|  |  |  |  |
| --- | --- | --- | --- |
| Research paper citation  **Literature Survey** | Overview | Tools/Algorithms | Problems Identified |
| 1.TY - JOUR  AU - Wagh, Dr. K  PY - 2018/08/01  SP - 2247  EP - 2253  T1 - A Survey: Data Leakage Detection Techniques  JO - International Journal of Electrical and Computer Engineering | The organizations are facing the problem of data leakage .The data may be seen in other laptops or websites. The data leakage detection system is largely useful for protecting the illegal use of data of various industries. So there is need to develop a content inspection method which detect leaks of important data in the content of files or network traffic. Also proposed system is useful to detect modification in data. In future such systems are necessary to detect data leak of personal, finance transactions, online shopping, and social media and so on. | Subsequence preserving sampling algorithm is used to generate sampling algorithm.Recurrenc-e relation dynamic programming  Algorithm. | Data-movement  tracking approached is not used. Time consuming Process.  This method not  work on allow for  localized reordering of Points. |
| 2.Praba, C.M. and Satyavathy, G., 2017. A technical review on data leakage detection and prevention approaches. *Journal of Network Communications and Emerging Technologies (JNCET)*, *7*(9). | Statistical data leakage prevention (DLP) model is presented to classify data on the basis of semantics. This study contributes to the data leakage prevention field by using data statistical analysis to detect evolved confidential data. The approach was based on using the well-known information retrieval function Term Frequency-Inverse Document Frequency (TF-IDF) to classify documents under certain topics. A Singular Value Decomposition (SVD) matrix was also used to visualize the classification results. The results showed that the proposed statistical DLP approach could correctly classify documents even in cases of extreme modification. | Statistical Data leakage prevention (DLP) classification approach.  Singular Value Decomposition (SVD) matrix was also used to visualize the classification results. | Only 60 percent of the modified documents were able to identify.  Scalability and integrity issues. |
| 3.Gupta, K. and Kush, A., 2017. A review on data leakage detection for secure. *Internationa -l Journal of Engineering and Advanced Technology (IJEAT)*, *7*(1), pp.153-159 | The industry for data leakage detection is extremely heterogeneous because from the study it has been cleared that the term data leakage detection is come into existence after using a [1]big number of technologies for example firewalls, encryption, access control, identity management, machine learning content/context-based detectors and much more. Watermarking is the most common technique for leakage detection, which implements a unique code in every copy of data so that origin of leakage can be traced with absolute certainty. Sometimes watermarking is not applicable to all the data, in case of such difficulties, [2]a concept of overlap of shared data is used which makes possible to find out the probability that whether an agent caused a leak. | Guilt Assessment Algorithm with data allocation strategy | [1]Identify the data leakage happening over encrypted channels becomes difficult.  Access Control Challenge.  Process may take long time.  [2]It is not adequate to capture diverse communication groups where people belong to multiple groups. At that time it is difficult to reveal a person leaking data (an outsider) in a communication or to detect persons having access to limited access data. |
| 4.Verma, R., Gautam, V., Yadav, C.P., Gupta, I. and Singh, A.K., 2020, May. A Survey on Data Leakage Detection and Prevention. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*. | [1]Distributor adds some fake object with the data and send to the agent. The fake object helps to detect the user who have leak the file.[2]We see mainly two aspects where the first one is fake objects including in-database and second is data allocation strategy to distribute the data to customers with the minimum transaction | Fake object Algorithm in Sample Data Request and Explicit Data request is used. | It’s difficult to analyze the data file.  Secure their data at the time of modification. Channels are other than specific ones such as USB, email and another format then it makes it difficult to secure their data.  Data size is huge so that monitoring, matching and accessing their data becomes more difficult. |
| 5.Shaj, V. and Kaliyamurthie, K.P., 2013. A review of Data Leakage Detection. *IJCSMC Journal*, *2*, pp.577-581. | The competitive benefits of developing a "one-stop-shop", silver bullet data leakage detection suite is mainly in facilitating effective orchestration of the enabling technologies to provide the highest degree of protection by ensuring an optimal fit of specific data leakage detection technologies with the "threat landscape" they operate in.[1]This landscape is characterized by types of leakage channels, data states, users, and IT platforms. | Evaluation of Explicit Data Request Algorithms.  Evaluation of Sample Data Request Algorithms | [1]Encryption:-  It difficult to identify  the data leaks  occurring over  encrypted channels.  [2]Access Control:-  Once the data is  retrieved from the  repository, it is  difficult to enforce  access control. |
| 6.[1]Mohammed Ghouse DM, Network and Cyber Security, Bharat Electronics Limited, Bangalore, India mohammedghouse@bel.co.in  ,[2] Manisha J. Nene Dept. of Computer Science and Engineering, Defense Institute of Advanced Technology, Pune, India mjnene@diat.ac.in,[3] VembuSelvi C MRS, Cyber Security Central Research Laboratory, BEL, Bangalore, India vembuselvic@bel.co.in | [1]The work in this  paper addresses a novel concept for prevention of data leakage for data in transit. The text under consideration is classified to confidential or non-confidential category based on the content and context using Machine Learning technique.  [2]A novel concept for DLP for data in transit is introduced, accomplished using techniques of Machine Learning employing the content and context based text classification and later employing the encryption technique to preserve the confidentiality . | [1]Scrambling,Perturbation&Interleaving of the datas. Understand on clustering and classification of texts/documents, and also various approaches are used eg: attribute reduction method, graph representation of texts/documents is presented. [2]A D-SeGATe architecture for data in transit is introduced to leakage of data via means of documents. This approach reduces the risk of confidential data leakage outside the organization without affecting the business life cycle of the project. | [2]Recently Data Diode was introduced to DLP for data in transit but in this case also problem arises for data from Secure domain to Unsecure domain(Internet Domain) is completely blocked. [3]Thus the user is unable to send any information to his/her peers which ultimately results in an adverse impact on the progress of the project and hence affects the business prospects of the organization and the reputation of the organization. |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |