International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 03 | June-2015 www.irjet.net p-ISSN: 2395-0072

Augmenting Data Warehouse Security Techniques - A Selective Survey

Smita Konda, Rohini More

[1] *Assistant Professor, Computer Science and Engineering Department, A.G.Patil Institute of Technology, Maharashtra, India*

[2] *Assistant Professor, Computer Science and Engineering Department ,A.G.Patil Institute of Technology, Maharashtra, India*

Abstract - ***Due to increasing use of Internet, large amount of data is stored in large databases which are data warehouses. Mainly in data warehouses confidential data of business or financial related is stored. So problem of maintaining security is a major concern. A broad survey is carried out in this paper which shows all the security techniques used till now. The data warehouse security is to achieve the entire system consistently from sources and their Export tables, to ware house stored tables and views defined over the warehouse. Certain problems may arise in data warehouse like data inconsistency, confidentiality, and slow throughput. So to overcome those problems we have discussed various data security techniques. An integrated data is provided from heterogeneous environment to analysts where they can have greater access to huge amount of data. Because of wide availability of massive amount of data, and to deliver a way to assimilate meaningful data from multiple sources, data warehouse has become a requisite for every organization.***

Key Words: *Data warehouse, Data Security, Data Encryption, Data Masking, OLAP*

1. INTRODUCTION

Formal Definition: "A data warehouse is a subject-oriented i.e. Store data regarding total Sales, Number of Customers, etc. and not general data on everyday operations., integrated, time variant and non-volatile collection of data in support of management decision making process."

Data warehouse are used by executive management, business analyst only. Data stored may not be current but varies with time and data have an element of time. Data warehousing systems helps enterprise managers to obtain and incorporate information from heterogeneous sources and to query very large databases efficiently. Advanced Encryption Standard technique provides strong encryption but encryption involves extra storage space of encrypted data and overhead in query response time.

Data Warehouse store substantial amounts of credit card numbers, organization secrets, financial information and other personal information which make it major target for attackers who desire access to their valuable data.[2]. There are chances for opponent to easily gain access to consolidated data in the data warehouse. So securing data in data warehouse is very challenging. There are various techniques to secure data in data warehouse-

Data Masking: Data masking or data obfuscations the process of hiding original data with random characters or data.

Basically applying masking to a data field is to protect data that is classified as personal identifiable data, personal sensitive data or commercially sensitive data.

Data Encryption: In cryptography, encryption is the process of encoding information in such a way that only authorized parties can read it.

Encryption does not prevent interception, but denies the message content to the interceptor. In an encryption, the plaintext is encrypted using an encryption algorithm, to produce cipher text that can be decrypted.

A legal user can easily decrypt the message with the key provided by the originator to recipients, but not to unauthorized interceptors. The purpose of encryption is to protect the integrity, confidentiality of digital data stored on computer systems or transmitted via the Internet or other computer networks. Modern encryption algorithms play an essential role in the security assurance of IT systems and communications as they can provide not only confidentiality, but also the following key elements of security: Authentication, Integrity, Non-repudiation.

2. LITERATURE SURVEY

Mr. Dishek Mankad & Mr. Preyash Dholakia et al. [1] proposed various approaches to the data ware house design and usage process and the steps involved.

Top-down & bottom-down approaches or combination of

both can be used to construct data warehouse.

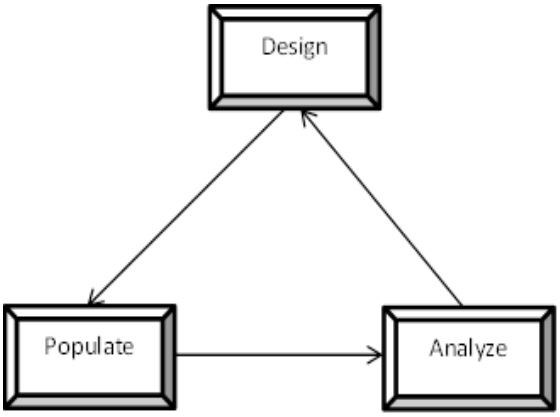

Fig -1: Data Warehouse Design Process

Data warehouse design process includes various steps as follows:

1. Choose a Business Process to Model
2. Choose the business process gain
3. Choose the dimension
4. Choose the measures

Data is extracted periodically from the applications and copied onto special computers. There it can be authenticated, reformatted, restructured, summarized, and enhanced with data from other sources.

Amritpal Singh and Nitin Umesh et al. [2] discussed solutions for securing data in data warehouse based on log implementation. Data Warehouse store huge amounts of business data, credit card numbers, secrets, and other personal information an opponent can easily access their valuable data. So, data stored in data warehouse need to be transformed to other form which should be unreadable by attacker. So to increase data security the authors have proposed a technique called data masking. Data Masking is a process to convert original data to some other form. For this MOD function/operator is used in SQL. Whenever the user sends request stored in log, so succeeding time when a query is sent, it is verified in the log and resend.

Figure 2 shows the basic entities in System Architecture. Firstly, Masked Database where all the data are encrypted in some unreadable format for security purpose. Second Modulous Based Data Masking Technique (MOBAT) which applies formula to mask the data by using MOD.

For example, consider 'Instructor' table where column instr_id is to be masked, it is carried this way,

ALTER TABLE Instructor ADD COLUMN K3

Let K1 and K2 be private keys, K3 is public key.

UPDATE Instructor SET instr_id = instr_id - K3 MOD K1MOD K2 +K2

For unmasking,

SELECT instr_id – K3 MOD K1 MOD K2 + K2 FROM Instructor

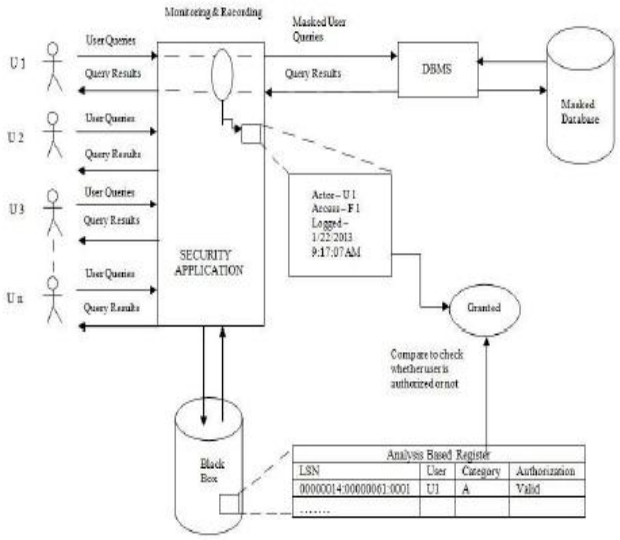

Fig -2: System Architecture

Akanksha, Akanksha, Rakheja, Ajay Singh et al. [3] found out another solution for securing data warehouse by using Model Driven Architecture where automatic transformations is carried out between models. Also, the Data Warehouse Striping (DWS) technique is a round-robin data partitioning approach especially designed for affordable data warehousing environments uses data encryption, spurious data, signatures, and redundancy to guarantee full data protection even when an attacker gets administrative access to one or more cluster nodes. The second technique used is DWS which uses round robin technique and partitions the data like star schema, fact tables to several other nodes in the network. Hence it achieves parallel query processing and load balancing for disks. OLAP queries are executed by all nodes in DWS cluster.

Slemo Warigon, et al.[4] proposed 7 phases for achieving stronger security in data warehouse, as it consists of very private, sensitive data.

Phase 1- In this phase the collected data is identified & data monitoring is done about all databases in the form of columns, rows & tables. Then data is properly structured for next phase.

Phase 2- In this phase data is classified according to sensitivity level of data i.e. low sensitive data, medium sensitive data & high sensitive data in order to achieve data confidentiality & integrity.

Phase 3- In this phase value of the data is given & various protective measures are used. If the value of data is very high then it consists of very confidential information & such information must be secured by the organization. For that purpose high risk factor is given to the data. The risk factor is nothing but probability of occurrence of attack to particular unit of data. If the probability of occurrence of attack is high then high risk factor is set to the data.

Phase 4- In this phase, some of the susceptibilities of data warehouse have been recognized like database DBMS limitations, dual security engines, inference attacks, insider threats.

Phase 5- In this phase, to overcome various problems found in previous phase, various cost-effective protective measures are used namely, Access control, Integrity controls , data encryption, partitioning.

Phase 6- In this phase, results of previous phase is evaluated. Cost-effective security measures are selected to protect the data against known vulnerabilities. In this 2 common techniques are used.
1. Economy of mechanism- principle edicts that a simple, well tested protective measure can be relied upon to control numerous susceptibilities in the DW.
2. Adequate data protection- edicts that the DW data can be secured with security measures that are effective and efficient enough for the short period of the data.

Phase 7- In this phase, Effectiveness of security measures is assessed to check whether security measures are: a) small, simple and straightforward, b) carefully examined, verified and proved, c) reasonably proficient in terms of time, memory space, and user-centric activities so that they do not harmfully disturb the secured computing properties. It is essential to confirm that the DW end-users realize the correctness of security measures. The data warehouse administrator (DWA) with the substitute authority from senior management is responsible for guaranteeing the effectiveness of security measures.

Dr. S.L. Gupta, Sonali Mathur, Palak Modi et al. [5] proposed a vision on some of the susceptibilities present in the data warehouse along with several security models and approaches to make the data warehouse secure.
Various security models have been proposed in order to secure data in data warehouse. As the data present is very large amount, the access is restricted. To restrict some of the access controls are used namely,

- Mandatory Access Control (MAC): This type of access control is managed by the administrator only. If any user tries to access the resources, he is first authenticated by verifying his category and classification. If these match then only the user is given access to the resources.
- Discretionary Access Control (DAC): In this a list of access control is provided where the user can decide to which user the permission can be given. As the whole authority is given to the user to decide which user can have access control, it's difficult for DAC policies to prevent from untrusted programs. This causes a limitation.
- Role Based Access Control (RBAC): In this the access control is given to individual user or group of users according to their role in the organization.
  As same permission is given to all users having same role, it is not possible for a single user to change the permission if he wants.
- Rule Based Access Control: In this some set of rules are defined. And according to the rules the access permission are allotted.

N. Katic, G. Quirchmayr, J. Schiefer, Stolba, M., Min Tjoa, A. et al. [6] introduced a metadata driven approach as a part of WWW-EIS-DWH project This protocol is implemented for technical realisation & not disclosed to use in different security policies. The main purpose of this paper is to confer requirements and decision about choice of correct security model for a data warehouse environment. This security model supports various features. This model provides access permissions to individual data items, selective encryption and original security processes. For the correct selection of the security model it is very necessary to focus on metadata of data warehouse. It consists of security details such as classifications of security objects or clearances of security subjects & access rules.
Data Warehouse & Security- The data warehouse is an essential part of the decision support system and does not usually entail data updating. It provides various benefits to an organization certain like data sharing, enabling staff to successfully and powerfully resolve active organizational problems, reducing operating costs and increasing revenue, managing market shares & reduce the employee turnover. It supports various mechanisms to support availability, confidentiality & integrity.

Availability- It ensures data will be available to legal users whenever they required.
Confidentiality- It ensures data is secured from illegal access.
Integrity-It means data should be protected from malicious alteration including devastation of data, corruption of data, and addition of wrong data.
Metadata & Security-A crucial thing about data warehouse are metadata -data about the data enclosed within the data warehouse. With no metadata, locating information present in the data warehouse becomes a difficult task. Metadata explains the contents of the data warehouse. Metadata plays very important role in the creation of operational data stores, multiple data marts & integrated data warehouses. Metadata repository management software is used to convert source data to destination database, produce code for data transformations, combine and convert the data, and manage moving data to the warehouse. Metadata can also explain protection mechanisms in a data warehouse environment.
Metadata Security Model- Metadata are very crucial part of data warehouse. For creating & maintaining data warehouse metadata are used by developers.
There are 2 types of metadata:

1. Structural metadata
2. Access metadata

Structural metadata- For creating & maintaining data warehouse structural metadata are used. It illustrates its structure & its contents. The fundamental constituent of structural metadata is a model which depicts their features ,their intermediate relationships & their data subjects.
Access metadata- To represent dynamic relationship between the data warehouse & end user applications access metadata are used. It consists of user defined names & aliases, measured values of an enterprise. These data consists of the explanation of the databases, tables, data warehouse servers thorough data and summarized data . Such metadata permit the security realization for an individual user, user groups or the entire enterprise, relating reading, modifying etc. of calculations, summed up data or analyses.
The M-View environment:
In this environment, there are two main components of this model.

1. Security manager
2. The Secure Query Management Layer (SQML)

The Security Manager is responsible for giving access rights to the users. He first identifies the user who tries to access the data and allows him to access only that part of data which he has requested. But the other part of data in data warehouse is kept hidden. Security Manager adds new user or deletes any user on request. This can be achieved by SQML.
Secure Query Management Layer (SQML) is a layer that accepts only valid queries. It means that all the queries coming from internet are filtered at this layer and takes care that only query is belonging to allocated data, not the unallocated data. If such type of queries arise then SQML throws or does not allow such queries into data warehouse. This really helps to build a strong security against data in data warehouse.

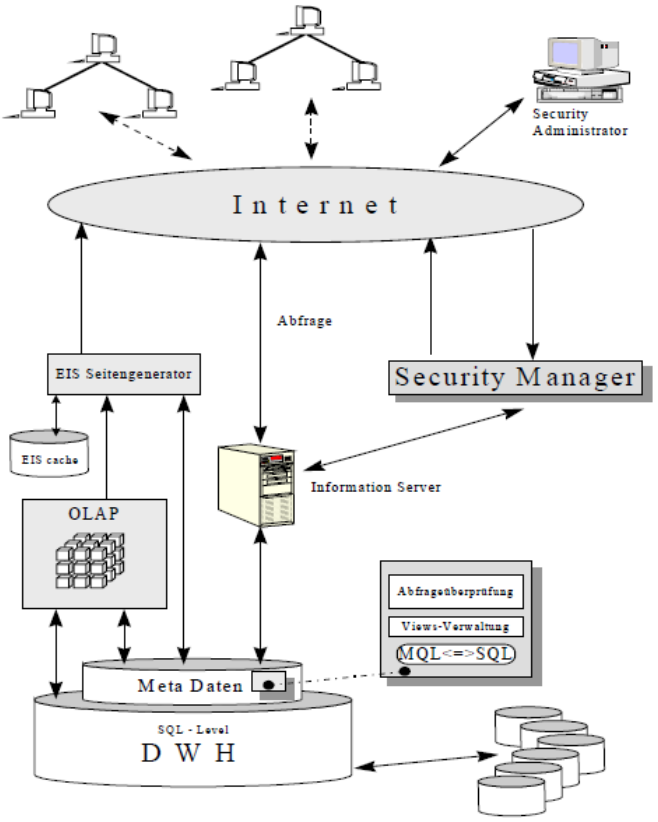

Fig 2: M- View Structure

This security model acts as a boundary to limit the scope of user queries to data that are not hazardous to the security rules of the information system. Due the Secure Query Management Layer (SQML) the user can access data limited to the predefined amount of data.

3. CONCLUSIONS

In this paper we have conferred several different security techniques to protect the sensitive data stored in data warehouse. A highly increased usability in terms of security is offered by data warehouses and OLAP.
A basic authorization model for data warehouses and OLAP offering greater expressiveness and highly increased usability with respect to security. Various models discussed considering security aspects like integrity, confidentiality and availability have been achieved. New

SQML model helps selected queries to access the data which is allowed to. Many other advanced techniques will be developed in future.

REFERENCES

[1] Mr. Dishek Mankad, Mr. Preyash Dholakia, "The Study on Data Warehouse Design and Usage," International Journal of Scientific and Research Publications, Volume 3, Issue 3, March 2013.

[2] Amritpal Singh and Nitin Umesh, "Implementing Log Based Security in Data Warehouse," International Journal of Advanced Computer Research Volume-3 Number-1 Issue-8 March-2013.

[3] Akanksha, Akanksha, Rakheja, Ajay Singh, "Data Warehouse Security," International Journal of Research in Engineering & Advanced Technology, Vol 1, Issue 5, Oct-Nov 2013.

[4] Slemo Warigon, "Data Warehouse Control And Security," Association of College and University Auditors LEDGER, Vol.41, No.2, April 1997; pp.3-7.

[5] Dr. S.L. Gupta, Sonali Mathur, Palak Modi, "Data Warehouse Vulnerability and Security," International Journal of Scientific and Research, Volume 3, Issue 5, May -2012.

[6] N. Katic, G. Quirchmayr, J. Schiefer, Stolba, M., Min Tjoa, A., "A Prototype Model for Data Warehouse Security Based on Metadata," in 9th Int. Workshop on Database and Expert Systems Applications,Vienna. IEEE Computer Society (1998).

BIOGRAPHIES

Mrs. Smita S. Konda working as Asst. Professor in AGPIT, Solapur. She has completed her BE in INFORMATION TECHNOLOGY from WALCHAND INSTITUTE OF TECHNOLOGY, SOLAPUR from Solapur University. She has completed her M.Tech in Computer Science and Engineering from JNTU University,Hyderabad. She has 5 years of teaching experience.

Ms. Rohini S. More working as Asst Professor in AGPIT, Solapur. She has completed her BE in Computer Science and Engineering from Bharat Ratna Indira Gandhi College of Engineering, Solapur from Solapur University. She has completed her M.Tech in Computer Science and Engineering from JNTU University, Hyderabad. She has 3.5 years of teaching experience.

