SURVEY PAPER ON DATA PROVENANCE METHODOLOGY FOR SECURE COMPUTING

V.UMARANI (ASST.PROF. OF CSE, SCHOOL OF IT, JNTUH),
DR.M.SREENIVASA RAO (PROF. OF CSE, SCHOOL OF IT, JNTUH),
FARHANA (M.TECH, SCHOOL OF IT, JNTUH)

Abstract:

In this technical era, the data value is incredible. In information technology massive database is shared with multiple people at a time. While sharing this data, there are more chances of data Vulnerability or alteration. So, to overcome these problems, a data leakage detection and data lineage system has been proposed. Generally, to protect private information of the clients, the agent who is responsible for the leakage should always be identified at an early stage. Always, the data detection or flow from the administrator to clients is mandatory. This project presents how to trace the origin of the data leakage using different method and which assess the chance that the leaked data came from one or more agents. For secure transactions, allowing only authorized users to access sensitive data through access control policies shall prevent data leakage by sharing information only with trusted parties and also the data should be detected from leaking by means of adding fake record`s in the data set and which improves probability of identifying leakages in the system.

Introduction:

The organization's Data security relies on upon workers by taking in the standards through training and awareness- building sessions. In any case, security must go beyond the knowledge of an employee and covers the physical and logical security mechanism that is adjusted to the necessities of the organization and to worker use then the technique for overseeing overhauls lastly it needs an up to date documented System. Data framework security is regularly the subject of metaphors. It is regularly contrasted with a chain in the illustration that a framework's security level is just as solid as the security level of its weakest connection. This demonstrates the issue of security must be tackled.

At a global level and should include the components like making users aware of security issues then the coherent security, i.e. security at the information level, quite organization information, applications and not withstanding working frameworks furthermore items utilized as a part of Information transfers security, for example, system advances, organization servers, access systems, and so on. Information leakage happens each day when confidential business information, for example, client or patient information, source code or design details, price lists, intellectual property and trade secrets, and forecasts and budget plans are spilled out. whenever these plans are leaked out by authorized or unauthorized user it leaves the organization unprotected and goes outside the jurisdiction of the company. This uncontrolled information leakage places business in a defenseless position. When this information is no more inside the specific domain, then the organization is at serious risk.

When digital criminals cash out or offer this information revenue driven it costs our association cash, harms the upper hand, brand, and notoriety and obliterates client trust. To address this issue, we build up a model for evaluating the guilt of specialists. The vendor distributor will intelligently offer information to specialists keeping in mind the end goal to enhance the odds of distinguishing blameworthy operators like adding the fake items to disseminated sets. Right now the distributor can survey the probability that the spilled information originated from one or more specialists, rather than having been autonomously accumulated by different means. On the off chance that the merchant sees enough confirmation that a specialists leaked information then they may quit working with him, or may start lawful procedures. Essentially it has one Furthermore, one objective. The distributor's imperative fulfills the users, by giving number of objectives they ask for that fulfills their confidence.

RELATED WORKS:

1: Rights Protection is provided for Relational Data

[1] RaduSion, Mikhail Attalla, and Sunil Prabhakar concentrate on giving the rights protection for relational database using watermarking technology. Rights protection relational data is of constantly expanding interest, specifically areas like where confidential information outsourced. Data Set is divided into partitions using markers and then varies the partition statistics to hide watermark bits. It proposes a watermark embedding algorithm such that it comprises of Sorting, Parceling used for marker location and bit embedding watermark embedding algorithm such that it comprises of Sorting, Parceling used for marker location and bit embedding watermark bits are embedded in the number set so as to provide a right security to the data that are stored into it the relational database.

It additionally builds up a watermark detection algorithm such that it comprises of Sorting, Dividing used For marker area and bit detection algorithm at the time of Retrieving data from the database in its customer side. The major flaw is that it should not deal on the territory of information security through watermarking in the system of non numeric encoding areas in this relational database.

2: Watermarking Technique for Multimedia Data

[2],[3] Hartung and Kuttur on the Multimedia watermarking technology that has developed rapidly. A recent development and success of the Web, with accessibility with the low cost, digital recording and storage devices has made a domain in which it turned out to be anything but difficult to get duplicate and computerized.

While the unauthorized access to digital content can be prevented by encryption techniques, it is clear that once content is decrypted the encryption has its constraints in ensuring intellectual property rights, and there's no technique to keep an authorized client from illegally duplicating digital content. Other technology was clearly expected to build up and demonstrate ownership rights, track content usage, guarantee approved access, encourage content verification and anticipate unlawful replication. A computerized watermark is data that is indistinctly and vigorously inserted in the host information such that it can’t be leaked.

A watermark ordinarily contains data about the origin, Status or beneficiary of the host information. It gives the Prerequisites and all the related applications for watermarking is reviewed. The application incorporates copyright Security, information monitoring, and information tracing. Robustness and security aspects are additionally examined in specific data source. Finally, a few remarks are made about the state of the art and conceivable future improvements in Watermarking innovation.

[4] J. Cox projected a secure (tamper-resistant) rule for watermarking images, and a technique for digital watermarking that will be generalized to audio, video, and multimedia system information. we have a tendency to advocate that a watermark ought to be created as an independent and identically distributed (i.i.d.) Gaussian random vector that's continuously inserted in an exceedingly spread-spectrum-like fashion into the perceptually most important spectral elements of the data. we have a tendency to argue that insertion of a watermark below this regime makes the watermark powerful to signal processing operations (such as lossy compression, filtering, digital-analog and analog-digital conversion, requantization, etc.), and general geometric transformations (such as cropping, scaling, translation, and rotation) on condition that the first image is out there which it is with successfully registered against the remodeled watermarked image. In these cases, the watermark detector unambiguously identifies the owner. Further, the utilization of Gaussian noise ensures robust resilience to multiple-document ,or collusion attacks [5] Rakesh Agrawal and Jerry Kiernan concentrate on watermarking the relational databases. It proposed that watermark can be applied to any database relation having attributes which are such that modifications in a few of their values don't affect the applications. They articulate the requirement for watermarking database relations to deflect their piracy, distinguish the different characteristics of relational data which pose new challenges for watermarking, and provide required properties of watermarking technique for relational data. While adding watermark, the changes in the attribute values of a relational database don’t affect the applications. Then this would be called as an effective watermarking for relational data.

3: Lineage Tracing General Data warehouse Transformations

[6] Yingwei Cui and Jennifer Widom Was proposed leakage tracing general data distribution transformations. In a warehousing domain, finding of the origin of data leakage issue is that of tracing warehouse data items back to the actual source items from which they were determined. It formally characterizes the tracking the lineage while changes made to the warehouse, and they apply algorithms for source tracking in this environment. These methodologies take the advantage of known structure or properties of transformations when present, additionally work without Specific attribute values.

Only the owner the data knows the private key which determines the tuples, attributes, bit positions and specific bit values algorithmically. This bit design constitutes the watermark, only if one has the access to the private key can the watermark be recognized with high likelihood. Identifying the watermark neither requires access to the original data, nor the watermark.

The watermark can be detected even in a small sub set only if the sample contains some of the marks. Our analysis shows that this technique is robust against malicious attacks this way, k-anonymity gives security assurance by ensuring and updates to the data.. Utilizing an execution running on DB2, They additionally demonstrate that the execution of the algorithms takes into consideration their utilization in real world appl1ications.

The real imperfection is that, it ought not to clarify how the knowledge about the schema and watermark will be given to the next client and not certain, how the data owner will identify the criticality of the data to be changed.

Their outcomes can be utilized as the premise for a lineage tracking tool in a general warehousing setting, furthermore can guide the design of warehouses that enable efficient lineage tracing. The major flaw is that it ought not concentrate on latest tools which will take care of this sort of issue automatically and there is no reasonable clarification is given at its security portion of this method.

4: Achieving K-Anonymity Privacy Protection

[7] Latanya Sweeney discussed about speculation and suppression techniques to shield the information from the data distributors utilizing k-anonymity privacy protection. The information in the system is dissected for speculation like Content (music, video, and picture) distributed commercial ventures, since technologies or techniques that could be utilized to ensure Intellectual property rights for computerized media, and prevent unauthorized copying did not exist.

Recoding a value with a less specific yet semantically consistent values and concealment includes not releasing a value at all. It accomplishes that the released records stick to k-anonymity, which implies each released record has at any rate (k-1) different records in the release whose values are indistinct over those fields that appear in external data . In that each released record will relate with in any event k-individuals regardless of the possibility that the Records are straightforwardly connected to outside data.

The favored Minimal Generalization Algorithm (MinGen), which is a theoretical algorithm which is presented here, consolidates these methods to provide k-anonymity protection with negligible distortion. This presents the real world algorithms Datafly and m-Argus is contrasted with MinGen. Both Data fly and m-Argus use heuristics to make approximations, thus they don't generally yield ideal results. It is demonstrated that Data fly can over distort information and m-Argus can’t provide adequate protection to the records which are stored.

Mainly, it focused towards suppression technique which is only it should not give the information to the client. The real Disadvantage in this framework is that, there is no reasonable clarification on, how the information will be secured in suppression technique. The another issue is that, while considering the data when it is not semantically linked then the suppression technique will not be effective.

Conclusion:

There are many leakage and lineage identification techniques and there by proposing a multi-edge approach which handles the various issues were all studied in detailed.

At the point when the event of handover confidential information happens it should always watermarks each object so that it could ready to follow its inceptions with absolute certainty, however certain information can't admit watermarks then it is conceivable to evaluate the probability that an user is responsible for a leak, based on the overlap of the data with the leaked data furthermore taking into account the likelihood that objects can be identified by other methodologies.

References:

[1] R.Sion, M.Atallah, and S.Prabhakar, ―Rights Protection for Relational Data, ‖Proc. ACMSIGMOD,pp.98-109,2003.

[2] R.Agrawal and J.Kiernan, ―Watermarking relational databases‖.InVLDB'02: Proceedings of the 28th international conference on Very Large Data Bases, pages 155–166.VLDB Endowment,2002.

[3] Hartung and kutter, watermarking technique for multimedia data‖2003.

[4] I. J. Cox ; NEC Res. Inst., Princeton, NJ, USA ;
J. Kilian ; F. T. Leighton ; T. Shamoon.

[5] L.Sweeney, ― Achieving K-Anonymity Privacy Protection Using Generalization And Suppression, http://en.scientificcommons.org/43196131,2002.

[6] Edward P.Holden, Jai W.Kang, Geoffrey R.Anderson, Dianne P.Bills, Databases in the Cloud: A Work in Progress,2012.

[7] Y.Cui and J.Widom. Lineage tracing for general data warehouse transformations. In The VLDB Journal, pages 471–480, 2001.