International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

Secure Mining of Association Rules in Horizontally Distributed Databases

Jeevan Kumar Kalluru & Veeresh
1 M.Tech Student (SED), Sir Vishveshwaraiah Institute Of Science & Technology Hyderabad
. Email: kjeevan007@gmail.com
2Sr. Assistant Professor, Department of CSE, Sir Vishveshwaraiah Institute Of Science & Technology Hyderabad.

ABSTRACT

We propose a protocol for secure mining of association rules in horizontally distributed databases. Our protocol, like theirs, is based on the Fast Distributed Mining (FDM) algorithm which is an unsecured distributed version of the Apriority algorithm. The main ingredients in our protocol are two novel secure multi-party algorithms — one that computes the union of private subsets that each of the interacting players hold, and another that tests the inclusion of an element held by one player in a subset held by another. Our protocol offers enhanced privacy with respect to the protocol. In addition, it is simpler and is significantly more efficient in terms of communication rounds, communication cost and computational cost.

INTRODUCTION TO PROJECT

We propose a protocol for secure mining of association rules in horizontally distributed databases. Our protocol, like theirs, is based on the Fast Distributed Mining (FDM) algorithm which is an unsecured distributed version of the Apriority algorithm. The main ingredients in our protocol are two novel secure multi-party algorithms — one that computes the union of private subsets that each of the interacting players hold, and another that tests the inclusion of an element held by one player in a subset held by another. Our protocol offers enhanced privacy with respect to the protocol. In addition, it is simpler and is significantly more efficient in terms of communication rounds, communication cost and computational cost

Problems in existing system:-In Existing System, the problem of secure mining of association rules in horizontally partitioned databases. In that setting, there are several sites (or players) that hold homogeneous databases, i.e., databases that share the same schema but hold information on different entities. The inputs are the partial databases, and the required output is the list of association rules that hold in the united database with support and confidence no smaller.

Disadvantage: Less number of features in previous system. Difficulty to get accurate item set.

Solution to these problems:-In Proposed System, propose an alternative protocol for the secure computation of the union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol does not depend on commutative encryption and oblivious transfer (what simplexes it significantly and contributes towards much reduced communication and computational costs). While our solution is still not perfectly secure, it leaks excess information only to a small number (three) of possible coalitions, unlike the protocol of that discloses information also to some single players. In addition, we claim that the excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol.

Advantage: As a rising subject, data mining is playing an increasingly important role in the decision support activity of every walk of life. Get Efficient Item set result based on the customer request. Diversion from solution.

SYSTEM ANALYSIS

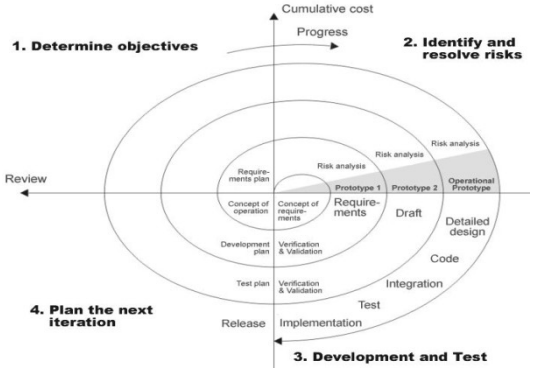
After analyzing the requirements of the task to be performed, the next step is to analyze the problem and understand its context. The first activity in the phase is studying the existing system and other is to understand the requirements and domain of the new system. Both the activities are equally important, but the first activity serves as a basis of giving the functional specifications and then successful design of the proposed system. Understanding the properties and requirements of a new system is more difficult and requires creative thinking and understanding of existing running system is also difficult, improper understanding of present system can lead diversion from solution.

Analysis model sdlc methodologies:-This document play a vital role in the development of life cycle (SDLC) as it describes the complete requirement of the system. It means for use by developers and will be the basic during testing phase. Any changes made to the requirements in the future will have to go through formal change approval process.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

Spiral model: - was defined by Barry Boehm in his 1988 article, "A spiral Model of Software Development and Enhancement. This model was not the first model to discuss iterative development, but it was the first model to explain why the iteration models. As originally envisioned, the iterations were typically 6 months to 2 years long. Each phase starts with a design goal and ends with a client reviewing the progress thus far. Analysis and engineering efforts are applied at each phase of the project, with an eye toward the end goal of the project. The steps for Spiral Model can be generalized as follows: The new system requirements are defined in as much details as possible. This usually involves interviewing a number of users representing all the external or internal users and other aspects of the existing system. A preliminary design is created for the new system. A first prototype of the new system is constructed from the preliminary design. This is usually a scaled-down system, and represents an approximation of the characteristics of the final product. A second prototype is evolved by a fourfold procedure: Evaluating the first prototype in terms of its strengths, weakness, and risks. Defining the requirements of the second prototype. Planning an designing the second prototype. Constructing and testing the second prototype. At the customer option, the entire project can be aborted if the risk is deemed too great. Risk factors might involve development cost overruns, operating-cost miscalculation, or any other factor that could, in the customer's judgment, result in a less-than-satisfactory final product. The existing prototype is evaluated in the same manner as was the previous prototype, and if necessary, another prototype is developed from it according to the fourfold procedure outlined above. The preceding steps are iterated until the customer is satisfied that the refined prototype represents the final product desired. The final system is constructed, based on the refined prototype. The final system is thoroughly evaluated and tested. Routine maintenance is carried on a continuing basis to prevent large scale failures and to minimize down time.

The following diagram shows how a spiral model acts like:-



Spiral Model

Study of the system:-In the flexibility of the uses the interface has been developed a graphics concept in mind, associated through a browser interface. The GUI'S at the top level have been categorized as. The administrative user interface concentrates on the consistent information that is practically, part of the organizational activities and which needs proper authentication for the data collection. The interfaces help the administrations with all the transactional states like Data insertion, Data deletion and Data updating along with the extensive data search capabilities. The operational or generic user interface helps the users upon the system in transactions through the existing data and required services. The operational user interface also helps the ordinary users in managing their own information helps the ordinary users in managing their own information in a customized manner as per the assisted flexibilities

Proposed system:-In Proposed System, propose an alternative protocol for the secure computation of the union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol does not depend on commutative encryption and oblivious transfer (what simplexes it significantly and contributes towards much reduced communication and computational costs). While our solution is still not perfectly secure, it leaks excess information only to a small number (three) of possible coalitions, unlike the protocol of that discloses information also to some single players. In addition, we claim that the excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol.

Modules

1. User Module.
2. Admin Module.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

3. Association Rule.
4. Apriority Algorithm.

Modules Description

User Module:- In this module, privacy preserving data mining has considered two related settings. One, in which the data owner and the data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the united corpus of data that they hold.

Admin Module:-In this module, is used to view user details. Admin to view the item set based on the user processing details using association role with Apriority algorithm.

Association Rule:-Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk."Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true.

Apriority Algorithm:-Apriority is designed to operate on databases containing transactions. The purpose of the Apriority Algorithm is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriority is sets of rules that tell us how often items are contained in sets of data.

SYSTEM SPECIFICATION

Hardware Requirements:

System	:	Pentium IV 2.4 GHz.
Hard Disk	:	40 GB.
Floppy Drive	:	1.44 Mb.
Monitor	:	14' Colour Monitor.
Mouse	:	Optical Mouse.
Ram	:	512 Mb.
Keyboard	:	101 Keyboards.

Software Requirements:

Operating system	:	Windows 7 Ultimate (32-bit)
Front End	:	VS2010
Coding Language	:	ASP.Net with C#
Data Base	:	SQL Server 2008

INPUT AND OUTPUT DESIGN

Input design:-The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

Output design:-A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

FEASIBILITY REPORT

Feasibility study: - The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential. Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

Economical feasibility:-This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical feasibility:-This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social feasibility:-The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

SOFTWAREREQUIREMENTSPECIFICATION

Scope of the project:-The software, Site Explorer is designed for management of web sites from a remote location.

Purpose: The main purpose for preparing this document is to give a general insight into the analysis and requirements of the existing system or situation and for determining the operating characteristics of the system.

Scope: This Document plays a vital role in the development life cycle (SDLC) and it describes the complete requirement of the system. It is meant for use by the developers and will be the basic during testing phase. Any changes made to the requirements in the future will have to go through formal change approval process.

Functional Requirements:-refer to very important system requirements in a software engineering process (or at micro level, a sub part of requirement engineering) such as technical specifications, system design parameters and guidelines, data manipulation, data processing and calculation modules etc. Functional Requirements are in contrast to other software design requirements referred to as Non-Functional Requirements which are primarily based on parameters of system performance, software quality attributes, reliability and security, cost, constraints in design/implementation etc. The key goal of determining “functional requirements” in a software product design and implementation is to capture the required behavior of a software system in terms of functionality and the technology implementation of the business processes.

Non nonfunctionalrequirements:- Alltheotherrequirements which do not form a part of the above specification are categorized as Non-Functional Requirements. A system may be required to present the user with a display of the number of records in a database. This is a functional requirement. How up-to-date this number needs to be is a non-functional requirement. If the number needs to be updated in real time, the system architects must ensure that the system is capable of updating the displayed record count within an acceptably short interval of the number of records changing. Sufficient network bandwidth may also be a non-functional requirement of a system.

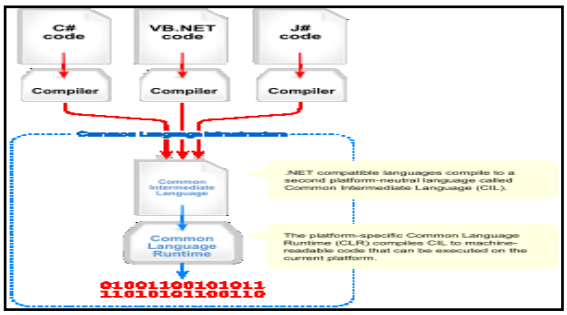
INTRODUCTIONTO.NETFRAMEWORK

The **Microsoft .NET Framework** is a software technology that is available with several Microsoft Windows operating systems. It includes a large library of pre-coded solutions to common programming problems and a virtual machine that manages the execution of programs written specifically for the framework. The .NET Framework is a key Microsoft offering and is intended to be used by most new applications created for the Windows platform. The pre-coded solutions that form the framework's Base Class Library cover a large range of programming needs in a number of areas, including user interface, data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. The class library is used by

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

programmers, who combine it with their own code to produce applications. Programs written for the .NET Framework execute in a software environment that manages the program's runtime requirements. Also part of the .NET Framework, this runtime environment is known as the Common Language Runtime (CLR). The CLR provides the appearance of an application virtual machine so that programmers need not consider the capabilities of the specific CPU that will execute the program. The CLR also provides other important services such as security, memory management, and exception handling

Architecture



Visual overview of the Common Language Infrastructure (CLI)

CommonLanguageInfrastructure:
Thecoreaspectsofthe.NETframeworkliewithintheCommonLanguage Infrastructure, or CLI. The purpose of the CLI is to provide a language-neutral platform for application development and execution, includingfunctionsforexceptionhandling, garbage collection, security, and interoperability. Microsoft's implementation of the CLI is called the Common Language Runtime or CLR.

SYSTEMDESIGN

Software design sits at the technical kernel of he software engineering process and is applied regardless of the development paradigm and area of application. Design is the first step in the development phase for any engineered product or system. The designer's goal is to produce a model or representation of an entity that will later be built. Beginning, once system requirement have been specified and analyzed, system design is the first of the three technical activities -design, code and test that is required to build and verify software. The importance can be stated with a single word "Quality". Design is the place where quality is fostered in software development. Design provides us with representations of software that can assess for quality. Design is the only way that we can accurately translate a customer's view into a finished software product or system. Software design serves as a foundation for all the software engineering steps that follow. Without a strong design we risk building an unstable system – one that will be difficult to test, one whose quality cannot be assessed until the last stage. During design, progressive refinement of data structure, program structure, and procedural details are developed reviewed and documented. System design can be viewed from either technical or project management perspective. From the technical point of view, design is comprised of four activities – architectural design, data structure design, interface design and procedural design.

Dataflow diagrams:-A data flow diagram is graphical tool used to describe and analyze movement of data through a system. These are the central tool and the basis from which the other components are developed. The transformation of data from input to output, through processed, may be described logically and independently of physical components associated with the system. These are known as the logical data flow diagrams. The physical data flow diagrams show the actual implements and movement of data between people, departments and workstations. A full description of a system actually consists of a set of data flow diagrams. Using two familiar notations Yourdon, Game and Samson notation develops the data flow diagrams. Each component in a DFD is labeled with a descriptive name. Process is further identified with a number that will be used for identification purpose. The development of DFD'S is done in several levels. Each process in lower level diagrams can be broken down into a more detailed DFD in the next level. The lop-level diagram is often called context diagram. It consists a single process bit, which plays vital role in studying the current system. The process in the context level diagram is exploded into other process at the first level DFD.The idea behind the explosion of a process into more process is that understanding at one level of detail is exploded into greater detail at the next level. This is done until further explosion is necessary and an adequate amount of detail is described for analyst to understand the process.

Data dictionary: - After carefully understanding the requirements of the client the entire data storage requirements are divided into tables. The below tables

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

are normalized to avoid any anomalies during the course of data entry.

DOTNETPC-02.....Account_Open

Column Name	Data Type	Allow Nulls
aid	numeric(18, 0)	☑
accountno	nvarchar(50)	☑
uname	nvarchar(50)	☑
accounttype	nvarchar(50)	☑
branch	nvarchar(50)	☑
rs	nvarchar(50)	☑
approval	nvarchar(50)	☑
dates	nvarchar(50)	☑
		☐

Column Name	Data Type	Allow Nulls
userid	nvarchar(50)	✓
pass	nvarchar(50)	✓

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS:-

Unit testing:- Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing:-

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components

Functional test:-Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

System Test:-System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

White Box Testing:-White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing:-Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

SYSTEM SECURITY

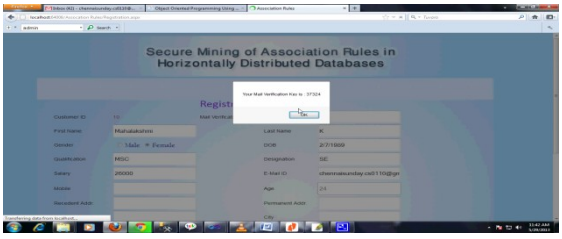
The protection of computer based resources that include hardware, software, data, procedures and people against unauthorized use or natural Disaster is known as System Security. System Security can be divided into four related issues: possible invasion of privacy. It is an attribute of information that characterizes its need for protection.

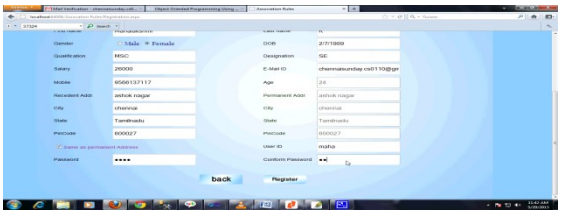
Security software:-System security refers to various validations on data in form of checks and controls to avoid the system from failing. It is always important to ensure that only valid data is entered and only valid operations are performed on the system. The system employees two types of checks and controls:

Client side validation:-Various client side validations are used to ensure on the client side that only valid data is entered. Client side validation saves server time and load to handle invalid data. Some checks imposed are:

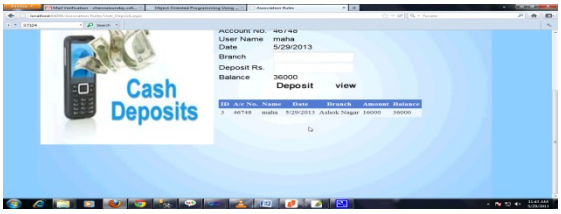
Server side validation:-Some checks cannot be applied at client side. Server side checks are necessary to save the system from failing and intimating the user that some invalid operation has been performed or the performed operation is restricted. Some of the server side checks imposed is: Server side constraint has been imposed to check for the validity of primary key and foreign key. A primary key value cannot be duplicated. Any attempt to duplicate the primary value results into a message intimating the user about those values through the forms using foreign key can be updated only of the existing foreign key values. User is intimating through appropriate messages about the successful operations or exceptions occurring at server side. Various Access Control Mechanisms have been built so that one user may not agitate upon another. Access permissions to various types of users are controlled according to the organizational structure. Only permitted users can log on to the system and can have access according to their category. User- name, passwords and permissions are controlled o the server side. Using server side validation, constraints on several restricted operations are imposed.

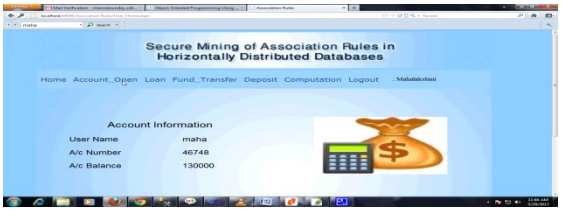
OUTPUT SCREENS

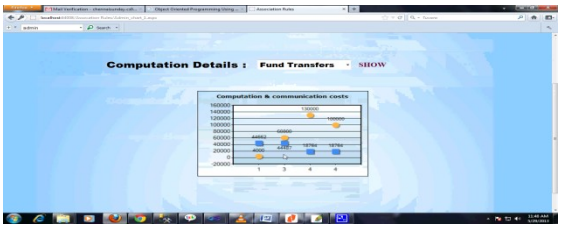


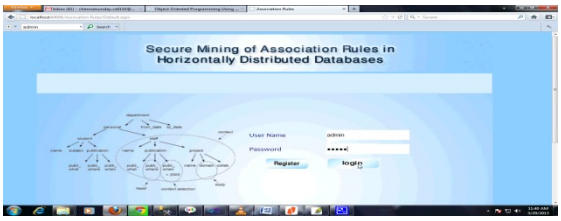


International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017









CONCLUSION

We proposed a protocol for secure mining of association rules in horizontally distributed databases that improves significantly upon the current leading protocol in terms of privacy and efficiency. One of the main ingredients in our proposed protocol is a novel secure multi-party protocol for computing the union (or intersection) of private subsets that each of the interacting players holds. Another ingredient is a protocol that tests the inclusion of an element held by one player in a subset held by another. Those protocols exploit the fact that the underlying problem is of interest only when the number of players is greater than two. One research problem that this study suggests. Namely, to devise an efficient protocol for inequality verifications that uses the existence of a semi honest third party. Such a protocol might enable to further improve upon the communication and computational costs of the second and third stages of the protocol. Other research problems that this study suggests is the implementation of the techniques presented here to the problem of distributed association rule mining in the vertical setting the problem of mining generalized association rules, and the problem of subgroup discovery in horizontally partitioned data

REFERENCES

*(*1). User Interfaces in C#: Windows Forms and Custom Controls by Matthew MacDonald.

(2).Applied Microsoft® .NET Framework Programming (Pro-Developer) by Jeffrey Richter.

(3). Practical .Net2 and C#2: Harness the Platform, the Language, and the Framework by Patrick Smacchia.
Data Communications and Networking, by Behrouz A Frozen.
(4). Computer Networking: A Top-Down Approach, by James F. Kurose.
(5). Operating System Concepts, by Abraham Silberschatz.
(6).M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, (7).A. Rabkin, I. Stoica, and M. Zaharias, "Above the clouds: A berkeley view of cloud computing," University of California, Berkeley, Tech. Rep. USB-EECS-2009-28, Feb 2009.
(8). Amazon Web Services (AWS), online at http://aws. amazon.com.
(9).Google App Engine, Online at http://code.google.com/appengine/.
(10).MicrosoftAzure,http://www.microsoft.com/azure/10.104th United States Congress, "Health Insurance Portability and Accountability Act of 1996 (HIPPA)," Online at http://aspe.hhs.gov/ admnsimp/pl104191.htm, 1996.
(11). H. Harney, A. Colgrove, and P. D. McDaniel, "Principles of policy in secure groups," in Proc. of NDSS'01, 2001.
(12). [7] P. D. McDaniel and A. Parkas, "Methods and limitations of security policy reconciliation," in Proc. of SP'02, 2002
(13). [8] T. Yu and M. Winslett, "A unified scheme for resource protection in automated trust negotiation," in Proc. of SP'03, 2003.
(14). [9] J. Li, N. Li, and W. H. Wins borough, "Automated trust negotiation using cryptographic credentials," in Proc. of CCS'05, 2005.
(15)J. Anderson, "Computer Security Technology Planning Study," Air Force Electronic Systems Division, Report ESD-TR-73-51, 1972, http: //seclab.cs.ucdavis.edu/projects/history/.