International Journal of Soft Computing 12 (2): 112-119, 2017
ISSN: 1816-9503

UML Modeling of Securing Sensitive Data by Inference Control Method

[1]Anurag, [2]Deepak Arora and [3]Upendra Kumar
[1]Amity Institute of Information Technology,
[2]Department of Computer Science and Engineering,
Amity University, Lucknow Campus, Noida, India
[3]Department of Computer Science and Engineering, Birla Institute of Technology,
Patna Campus, Noida, India

Abstract: With the advancement of internet and data technologies, the information explosion can be seen everywhere in the industry. The huge corpus of statistical data are periodically generated, updated and maintained by various organizations. It has evolved not only as a speed of exponential growth of data but as an impressive tool to understand the deep insights of any business future planning and market strategy. Various web repositories are being employed to sustain this data for further findings of different business patterns which could be very important and private for a particular industry. This sensitive data when exposed to legitimate but untrustworthy owners may possess the grave risks to the owner's privacy. Secure access of organization database during mining process has been a major challenge of the database designers which consequences in evolution the new research disciple privacy preserving data mining. It deals with preventing the confidential information of an individual or entity from inferring by the malicious data miners during the mining process. In this study, researchers have presented an object oriented modeling of the proposed system by applying inference control techniques based on unified modeling language. Different UML diagrams like, class, sequence and activity diagram have been designed in this research work for modeling proposed PPDM system. Its generalized UML model will act as a prototype for the software engineers to conceive and develop the protocol embeds in the complicated privacy preserving data mining systems. This object oriented UML modeling will ensure successful consummation of any large and complex PPDM software projects in both optimal time and cost efficient manner.

Key words: Object oriented UML inference control approach, privacy preserving data mining, UML modeling, object oriented modeling, query based inference, perturbation based inference

INTRODUCTION

With the purpose to insight hidden patterns and trends from the vast datasets and exploiting it in various fields such as in scientific discovery, web search, digital library, etc. (Xu *et al*., 2014) the scope of the data mining has increased tremendously in recent decades. The huge corpus of data of a particular organization has to be exposed to numerous untrustworthy parties during mining process. The intention to prevent its misuse from the unauthorized owners causes the evolution of new research area called privacy preserving data mining which involves concealing the exposure of true sensitive data as well as information from unauthorized owners during mining process and preserving data utility. Various techniques have proposed by the researchers in this area for each suitable and specific circumstances and applications in which inference control is one such well known approach (Thuraisingham *et al*., 1993; Zhang and Zhao, 2007). Its major application is in the field of the statistical database. Statistical database gives the aggregate (sum, max, min, etc.) information about the entity sets, rather revealing accurate information about particular individual or entity during query response. Field such as credit ratings, test-score and medical datasets which uses aggregate data relies on inference based approach for concealing personal sensitive data. It has attracted the attention of researchers throughout the world since 1970's and is still a continuously evolving field. Inference simply means the process of deducing useful information from the given set of premises (Zhang and Zhao, 2007). It is the technique ofderiving the useful information on the basis of certain logic and reasoning. The set of successive query responses infers by the malicious adversary could reveal sensitive data or information. So, inference control protocol deals with

Corresponding Author: Anurag, Amity Institute of Information Technology, Amity University, Lucknow Campus, Noida, India

preventing the sensitive information from disclosure while query handling. The queries sets are checked priory before replying so that it could not reveal any sensitive information (Zhang and Zhao, 2007).

Modeling plays a vital role in overall software designing and also towards verifying and optimizing the entire software development process to a more robust and with less complexity. It helps the analysts in understanding the functions and behavior of the entire system and makes the requirement analysis tasks easier and performs in systematic manner (Pressman, 2005).

Object oriented system development has been gaining popularity since decades because of its robustness in all fields of development phase (Sommerville, 2004). It has various indispensible features such as minimizing development costs (Waman, 2009) easing system design, saving development time and minimizing development costs. UML is a popular global standard notion for expressing Object Oriented solutions. It describes numerous design options and documenting design artifacts (Lawrence, 2003) thus helpful in understanding and designing of entire system (Jalote, 1997). Researchers model the entire system through rational software architecture as it provides a modeling environment which uses UML for designing architecture for object oriented applications and web based services. It is a tool for complete software delivery and provides the fastest way for clients to become familiar with the entire system.

Literature review: Research work presented in this research study focuses on UML modeling of inference control approach in PPDM systems. Through massive literature survey it has been found that very little research work has been proposed in this area. Most of the studys very specific to exploring different approaches related to PPDM systems. Inference control approach is applied for the protection against tabular data as well as micro data. Software for the inference control approach has also been developed. Inference control approach is applied for the protection against tabular data as well as micro data.

Accorsi and Muller (2013) deal with protecting inference in data centric business model. Geetha and Rebacca describe Private Inference Queries for aggregated database (Jagannathan and Wright, 2007). Client learns the value of the particular queries only if a query passes a specific inference control rules. Server infers nothing about queries and client learns nothing other than the query output for each input query. Federica and Nicola present an access control method for preventing inference of sensitive information. The

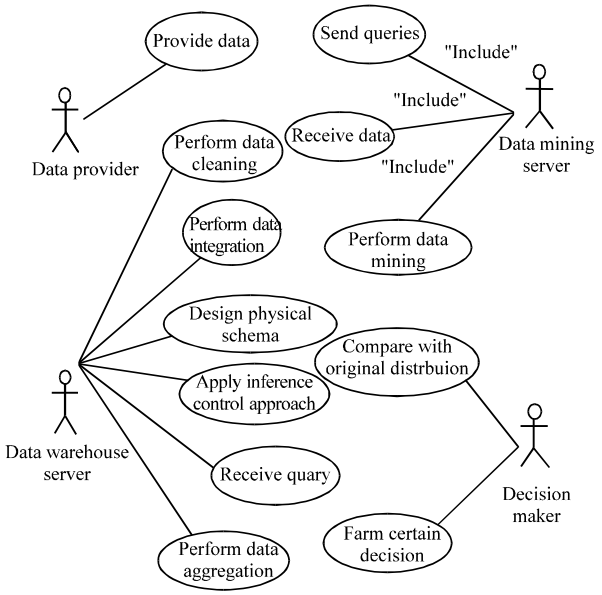

Fig. 1: Use case diagram of PPDM system

semantic approach is based on data model that encode domain knowledge. It organizes the data model into hierarchical structure that makes explicit inference relation between data (Paci and Zannone, 2015). Jessica deals with applying inference control approach on multilevel database. As the knowledge of some objects allows the information of the high security level to be inferred, tokens are associated with every object for enabling access. The same tokens cannot be used to query another object in inference channel. Hence, it facilitates flexible information access as the same tokens associated with each object. Secure inference control approach on provenance has been introduced by Bhavani. As vast quantities of data and information are available on the web, it maintains the data history so that malicious operation on it could be tracked and detected (Thuraisingham *et al*., 2014). As the adversary having high domain knowldge and skills could exploit inference vulnerability, Muhamed in deals with the situation in which adversary profile is deployed on the database for security (Turkanovic *et al*., 2014).

UML MODELING OF PROPOSED PPDM SYSTEM

Inference based control involves various stakeholders. Use case diagram has been drawn for depicting important scenario and the cases of the system and interaction of different actors within the system for performing different operations. Figure 1 illustrates the use case diagram of privacy preserving system. The major

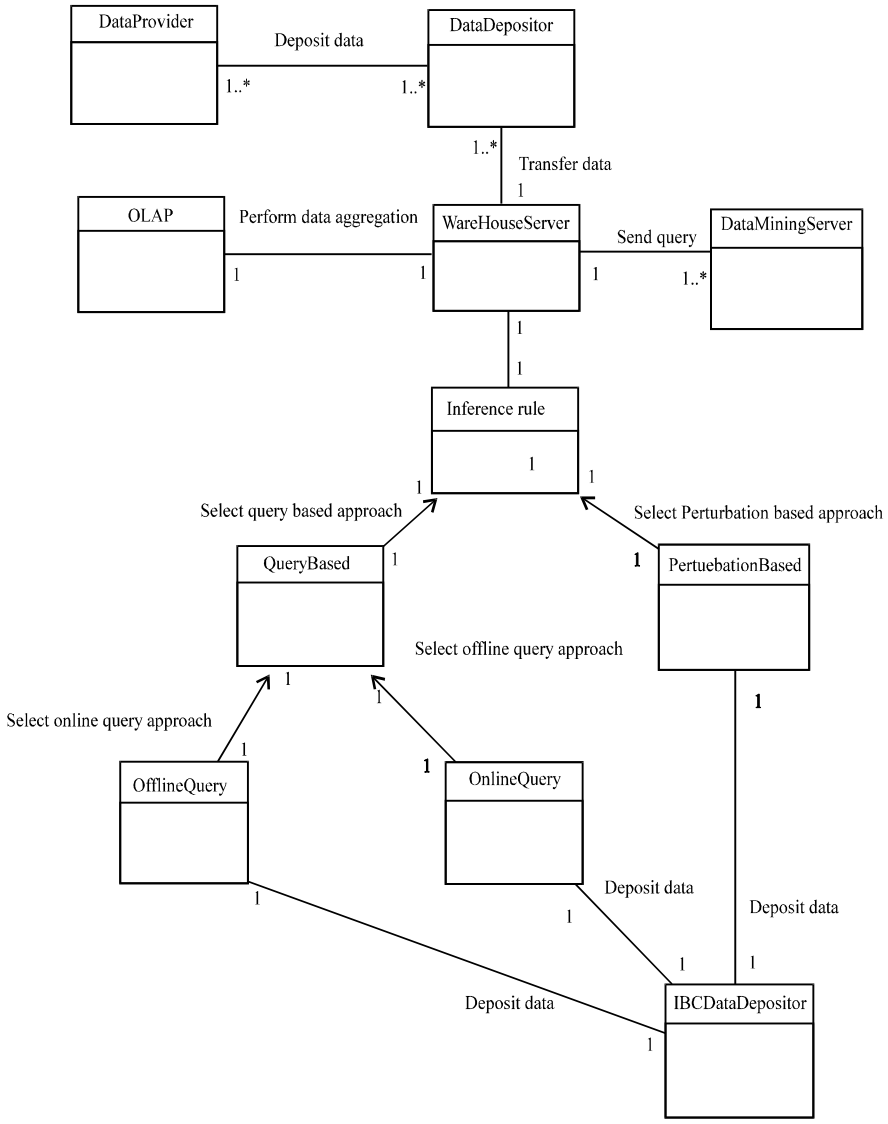

Fig. 2: Class diagram of PPDM system with inference control approach

stakeholders have been identified for the system-data provider, data warehouse server, data mining server and decision makers. Data provider has the responsibility of providing data to the privacy preserving system. Data Warehouse Server will perform the tasks of data preprocessing, receiving and analyzing the queries and applying specific inference based control approach for privacy preserving, converting data into aggregate form and then sends the aggregated data to the data mining server.

Data mining server performs the tasks of sending the data to warehouse server, receiving aggregated data and applying data mining operations on the received data. Figure 2 shows the generalized class diagram of the inference based control approach applied in the system. Different classes each with some specific roles and responsibilitties have been identified. Class DataProvider has the responsibility of depositing data to the DataDepositor class. DataDepositor transfers the data to the class WarehouseServer. The Class WarehouseServer performs data preprocessing, design physical schema. The data then send to OLAP class for converting in to aggregated form. After that the data storage to the warehouse server will be handled by WarehouseDataDepositor class. ClassDataMiningServer has the responsibility of sending query to the WarehouseServer class.

Class WarehouseServer after receiving query, starts fetching the appropriate data from the class WarehouserDataDepositor and further communicate with class InferenceRule to apply best suitable Inference Control approach. Class QueryBased has the

responsibility of applying query based approach on Warehouse Data. Class PerturbationBased has the responsibility of applying Inference Control based perturbation, and transfers it to the class DataMiningServer. Class DataMiningServer received aggregated data and applies data mining operations on it.

MAJOR ATTRIBUTES AND METHODS USED IN PROPOSED SYSTEM

Major classes with their attributes and methods defined for the proposed PPDM system are as follows.

Algorithm; attibutes and methods of each of these classes:

DataProvider
name: string
source id: int
Set_Connection()
Putdata()

DataDepositor
data: void
Getdata()

WareHouseServer
data: void
preprocess_data()
Physical_schema_design()
put data()
receives_query()
Getdata()

WareHouseDataDepositor
data: void
Get data()
Putdata()

QueryBased
data: void
Get data()
Cheak_safe_sequence()
accept_query(), reject_query()
sendQuery()

OnlineQuery
data: void
retrieve_query history Table()
update_query_history Table()
check_safe_sequence()

OfflineQuery
data: void
cheak_safe_sequence()

PerturbationBased
data: void
Get data(), perturb_query()
send_perturbed_query()

DataMiningServer
data: void
Get data()
Receive_aggregated_data()
Apply_data_minig()

MiningDataDepositor
data: void
Put data()

OLAP
data: void
get data()
put data()
data aggregation()

InferenceRule
set rule()
get rule()
apply rule()

UML ACTIVITY MODELING PROPOSED SYSTEM

Figure 3 illustrates the activity diagram of inference based control. Query-oriented approach is suitable for the situation when the construction of data mining models requires highly accurate query results. Query perturbation oriented approach is applicable for the situations in which preserving sensitive data is utmost important than accurate data model and queries rejections should not affect the mining model to significant extent (Thuraisingham *et al*., 1993).

Figure 4 shows the series of activities takes place in online query based approach applied in privacy preserving system. Class data mining server sends the query sets to the class data ware house server. Class data ware house server users transfer control to the class inference rule for applying suitable inference control approach. Class InferenceRule applies query based approach and transfers control to class online query. Class online query fetches the appropriate data after receiving the queries. It then accesses the query history table for checking the history of previous queries already answered. These queries along with currently received queries data are then checked for safe sequence. It accepts the queries when safety condition satisfies, otherwise reject the queries. The accepted query data are sent to class ware house data depositor. Figure 5 shows the series of activities takes place within Offline Query approach. Class InferenceRule applies Offline Query approach and the control transmits to the class Offline Queries are checked for safe sequence whether the queries received belongs to the subsets of query history.

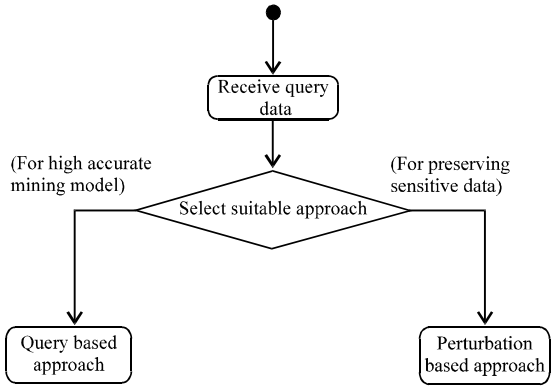

Fig. 3: Activity diagram of inference based selection

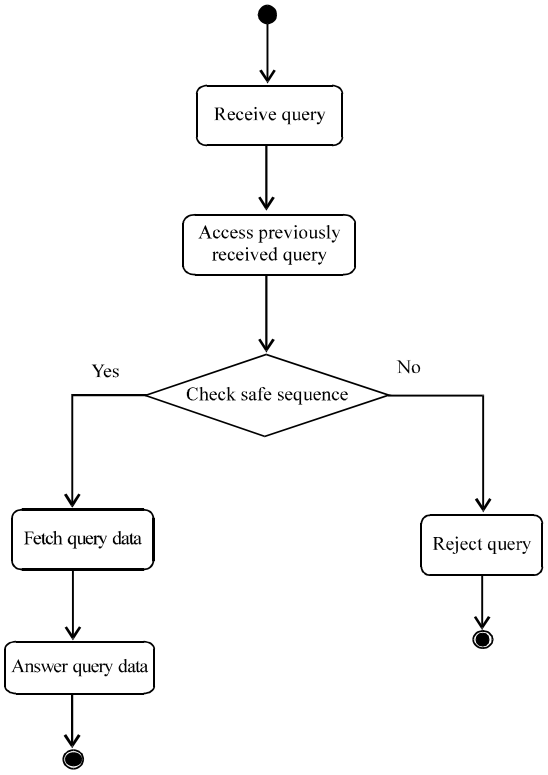

Fig. 4: Activity diagram for online query handling

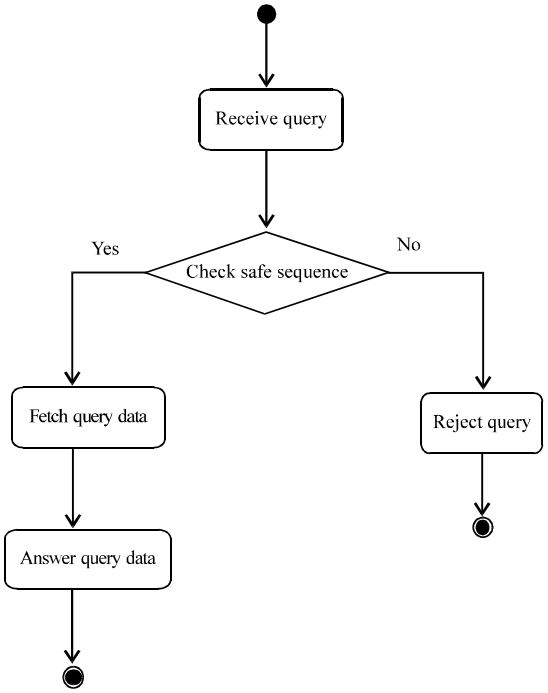

Fig. 5: Activity diagram for handling OfflineQuery

It fetches the query data if the safe sequence property satisfy otherwise reject the query. The query

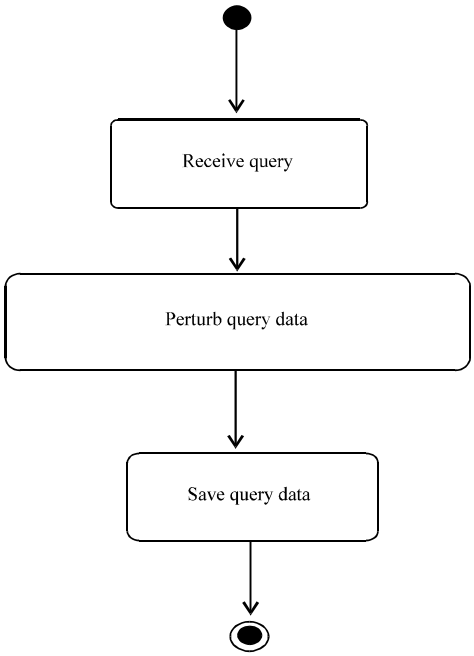

Fig. 6: Activity diagram of applying perturbation

data is then sent to DataMiningServer class. Figure 6 represents the series of activities takes place within Perturbation based approach. Class DataMiningServer sends the queries to the ware house server class. Class WareHouseServer selects query perturbation approach and transfers control to the class perturbationbased. Class PerturbationBased perturbs each received query data and transfers it to the class data mining server.

INTERACTIONS BETWEEN INFERENCE RULE OBJECT WITH OTHER SYSTEM OBJECTS

Figure 7 shows the sequence diagram of activities takes place between DataProvider and DataWarehouseServer. Class DataProvider saves its data in the DataDepositor class. The data from the DataDepositor is transmitted to the DataWarehouseServer for further preprocessing and then designed it into physical schematic form for making it suitable for mining. It transfer control to the OLAP class for performing data aggregation. The aggregated data is saved in the WarehouseDataDepositor class

Figure 8 shows the Sequence diagram between DataWarehouseServer, InferenceRule and DataMiningServer. Class DataMiningServer send queries to the Class DataWarehouseServer. Class DataWarehouseServer transfers control to the class InferenceRule for applying suitable Inference Control approach.

Class inferencerule transfer its control to the class query based which further transfers the control to the

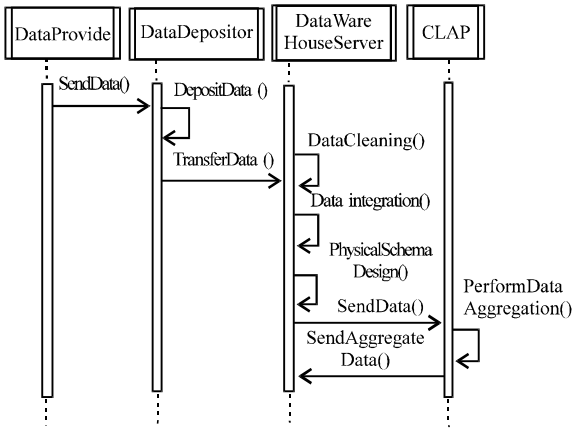

Fig. 7: Sequence diagram for processing of data among data provider and data warehouse server objects

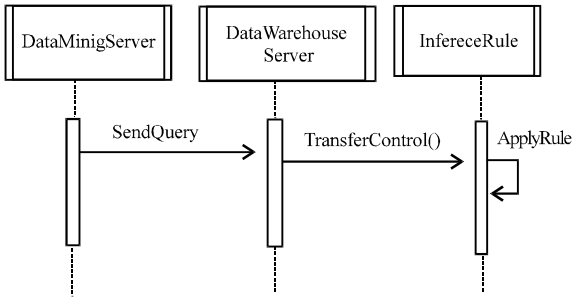

Fig. 8: Sequence diagram for applying inference rule

class OfflinQquery. Class OfflineQuery checks the safe sequence and accepts the query if condition satisfies, otherwise reject the query. The required data is fetched from the WarehouseData depositor class and transfer it to the class DataMiningServer.

Figure 11 shows the Sequence diagram of activities takes place within Online Query Approach. Class DataWarehouseServer after receiving queries from class Data MiningSever transfer its control to the Class InferenceRule. Class InferenceRule applies Online Query approach and the control is transferred to the class OnlineQuery. Class OnlineQuery access the Query History table and safe sequence is checked along with received queries and accepted if the safety condition satisfies.The query data is fetched from WarehouseDataDepositor class and sent it to the class DataMiningServer for further mining operations (Fig. 9 and 10).

Figure 11 shows the sequence diagram of activities takes place in query perturbation approach. Class DataMiningServer sends the queries to the class ware house server. Class WarehouseServer transfers the control to the InferenceRule class. It applies query

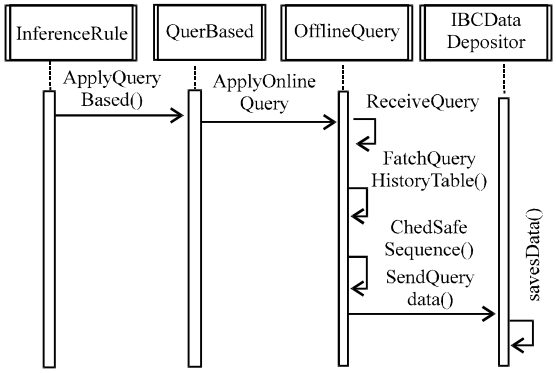

Fig. 9: UML sequence diagram of applying inference rule for OfflineQueries

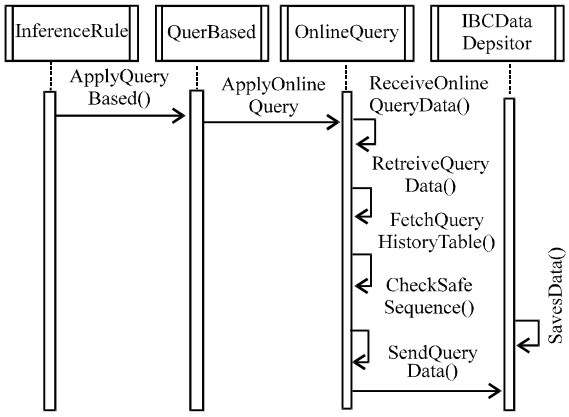

Fig. 10: Sequence diagram of applying inference rule on for OnlineQuery

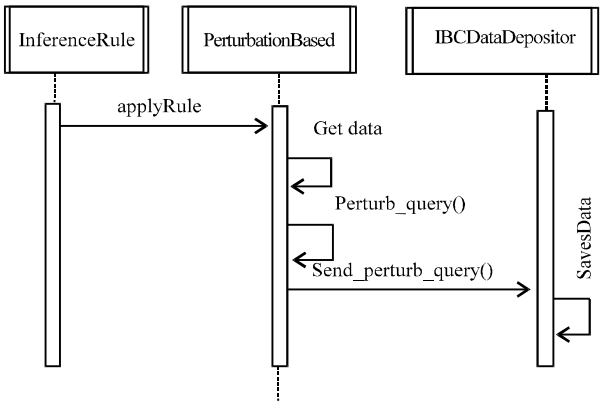

Fig. 11: Sequence diagram of applying inference rule for Perturbation

perturbation approach. Class PerturbationBased fetches the appropriate data from the WarehouseDataDepositor

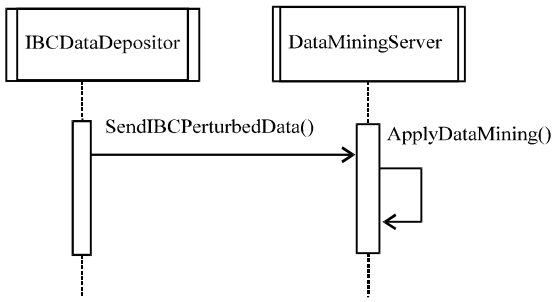

Fig. 12: Interaction among WarehouseServer and DataMiningServerObjects

class and perturb it. The Perturbed data is transferred to class DataMiningServer. Figure 12 shows the sequence of activities takes place between IBCDataDepositor and DataMiningSever class. Class DataMiningServer applies data mining algorithm on the received data. This interaction has been represented in Fig. 12. Interaction among warehouse server and data mining server objects.

CONCLUSION

This study represents the uml based modeling of inference based control approach in privacy preserving data mining system. The proposed system has been implemented and designed on ibm rational software architect tool. This object-oriented modeling of the PPDM system will facilitate the programmers to develop robust and efficient software applications for similar security systems for preserving and securing sensitive data for an organization. Different graphical uml diagrams have been designed and major classes and interactions among these classes also represented in this research work which provides high levels of abstraction. This research work not only enriches uml profiles but also presented a framework to develop such applications using any object oriented platform.

RECOMMENDATIONS

This research will further ease to trace different design complexities at different levels of software development of these systems at early phases of its development. The design presented in this work is also helpful in reducing overall costing and development effort of incurred towards any PPDM system development. Further this research can be extended towards more approaches of PPDM system development and securing sensitive business data.

ACKNOWLEDGEMENTS

The researchers are very grateful to respected Mr. Aseem Chauhan, Chairman, Amity University Lucknow Campus and Maj. Gen. K.K. Ohri, AVSM (Retd.) Pro-VC, Amity University, Uttar Pradesh, Lucknow Campus India, for providing excellent research infrastructure in the university campus. Researchers also pay their best regards to brig UK Chopra, Director Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow Campus, for giving their motivational support and help to carry out the present research work.

REFERENCES

Accorsi, R. and G. Muller, 2013. Preventive inference control in data-centric business models. Proceedings of the Workshops on Security and Privacy (SPW), May 23-24, 2013, IEEE, New York, USA., ISBN:978-1-4799-0458-7, pp: 28-33.

Jagannathan, G. and R.N. Wright, 2007. Private inference control for aggregate database queries. Proceedings of the 7th IEEE International Conference on Data Mining Workshops, October 28-31, 2007, IEEE, New York, USA., ISBN:978-0-7695-3019-2, pp: 711-716.

Jalote, P., 1997. An Integrated Approach to Software Engineering. 2nd Edn., Springer, Berlin, Heidelberg, New York, ISBN: 9780387948997, Pages: 497.

Lawrence, S., 2003. Software Engineering Theory and Practices. 2nd Edn., Pearson Education, Upper Saddle River, New Jersey, USA.

Paci, F. and N. Zannone, 2015. Preventing information inference in access control. Proceedings of the 20th ACM Symposium on Access Control Models and Technologies, June 1-3, 2015, ACM, New York, USA., ISBN:978-1-4503-3556-0, pp: 87-97.

Pressman, R.S., 2005. Software Engineering: A Practitioners Approach. 6th Edn., McGraw-Hill, Boston, ISBN: 978-0071238403, pp: 1-880.

Sommerville, I., 2004. Software Engineering. 6th Edn., Pearson Education, Upper Saddle River, New Jersey, USA.,.

Thuraisingham, B., T. Cadenhead, M. Kantarcioglu and V. Khadilkar, 2014. Secure Data Provenance and Inference Control with Semantic Web. CRC Press, Boca Raton, Florida, USA., Pages: 429.

Thuraisingham, B., W. Ford, M. Collins and J. O'Keeffe, 1993. Design and implementation of a database inference controller. Data Knowl. Eng., 11: 271-297.

Turkanovic, M., T.W. Druzovec and M. Holbl, 2015. Inference attacks and control on database structures. TEM. J., 4: 1-13.

Waman, S.J., 2009. Software Engineering Principal and Practices. McGraw-Hill, New York, USA.

Xu, L., C. Jiang, J. Wang, J. Yuan and Y. Ren, 2014. Information security in big data: Privacy and data mining. IEEE. Access, 2: 1149-1176.

Zhang, N. and W. Zhao, 2007. Privacy-preserving data mining systems. Comput., 40: 52-58.