SMC Protocol for Naïve Bayes Classification over Grid Partitioned Data using Multiple UTPs

Alka Gangrade
Associate Professor
Technocrats Institute of Technology,
Bhopal, MP, India

Ravindra Patel
Head, Deptt. of MCA,
U.I.T., R.G.P.V.
Bhopal, MP, India

ABSTRACT

The case where data is distributed horizontally as well as vertically, it refers as grid partitioned data. SMC protocol for Naïve Bayes classification over grid partitioned data is offered in this paper. Also present a solution of the Secure Multi-party Computation (SMC) problem in the form of a protocol that preserves privacy. In this system, a protocol with several Un-trusted Third Parties (UTPs) is used, where there is almost impossible of privacy leakage. Multiple UTPs will calculate the model parameters for integrating the horizontally partitioned data. After that secure multiplication protocol will apply on vertical partitioned data (multiple UTPs) to classify the new tuple. The main contribution of this paper is that it shows a simple and easy calculation for developing Naïve Bayes classifier for grid partitioned data. The evaluation method is simple and more efficient.

General Terms

SMC, Naïve Bayes Classification.

Keywords

Privacy preserving, probability, secure multiplication protocol, grid partitioned, UTP.

1. INTRODUCTION

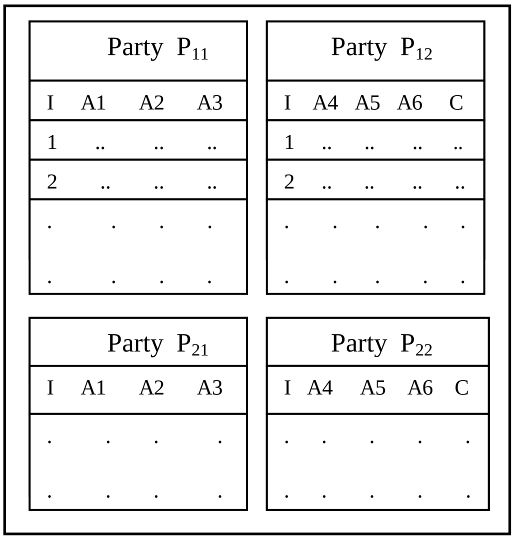
In modern years privacy preserving data mining has emerged as a very active research area in data mining. Over the past few years the research in privacy preserving data mining has concentrated on two major issues: data which is horizontally partitioned and data which is vertically partitioned. Horizontal partitioned data which is homogeneously distributed, meaning that all data tuples have the same attribute set. Vertical partitioned data which is heterogeneously distributed. Basically this means that data is collected by different sites or parties on the same individuals but with different attribute sets. Consider for instance financial institutions as banks and credit card companies, they both collect data on customers having a credit card but with differing attribute sets.

1.1 Grid Partitioned Data

In this paper, data which is both horizontally and vertically distributed is considered, which term as grid partitioned data. There has been limited research till now in privacy preserving data mining that considers grid partitioned data. However this kind of situation seems highly relevant and significant. Consider for instance the situation where different financial institutions gather data on clients concerning savings account, credit cards, investments etc. This situation clearly considers data which is grid partitioned, since some institutions deal with credit cards and not with investments and vice versa and since financial institutions typically have data emerging from various branches of a bank. In this section, a proper definition of horizontally, vertically and grid partitioned data [1] is provided. Let us consider:

1. A relation or data set R over the schema I, A_1, ..., A_a, C consisting of a finite number of tuples. The attribute I is supposed to be a key or identifier and is not considered as an attribute to calculate the model parameters. The only purpose of the attribute I is to be able to join or identified the vertically distributed data. The attribute C will be referred to as the class attribute.
2. Parties P_{ij} where i = 1, .., m, j = 1, .., n and n smaller than the number of attributes.
3. Each party P_{ij} is holding a part R_{ij} containing information about certain attributes (including I) and certain tuples. The R_{ij} are such that

- R_{ij} is a partition of R;
- R_{ij} and $R_{i'j}$ have the same attributes but having different tuples of R when i != i';
- R_{ij} and $R_{ij'}$ have different attributes but contain information about the same tuples of R when j != j';



Party P_{11}

I	A1	A2	A3
1	..	..	..
2	..	..	..
.	.	.	.
.	.	.	.

Party P_{12}

I	A4	A5	A6	C
1	..	..	..	..
2	..	..	..	..
.	.	.	.	.
.	.	.	.	.

Party P_{21}

I	A1	A2	A3
.	.	.	.
.	.	.	.

Party P_{22}

I	A4	A5	A6	C
.	.	.	.	.
.	.	.	.	.

Figure 1: Grid partitioned data.

Relation R is described as follows:

- Horizontally partitioned if and only if n = 1;
- Vertically partitioned if and only if m = 1; and
- Grid partitioned if and only if m and n >= 2. Fig.1 shows the grid partitioned data.

1.2 Classification Rule Mining

Classification is a popular data mining technique used to predict group membership for data tuples. In classification rule mining, a set of database tuples act as a training sample and it is analyzed to produce a model of the data or classifier that can be used for classifying a new tuple. The popular classification rule mining techniques are decision trees, neural networks, Naïve Bayesian classifiers etc. Privacy preserving data mining is the emerging field that protects susceptible data. The goal of privacy preserving classification is to build precise classifiers without disclosing personal information in the data being mined.

In this paper, a new algorithm is proposed to preserve privacy while calculating the model parameters and classifying a new tuple over grid partitioned data using Naïve Bayes classification, involving multiple parties.

1.3 Naïve Bayesian Classification

Bayesian classification is based on Bayes' theorem [2]. Bayes' theorem is

$$P(H|X) = \frac{P(X|H)\,P(H)}{P(X)} \qquad (1)$$

Where H is some hypothesis, such as that the data tuple X belongs to a specified class 'C'.

The posterior probability of H conditioned on X is P (H|X).
The prior probability of H is P (H).
The posterior probability of X conditioned on H is P (X|H).

Naïve Bayes is extremely effective but straightforward classifier. Due to this combination of straightforward and effectiveness it is used as a baseline standard by which other classifiers are measured. A simple Bayesian classifier is known as the Naïve Bayesian classifier, to be comparable in performance with decision tree and selected neural network classifier. It represents each class with a probabilistic summary and to classify each new tuple with the most likely class. It provides a flexible way for dealing with any number of attributes or classes, and is based on probability theory. It is fast learning algorithm that examines all its training input. It has been established to achieve unexpectedly well in a wide variety of problems despite of the simple nature of the model. With various enhancements it is highly effective, and receives practical use in many applications for example content based filtering and text categorization [3].

The framework is used for preserving privacy is defined in Secure Multiparty Computation [4], and several primitives from the Secure Multiparty Computation contents. Complete details of Naïve Bayes classification algorithms can be found in [3]. Here assume that the basic formulae are well known. In order to construct SMC protocol for Naïve Bayesian classifier, must concentrate on two issues, calculation of the probability or model parameter for each attribute and classification of a new tuple [5, 6, 7]. The protocol presented below is quite efficient.

1.4 Main Contributions

Main contributions in this paper are as follows:

- Present a SMC protocol for Naïve Bayes classifier over grid partitioned databases.
- First integrate horizontally partitioned data using multiple UTPs.
- Classify new tuple by applying secure multiplication protocol on vertically partitioned data (multiple UTPs).

1.5 Organization of the paper

The rest of the paper is organized as follows. Section 2, discuss the background study. Section 3 describes proposed work of new SMC protocol for Naïve Bayes classification architecture for grid partitioned data. Section 3.1 describes the system architecture. Section 3.2 sets some assumptions. Section 3.3 and 3.4 describe informal and formal description of proposed protocol respectively. Section 4, present results that are conducted by using proposed architecture on real-world data sets. Section 5, conclude the paper with the discussion of the future work.

2. BACKGROUND

Privacy preserving data mining has been an active research area for a decade. A lot of research work is going on privacy preserving classification in distributed data mining. Yao described the first Secure Multiparty Computation (SMC) problem [8]. SMC allows parties with similar background to compute result upon their private data, minimizing the threat of disclosure was explained [9].

There have been several approaches to support privacy preserving data mining over multi-party without using third parties [5, 10]. Some techniques, review and evaluation of privacy preserving algorithms also presented in [10]. Various tools discussed and how they can be used to solve several privacy preserving data mining problems [11]. Now give some of the related work in this area. The aim is to preserve customer privacy by distorting the data values presented in [12]. Agrawal D. and Aggarwal C. C. designed various algorithms for improving this approach [13].

Classification is one of the most extensive data mining problems come across in real life. General classification techniques have been extensively studied for over twenty years. The classifier is usually represented by classification rules, decision trees, Naïve Bayes classification and neural networks. First ID3 decision tree classification algorithm is proposed by Quinlan [14]. A secure algorithm to build a decision tree using ID3 over horizontally partitioned data between two parties using SMC is proposed by Lindell and Pinkas [15]. A novel privacy preserving distributed decision tree learning algorithm [16] that is based on Shamir [17]. The ID3 algorithm is scalable in terms of computation and communication cost, and therefore it can be run even when there is a large number of parties involved and eliminate the need for third party and propose a new method without using third party. A generalized privacy preserving variant of the ID3 algorithm for vertically partitioned data distributed over two or more parties introduced in [18, 19, 20, 21] and horizontally partitioned data distributed over multi parties introduced in [22, 23]. Privacy preserving Naïve Bayes classification for horizontally partitioned data introduced in [5, 24] and vertically partitioned data introduced in [6, 7, 25]. Vaidya J. et. al. proposed both the partitioned in [26]. Centralized Naïve Bayes classification probability calculation is introduced in [27].

3. SMC PROTOCOL FOR NAÏVE BAYES CLASSIFICATION OVER GRID PARTITIONED DATA

In this section, system architecture, informal and formal description of proposed protocol for Naïve Bayes classification over grid partitioned data is introduced. First, multiple UTP calculate model parameters to integrate horizontal partitioned data and then secure multiplication protocol will apply on vertical partitioned data to classify the new tuple. A concept of multiple UTP is introduced in computation layer$_1$ to achieve full security. All the layers are having their own predefined functionality. Each layer communicates with its next layer.

3.1 System Architecture

Proposed architecture of four-layer SMC protocol for Naïve Bayes Classifier over grid partitioned databases is shown in Fig. 2. The four layers are named input layer, computation layer$_1$, computation layer$_2$ and output layer. The last layer of this system is output layer. It finds out the class having maximum probability, announces the result publicly and sends the result to all the UTPs.

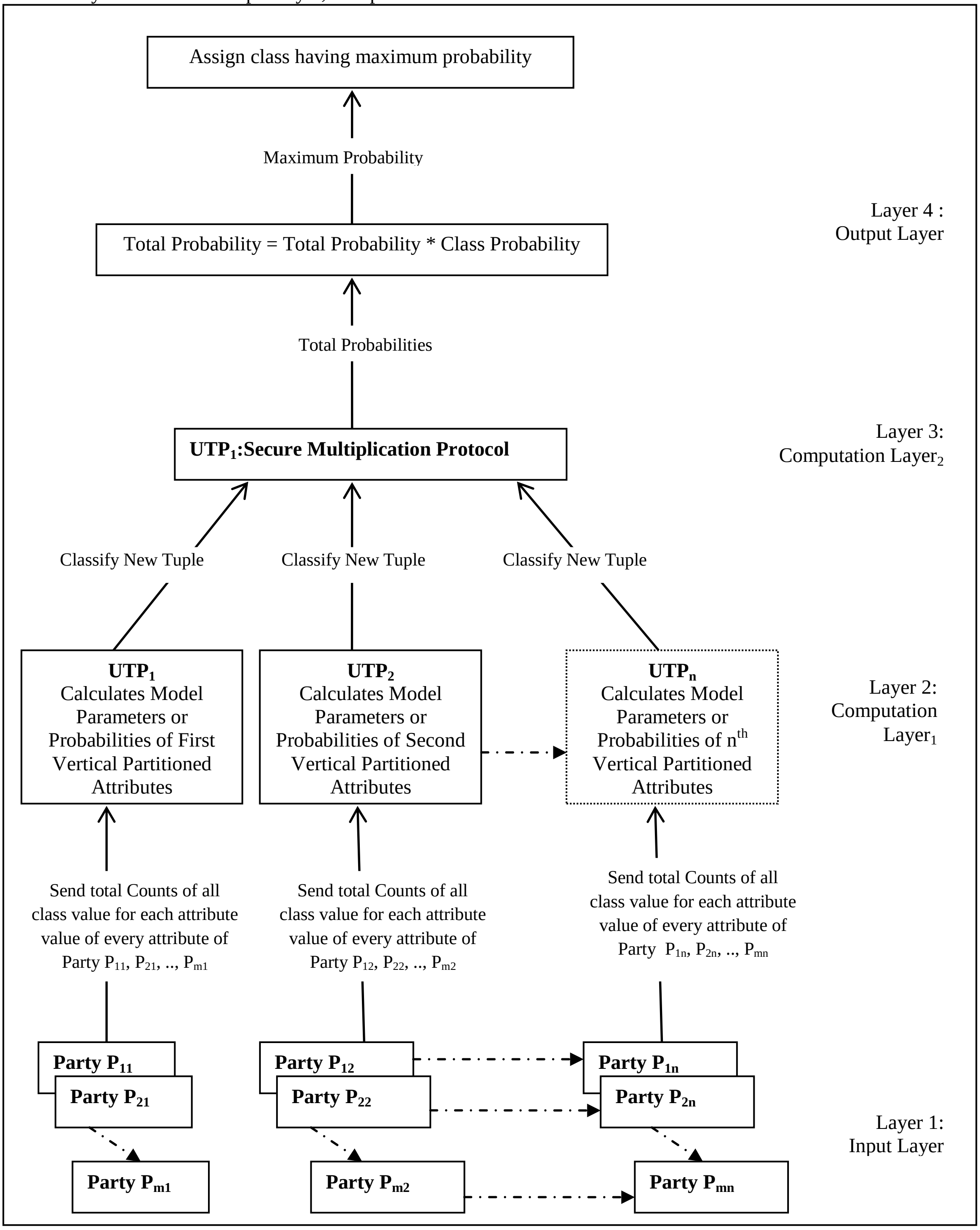

Assign class having maximum probability
Maximum Probability
Layer 4 : Output Layer
Total Probability = Total Probability * Class Probability
Total Probabilities
Layer 3: Computation Layer$_2$
UTP$_1$:Secure Multiplication Protocol
Classify New Tuple
Classify New Tuple
Classify New Tuple
UTP$_1$ Calculates Model Parameters or Probabilities of First Vertical Partitioned Attributes
UTP$_2$ Calculates Model Parameters or Probabilities of Second Vertical Partitioned Attributes
UTP$_n$ Calculates Model Parameters or Probabilities of nth Vertical Partitioned Attributes
Layer 2: Computation Layer$_1$
Send total Counts of all class value for each attribute value of every attribute of Party P$_{11}$, P$_{21}$, .., P$_{m1}$
Send total Counts of all class value for each attribute value of every attribute of Party P$_{12}$, P$_{22}$, .., P$_{m2}$
Send total Counts of all class value for each attribute value of every attribute of Party P$_{1n}$, P$_{2n}$, .., P$_{mn}$
Party P$_{11}$
Party P$_{21}$
Party P$_{m1}$
Party P$_{12}$
Party P$_{22}$
Party P$_{m2}$
Party P$_{1n}$
Party P$_{2n}$
Party P$_{mn}$
Layer 1: Input Layer

Figure 2: Four Layer architecture for Naïve Bayes Classification over Grid Partitioned data.

3.2 Assumptions

- All participating parties know the class value of all tuple they have and individually calculate all counts.
- Number of UTPs is equal to the number of vertical partition i.e. n.
- UTP$_1$ (first un-trusted third party) drives the secure multiplication protocol.
- UTP$_1$ (first un-trusted third party) calculates the total probabilities for new tuple to be classified.
- Send class value to all parties via all UTPs.

- Input data of all parties are secured and privacy is preserved.
- The Secure Multiplication Protocol used by the UTPs is secured.
- The communication networks used by the input parties to communicate with the particular UTP are secure.

3.3 Informal Description of Protocol

The proposed protocol is divided into four layers. Each layer is having some unique functions. Study and analyze separately the functionality of each layer of the protocol. The functions of layers are summarized as follows:

3.3.1 Input Layer:

- Input layer comprises of all the parties that are involved in the classification process.
- Each party calculates all counts separately.
- Send all counts to corresponding UTP to integrate horizontally partitioned result.

3.3.2 Computation Layer$_1$: (Multiple UTPs – UTP$_1$, UTP$_2$, .., UTP$_n$)

- Receives all counts.
- Calculates model parameters or probabilities for all class value of each attribute value for every attribute.
- Integrate horizontally partitioned data.

3.3.3 Computation Layer$_2$: A new tuple to be classified.

- Apply secure multiplication protocol [25].
- First UTP drives the secure multiplication protocol.
- Calculate total probability for all class value for new tuple to be classified.

3.3.4 Output Layer:

- Based on the total probability of all class value, first UTP will find the class with the highest total probability and finally classify the new tuple.
- Announces the class value publicly as well as send back to the all UTPs.
- Each UTP sends the class value to respective parties.

3.4 Formal Description of Protocol

Require:

1. P_{ij} parties, where i = 1, 2, ..,m and j = 1, 2, ..,n
2. c class values i.e. c_1, c_2, .., c_c,
3. a attribute name, where a = a1 + a2 + ... + an and A_c is the class attribute.

$$P_{ij} \rightarrow P_{ij}.A_{j1}, P_{ij}.A_{j2}, .., P_{ij}.A_{aj}$$

Note:

- C^{ij}_{xyz}: represents number of tuples with party P_{ij} having class z, attribute value y of attribute A_x.
- $UTP_j.C_{xyz}$: represents number of tuples having class z, attribute value y of attribute A_x of UTP_j.
- A_{jx}: represents attribute name A_x of party P_{ij}.
- $UTP_j.A_{xy}$: represents attribute name A_x with attribute value y of UTP_j.
- N^i_z: represents number of tuples with ith horizontally participating party having class z.
- N_z: represents number of tuples having class z.
- T: represents total tuples with all participating parties.
- $UTP_j.Prob_{xyz}$: represents probability of attribute A_x with attribute value y having class z of UTP_j.
- $Prob_z$: represents probability of class z.
- $New.UTP_j.A_{xy}$: represents new tuple of attribute name A_x with attribute value y of UTP_j to be classified.

Algorithm 1: 4LPPGPNBC ( ) – Four-layer privacy preserving grid partitioned NBC to calculate model parameters.

1. Local_Att_Count ( )
2. Local_Class_Count ( )
3. Global_Class_Count ( )
4. Cal_Att_Prob ( )

3.4.1 Input Layer:

Algorithm 2: Local_Att_Count ( ) - Calculate local counts for each party for each attribute value for every attribute for all class value.

```
For Party $P_{ij}$ where j = 1 to n do
  For  i = 1 to m do
    For Attribute $A_{jx}$ where x = 1 to $a_j$ do
      For Attribute value $v_y$ where y = 1 to $v_x$ do
        For Class value $c_z$ where z = 1 to c do
          $C^{ij}_{xkz} = 0$
          For all tuples having class value z
            $C^{ij}_{xyz} = C^{ij}_{xyz} + 1$
          End for
        End for
      End for
    End for
  End for
End for.
```

Algorithm 3: Local_Class_Count ( ) - Calculate local counts for each party for all class value and total number of tuples with all horizontally participating parties.

```
T = 0, j = 1
For Party $P_{ij}$ where i = 1 to m do
  For Class value $c_z$ where z = 1 to c do
    $N^i_z = 0$
    For all tuples having class value z
      T = T + 1
      $N^i_z = N^i_z + 1$
    End for
  End for
End for.
```

3.4.2 Computation Layer$_1$

Algorithm 4: Global_Class_Count ( ) - Calculate total or global counts and probabilities for all class value.

```
j = 1
For Class value $c_z$ where z = 1 to c do
  $N_z = 0$
  For Party $P_{ij}$ where i = 1 to m do
    $N_z = N_z + N^i_z$
  End for
  $Prob_z = N_z / T$
End for.
```

Algorithm 5: Cal_Att_Prob ( ) - Calculate probability of each attribute value for every attribute for all class value for all vertically participating parties i.e. integrating horizontally.

```
For Party $P_{ij}$ where j = 1 to n do
  For Attribute $A_{jx}$ where x = 1 to $a_j$ do
    For Attribute value $v_y$ where y = 1 to $v_x$ do
      For Class value $c_z$ where z = 1 to c do
        $UTP_j.C_{xyz} = 0$
        For Party $P_{ij}$ where i = 1 to m do
          $UTP_j.C_{xyz} = UTP_j.C_{xyz} + C^{ij}_{xyz}$
        End for
```

```
            i.  $UTP_j.Prob_{xyz} = UTP_j.C_{xyz} / N_z$
 7.         End for
 8.       End for
 9.     End for
10.   End for
```

3.4.3 *Computation Layer$_2$: Secure Multiplication Protocol*

Algorithm 6: Cal_Total_Prob (c_z) : Calculate total Probability for all class value of new tuple by using secure multiplication protocol [25].

```
1.   For Class value $c_z$ where z = 1 to c do
     i. $Total\_Prob_z = 1$
2.     For $UTP_j$ where j = 1 to n do
3.       For Attribute $A_{jx}$ where x = 1 to $a_j$ do
4.         For Attribute value $v_y$ where y = 1 to $v_x$ do
5.           If $UTP_j.A_{xy} = New.UTP_j.A_{xval}$ then
               i.   $Total\_Prob_z = Total\_Prob_z * UTP_j.Prob_{xyz}$
              ii.   Break
6.           End if
7.         End for
8.       End for
9.     End for
10.    Return $Total\_Prob_z$
11.  End for.
```

3.4.4 *Output Layer:*

Algorithm 7: Classify_Tuple ( ): Find the maximum probability and classify the tuple [24, 25].

```
1.  Max_Prob = 0
2.  Class = Null
3.  For Class value $c_z$ where z = 1 to c do
    i. Prob = Cal_Total_Prob ($c_z$) * $N_z$/T
4.    If Prob > Max_Prob then
        i.   Max_Prob = Prob
       ii.   Class = $c_z$
5.    End if
6.  End for
7.  For $UTP_j$ where j = 1 to n do
    i. $A_c$ = Class
8.  End for.
```

4. EVALUATION AND RESULTS

In this Naïve Bayes classifier for grid partitioned data, first integrate horizontal partitioned data using multiple UTPs and then applying secure multiplication protocol on these UTPs i.e. on vertical partitioned data. Table 1 shows Student data set. Here we are addressing four parties, Student data set is divided into four parties, where parties are distributed horizontally as well as vertically i.e. grid partitioned. Each party has three attributes including class attribute. Class attribute has two values. For the real word data experiment results, 400 records are generated, and randomly choose 200 records for training sample, and remaining 200 records for testing purpose. WEKA [27] data mining software is used to run existing NBC and proposed four-layer grid partitioned NBC, and reported the experiment results on the test data. Experiment results show total time taken to calculate the probabilities or model parameters on training data and accuracy on test data. Number of parties as well as the number of attributes could be extended. In this proposed system parties are communicating their intermediate results only not the actual data thus the protocol protect the actual data of parties in the process. Thus, privacy is being maintained. Its execution time for calculating the model parameters or probabilities are less than the existing Naïve Bayes classifier, three-layer privacy preserving horizontal partitioned NBC (3LPPHPNBC) [24] and three-layer privacy preserving vertical partitioned NBC (3LPPVPNBC) [25] with almost same accuracy. Execution time comparison is shown by table 2 and accuracy on test data is shown by table 3. Execution time comparison graph is shown by fig 3. But the time for classifying a new tuple by grid partitioned is same as 3LPPVPNBC but greater than 3LPPHPNBC.

Table 1: Student Data set Description

Attribute Name	No. of values	Category
Age	3	<=30, 31..40, >40
Income	3	Low, Medium, High
Student	2	Yes, No
Credit_rating	2	Fair, Excellent
Buys_computer (Class)	2	Yes, No

Table 2: Execution time for Calculating Model Parameters

S. No.	No. of Tuples	Existing NBC (ms)	3-L PP HPNBC (ms)	3-L PP VPNBC (ms)	4-L PP GPNBC (ms)
1	14	69	20	14	13
2	25	82	32	16	15
3	50	97	42	18	16
4	100	111	50	30	26
5	200	133	57	34	29

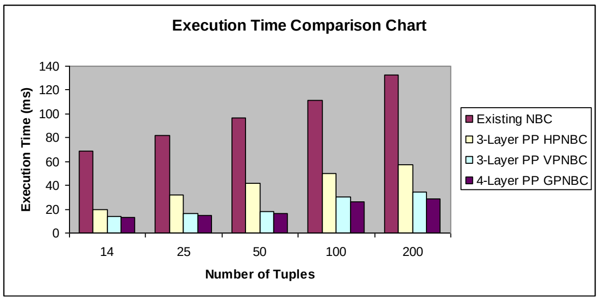


Figure 3: Execution time comparison chart

Table 3: Test data Accuracy

S. No.	Number of Tuples	Accuracy (%)
1	14	78.57%
2	25	80%
3	50	82%
4	100	83%
5	200	84%

5. CONCLUSIONS

In this paper proposed system is divided into four layers. This helped us to analyze the problem step wise. Instead of using data transformation, multiple UTPs are used to integrate the horizontally partitioned data by transferring the model parameters to particular UTP while keeping the actual data secure and apply secure multiplication protocol on these UTPs to classify the new tuple. Proposed classification system is quite efficient and fast. It is also much faster than ID3 and C4.5 decision tree classifier because Bayesian classifier only needs to go through the whole training data once. They are also space efficient because they build up various frequency tables only. The authors are continuing work in this field to develop decision tree classifier for grid partitioned databases and also analysis new as well as existing classifiers.

6. ACKNOWLEDGMENTS

The authors are thankful to the University and College for their valuable support.

7. REFERENCES

[1] Kuijpers, B., Lemmens, V. and Moelans, B. 2008. Privacy Preserving ID3 over Horizontally, Vertically and Grid Partitioned Data, [cs.DB], 11 march 2008.

[2] Han, J. and Kamber, M. 2001. Data Mining: Concepts and Techniques, Elsevier, India.

[3] Mitchell, T. 1997. Machine Learning, 1st edn. McGraw-Hill Science/Engineering/Math, New York.

[4] Goldreich O. 1998. Secure multi-party computation, Sep 1998. (working draft).

[5] Kantarcioglu, M. and Vaidya, J. 2003. Privacy preserving naive Bayes classifier for horizontally partitioned data, In IEEE ICDM Workshop on Privacy Preserving Data Mining, Melbourne, FL, pp. 3-9, November 2003.

[6] Vaidya, J. and Clifton, C. 2004. Privacy preserving naive Bayes classifier on vertically partitioned data, Proc. SIAM International Conference on Data Mining, Lake Buena Vista, Florida, pp. 22-24, April 2004.

[7] Yang, Z. and Wright, R. 2006. Privacy-Preserving Computation of Bayesian Networks on Vertically Partitioned Data, IEEE Transactions on Data Knowledge Engineering, 18(9), April 2006, pp. 1253-1264.

[8] Yao, A. C. 1982. Protocols for secure computation, Proc. of 23rd IEEE Symposium on Foundations of Computer Science (FOCS), pp. 160-164.

[9] Du, W. and Attalah, M. J. 2001. Secure multi-problem computation problems and their applications: A review and open problems, Tech. Report CERIAS Tech Report 2001-51, Center for Education and Research in Information Assurance and Security and Department of Computer Sciences, Purdue University, West Lafayette, IN 47906.

[10] Verykios, V. and Bertino, E. 2004. State-of-the-art in Privacy preserving Data Mining, SIGMOD Record, 33(1), pp. 50-57.

[11] Cliffton, C., Kantarcioglu, M. and Vaidya, J. 2004. Tools for privacy preserving distributed data mining, ACM SIGKDD Explorations Newsletter, 4(2), pp. 28-34.

[12] Agrawal, R. and Srikant, R. 2000. Privacy preserving data mining, Proc. of the ACM SIGMOD on Management of data, Dallas, TX USA, pp. 439-450, May 15-18.

[13] Agrawal, D. and Aggarwal, C. C. 2001. On the design and quantification of privacy preserving data mining algorithms, Proc. of the Twentieth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Santa Barbara, California, USA, pp. 247–255, May 21-23 2001.

[14] Quinlan, J. R. 1990. Induction of decision trees, in: Jude W. Shavlik, Thomas G. Dietterich, (Eds.), Readings in Machine Learning, Morgan Kaufmann, 1, pp. 81–106.

[15] Lindell, Y. and Pinkas, B. 2002. Privacy preserving data mining, Journal of Cryptology, 15(3), pp. 177–206.

[16] Emekci, F., Sahin, O. D., Agrawal, D. and Abbadi, A. E. 2007. Privacy preserving decision tree learning over multiple parties, Data & Knowledge Engineering 63, pp. 348-361.

[17] Shamir, A. 1979. How to share a secret, Communications of the ACM, 22(11), pp. 612-613.

[18] Du, W. and Zhan, Z. 2002. Building decision tree classifier on private data, In CRPITS, pp. 1–8.

[19] Fang, W. and Yang, B. 2008. Privacy Preserving Decision Tree Learning Over Vertically Partitioned Data, Proc. of the 2008 International Conference on Computer Science & Software Engineering, pp. 1049-1052.

[20] Vaidya, J. 2004. Privacy preserving data mining over vertically partitioned data, doctoral diss., Purdue University, August 2004.

[21] Vaidya, J., Clifton, C., Kantarcioglu, M. and Patterson, A. S. 2008. Privacy-preserving decision trees over vertically partitioned data, Proc. of the 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security, pp. 139–152.

[22] Gangrade, A. and Patel, R. 2011. A novel protocol for privacy preserving decision tree over horizontally partitioned data, International Journal of Advanced Research in Computer Science, 2 (1), pp. 305-309.

[23] Gangrade, A. and Patel, R. 2012. Privacy preserving two-layer decision tree classifier for multiparty databases, International Journal of Computer and Information Technology (2277 – 0764), 1(1), pp. 77-82.

[24] Gangrade, A. and Patel, R. 2012. Privacy preserving Naïve Bayes classifier for horizontally distribution scenario using Un-trusted Third Party, *IOSR Journal of Computer Engineering (IOSRJCE)* ISSN: 2278-0661, ISBN: 2278-8727, Volume 7, Issue 6 (Nov. - Dec. 2012), pp 04-12.

[25] Gangrade, A. and Patel, R. 2012. Privacy preserving three-layer Naïve Bayes classifier for vertically partitioned databases, paper communicated in Journal of Information and computing Science (JIC), ISSN 1746-7659 (print).

[26] Vaidya, J., Kantarcioglu, M. and Clifton, C. 2008. Privacy-preserving Naïve Bayes classification. The VLDB Journal (2008) 17, pp. 879–898.

[27] Witten, I. H., Frank, E. and Hall, M. A. 2011. Data Mining Practical Machine Learning Tools and Techniques, Burlington, MA, Morgan Kaufmann.