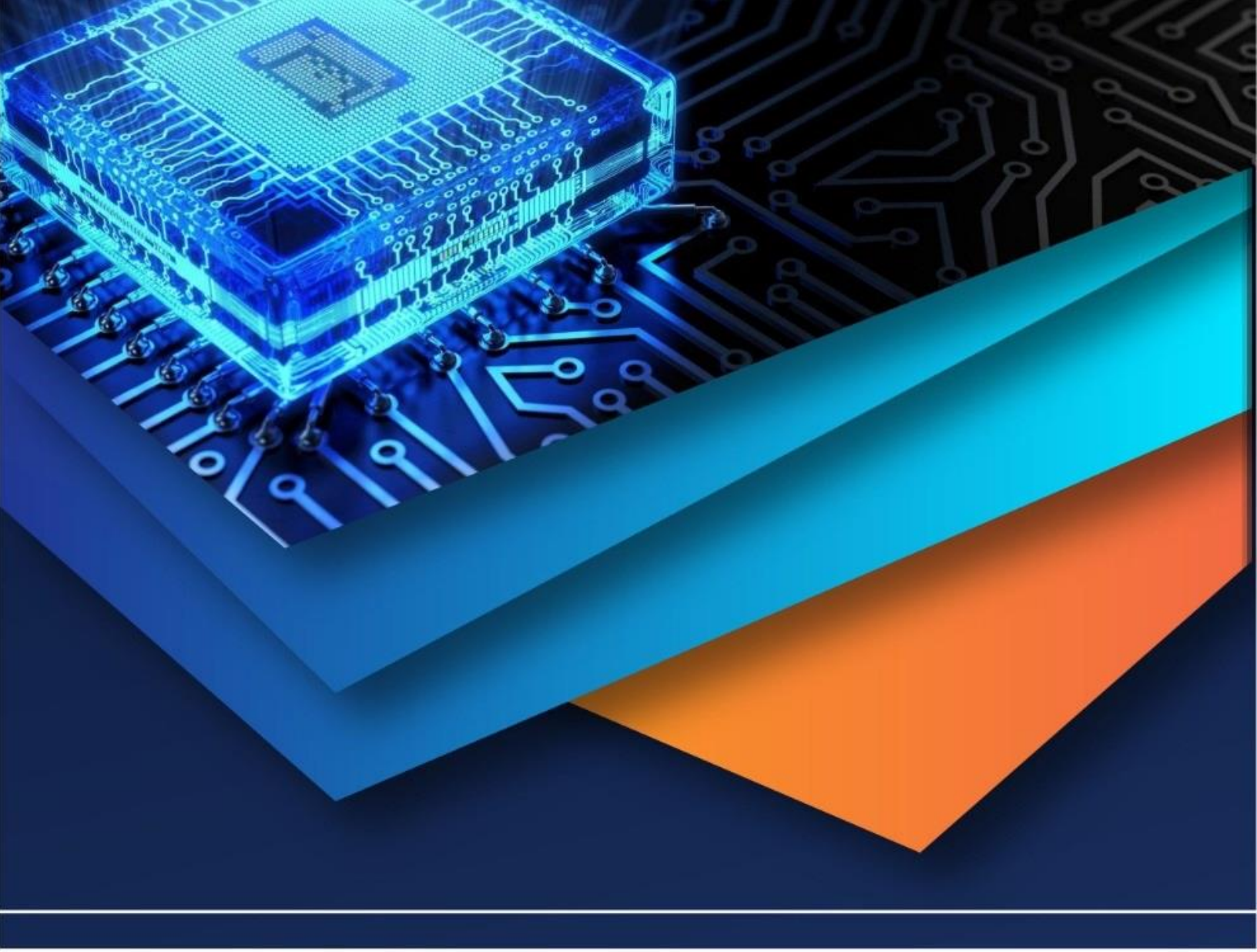
V. CONCLUSION

Data leakage is a problem which is faced by every other organization in real world. It is very convenient for an unauthorized user to access leaked data and harm any organization or an individual. This data leakage system will be efficient for protecting the leaks of data and alarming the owner of the data for a potential threat for data leakage. Data Leakage is a very big industry and it causes a crisis in information security. It is very important to protect data from any kind of leak. If a confidential information of an organization is leaked than it may cause loss to the organization financially and effect its brand-value. This paper concludes that data leakage detection and prevention is extremely useful and is the need of the hour.

REFERENCES

[1] Fast Detection of Transformed Data Leaks(2014)
[2] Privacy-preserving detection of sensitive data exposure(2015)
[3] Privacy-preserving scanning of big content for sensitive data exposure with MapReduce(2015).
[4] A. Nadkarni and W. Enck, Preventing accidental data disclosure in modern operating systems(2013).
[5] R. Hoyle, et al., Attire(2013)
[6] H. A. Kholidy, et al., DDSGA(2015).
[7] Y. Jang, et al., Gyrus(2014).
[8] L. D. Carli, et al., Beyond pattern matching (2014)
[9] X. Shu, et al., Rapid and parallel content screening for detecting transformed data exposure (2015).
[10] S. E. Coull, et al., On measuring the similarity of network hosts: Pitfalls, new metrics, and empirical analyses (2011)
[11] S. Yin, et al., Distributed Searchable Asymmetric Encryption(2016).
[12] H. Kaur and M. Kaur, KAMAN (2015).
[13] M. Altayeb, et al., Wireless Sensor Network for Radiation Detection (2017)

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)