Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Performance Improvement of Privacy Preserving in K-anonymous Databases Using Advanced Encryption Standard Technique

Mr. Mahesh T.Dhande[1]
Research Scholar,
S.S.G.B.C.O.E.T, Bhusawal, India

Mrs. Neeta A.Nemade[2]
Guide & Assi.Professor,
S.S.G.B.C.O.E.T, Bhusawal, India

Abstract— it's habit to collect data in the format of digital information by governments, corporations, and individuals. So to protect that information by sectors uses different techniques such as k-anonymization. In k-anonymization databases, its needs to determine when someone inserted a tuple and database is still k-anonymous. For Accessing database is in controlled. Because for certain experiments that need to be maintained data confidentially. Whether a k-anonymous database retains anonymity once a new tuple is being inserted to it without violets the sensitive attributes and the database. In this paper, we propose a protocol solving this problem on suppression based k-anonymous and generalization-based k-anonymous and confidential Databases with suitable cryptographic assumptions.

Keywords— Anonymization, Encryption, security, Data confidentiality, Privacy-preserving.

I. INTRODUCTION

The problem of privacy-preserving data mining has become more important in recent years because of the increasing ability to store personal data about users is in the format of digital information, and to provide security to the personal database a number of ways and techniques are available such as randomization, cryptography and k-anonymity have been suggested in recent years in order to perform privacy-preserving data mining. The important goal is to protect individual's information without violating the data owner's privacy [6]. Today there is an increased concern for privacy. Because availability of huge numbers of databases and that contain a large variety of information about individuals, detailed person-specific data in its original form often contains sensitive information about individuals, and publishing such data immediately violates individual privacy [2]. It's important task of develop methods and tools for secured data in a more protective environment, so that the secured access data remains practically useful for other and while individual privacy is also preserved.

1.1 Problem Statement

The general participating of our application is data provider enter data onto the system is in the form of tuple then system perform anonymous authentication and update operation on tuple and finally stored that tuple in database The user can also access the stored record by simply using query. However, a general problem is that if we consider that the stored data is confidential, then how it possible to grant for update and preserve anonymity of database. In such a condition of modification process of adding individuals information, two problems are introduces according to anonymity and confidentiality of the data in the database and the privacy of data provider. 1) Is the modified database still preserving the privacy? And 2) is it necessary to know the actual inserted data to database owner? For understanding Figure 1 captures the exact participating of our application [2][4].

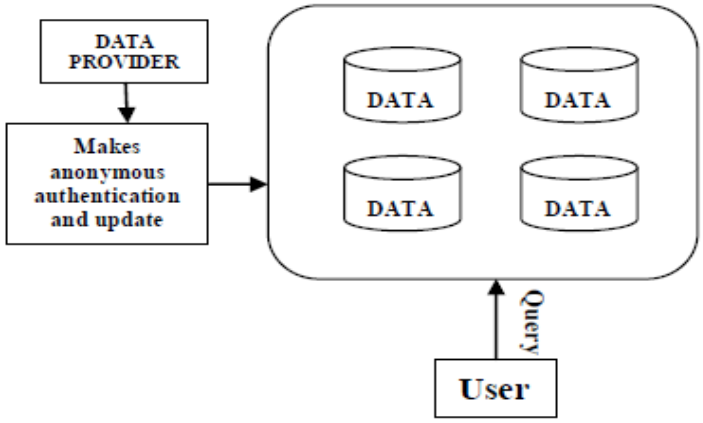


Figure 1: Anonymous Database System

Figure 1 shows Anonymous database system, in this system data provider provides the data then performing authentication by using private checker and then perform the anonymization and updation of database and finally stored that record in database. User can access the anonymous database by using query. Here we consider the example of educational record in this single student is a data provider and stored the information of student in a tuple and stored in database at server. Only institution have authorized to access to information of student for authentication purpose. Since database is in anonymous form, the student's (data provider's) privacy is secured from other users. Now here database have private sensitive data of student, so it's very risky to revealing the data of student by unauthorized person so to protect privacy of individual student we use anonymization technique. If database is anonymous by using suppression based and generalization it is difficult to search student's identity in database. Now for updation of database, Suppose newly admitted student wants to enter his/her information, but inserting record in anonymous database can be done as follows: here in this example institute (authorized person) checks whether the inserted tuple is properly anonymous and after inserted database cannot be disturb and still anonymous then the tuple is inserted. Under this approach, the entire information to be inserted revealing by institute, thus it violating the privacy of the student. Another possibility when student do the verification then shall they verify himself/herself? Like this several problem needs to be addressed: The first problem is: if the data owner does not knowing the contents of inserted tuple t, how to preserve data integrity by directly inserting tuple and establishing anonymity of DB U {t}.The second problem is if database is perfect anonymous and forms anonymity, then it is possible to update? Third problem is: what is the situation when database is not preserved anonymity? Finally, forth problem is: when database is empty then what is the initial content of DB. In this paper, we propose a protocol solving this problem on suppression based and generalization-based k-anonymous database.

II. RELATED WORK

Early research efforts in the area of privacy preserving and confidentiality used protocols have some limitations, by using anonymous database for privacy preserving and if the inserted tuple is not match according to attributes of database then insertion can be rejected. The first research is based on algorithms for database anonymization. Database anonymization techniques are given by hiding sensitive attribute value by using data reduction, data perturbation technique. In this data is divided in two category first the stream of data is released continuously and anonymized or released in different fashion and anonymized [2]. The second research is based on to Secure Multi-party Computation (SMC) techniques. Which is related to cryptography? Cryptographers seek to make secure protocols as efficient as possible in order to minimize the performance gap between secure and naive protocols. The secure protocol for computing a certain function will always be more costly than a naive protocol that does not provide any security and also observe some problem. The third research is based on to the private information retrieval that is application of the secure multiparty computation (SMC) techniques area of data management. In this area user can use query to see database [2].

Finally, the fourth research direction is based on to query processing techniques for encrypted data.

III.PROPOSED TECHNOLOGY

The proposed protocol based on to solve Problem 1 that is the original database cannot be disturb and gain it's anonymity by inserting new record or tuple in it and if the inserted record is match with record stored Database Then, Problem 1 is equivalent to privately checking data deduplication for avoiding repetition and finding the data Integrity problem. The first purpose of proposed system at suppression-based anonymous databases, the database owner check inserted tuple then properly anonymize for inserting it to database, without losing its contents and also without giving any information about insertion to data provider. To satisfy such goal, the data provider must take care about his/her inserted information and their secured communication by encrypting information with their own private key to get the higher level privacy-preserving verification of the database anonymity. The second purpose of proposed system at generalization-based anonymous databases, it relies on a secure set of intersection protocol and Cryptographic Primitives such as the more general value can found in the Cryptographic assumptions for managing original dataset in generalization based k-anonymous DB.

A. Overview of Proposed System

Figure 2 shows the Overview of Proposed System. In the system user has ability to insert a tuples or recoed in the database but before inserting a tuple he/she can encrypt the data using shared secrete key by using AES algorithm. The shared secrete key is added because data provider can communicate with each other. The data provider can share a secrete key with each other based on a commutative encryption function using Diffie-Hellman Algorithm then the inserted tuple compares with existing data if any match found then the tuple is inserted in the database otherwise it is rejected[1]. The purpose for checking already existing data with new inserted data for avoiding no redundancy and solving the data integrity problem. In K-Anonymization the database is anonymous using suppressed and generalized technique that is much secured. After inserting a tuple into the database user will able to see the anonymous database by using query.

The Diffie-Hellman Algorithm

The data provider can communicate with each other by using Public key encryption scheme that work on a commutative encryption function.

1. Alice encrypts message M with her key (ka):- $ka \rightarrow \{M\}ka$.
2. Alice sends {M}ka to Bob.

3. Bob in his turn encrypts the received message$\{M\}ka$ with her key (kb): → $\{\{M\}ka\}$kb
4. Bob sends {{M}ka}kb back to Alice.
5. Alice is able to decrypt the received message due to commutativity $\{\{M\}ka\}\text{kb} = \{\{M\}kb\}ka :\rightarrow \{M\}kb$
6. Alice sends {M} kb to Bob, who can decrypt it using his key $kb \rightarrow M$.

Diffie and Hellman use a commutative encryption function based on discrete algorithm:

Appropriate prime p and generator g are chosen, and common for all users.

1. Alice chooses a secret random number xa(her private key) and publishes $ya = g^{xa}$ (Her public key).
2. Bob does the same with xb secret and $yb = g^{xb}$ public.
3. Alice uses $yb^{xa} = g^{xa\ xb}$ to encrypt a message to Bob.
4. Bob uses $ya^{xb} = g^{xa\ xb}$ to decrypt the received message [1].

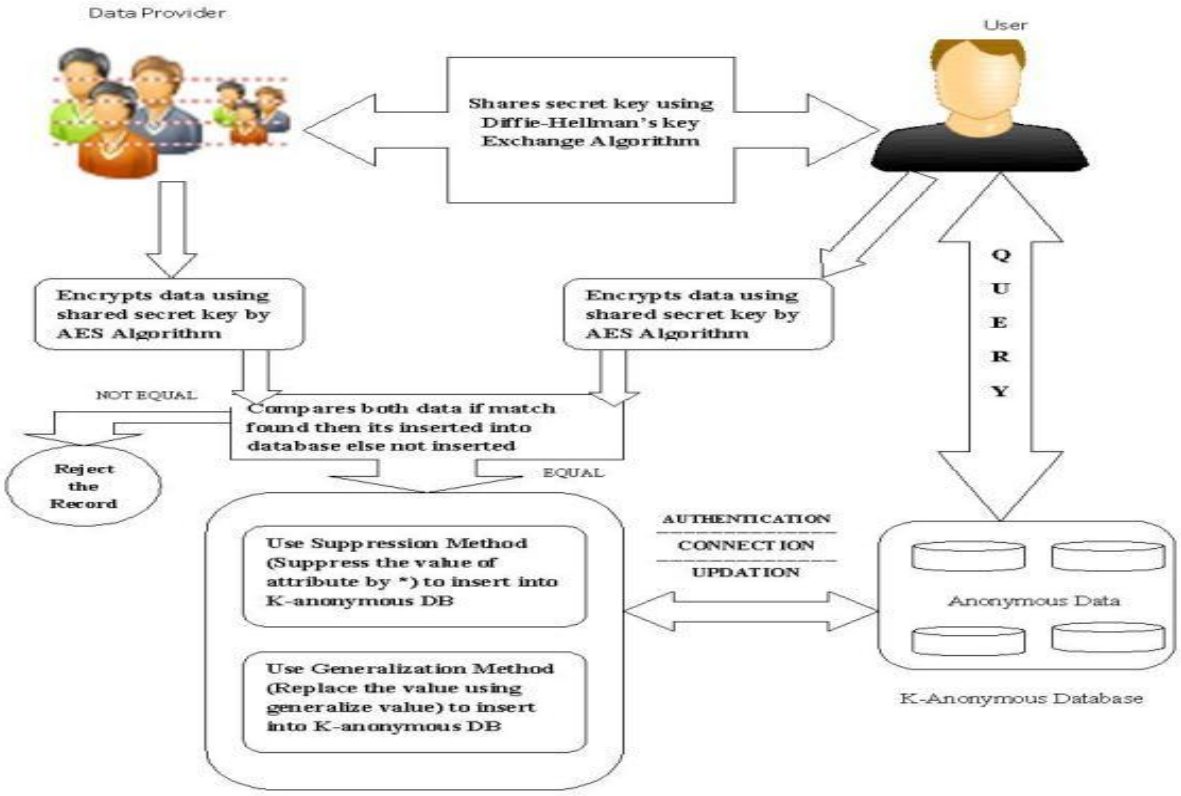


Figure 2: Overview of Proposed System

IV. SUPPRESSION BASED AND GENERALIZATION BASED K-ANONYMOUS TECHNIQUE

In the anonymous databases it is clear to understand the meaning of Anonymization. Anonymization technique is hiding sensitive attribute value in such a way that it can't identify back. Numbers of ways are available to do anonymization. We consider k-anonymization approach in that k is the total number of rows and we can't differentiate k with other k-1 rows by only considering a set of attributes, then this table is K-anonymized [8]. Privacy can be preserved by simply using k-anonymization approach on suppression based and generalized based techniques. Suppressing all sensitive data from database by using'*', and *Generalization* replaces a value with a "less-specific but semantically consistent" value according to apriori established value generalization hierarchies (VGHs)[9].

A. Suppression Based Anonymous Technique

When using a suppression-based anonymization method, we consider a table $T = \{t_1, \ldots \ldots \ldots, t_n\}$ tuples over the attribute set A. In *suppression* method, the aim is to form subsets of indistinguishable tuples by masking the values of some well-chosen attributes. We mask with the special value '*'. Now forming the subset and classify that subsets we use Quasi-Identifier (QI).

Quasi-Identifier (QI): Each record has a number of attributes: some attributes unique and personal(such as *disease* and *salary*) and some may be repeated and general that is quasi-identifiers (called QI, such as *zipcode*, *age*, and *gender*) by taking this we can easily identify someone [3][7].

To better understand the k-anonymization approach consider the example of patient As shown in table 1 which contains original database (Table T) having Quasi-Identifier *QI={Disease, gender, age}* or more sensitive three attributes value.

Table 1: original dataset

DISEASE	*GENDER*	*AGE*
Typhoid	Girl	19
HIV	Man	50
Typhoid	Women	45
Exothrix	Man	68
HIV	Boy	20
Exothrix	Women	63

Table 2: Suppressed Data with k = 2

DISEASE	*GENDER*	*AGE*
*	Girl	*
*	Man	*
*	Women	*
*	Man	*
*	Boy	*
*	Women	*

Table.3: Generalized Data with k = 2

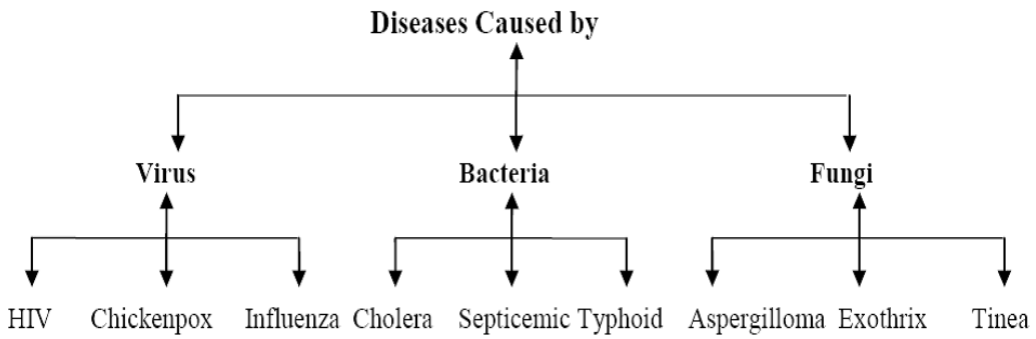
DISEASE	*GENDER*	*AGE*
Disease Caused by Bacteria	Female	[1-30]
Disease Caused by Virus	Male	[31-60]
Disease Caused by Bacteria	Female	[31-60]
Disease Caused by Fungi	Male	[61-100]
Disease Caused by Virus	Male	[1-30]
Disease Caused by Fungi	Female	[61-100]

After applying suppression based technique on original dataset the original dataset is anonymized and Table 2 shows a suppression based k-anonymization with k=2 it means that at least k=2 tuples should be indistinguishable by masking values.

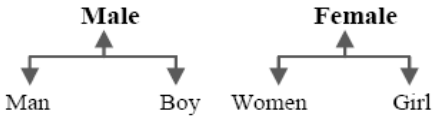
B. Generalization Based Anonymous Technique

In generalization-based anonymization consists in substituting the values of a given attribute with more general values in the database, according to a priori established value generalization hierarchies (VGHs) with some Cryptographic Primitives[3][5]. table 1 contains original information and after performing generalization based techniques original dataset the original dataset is anonymized and table 3 shows generalized data with k=2.

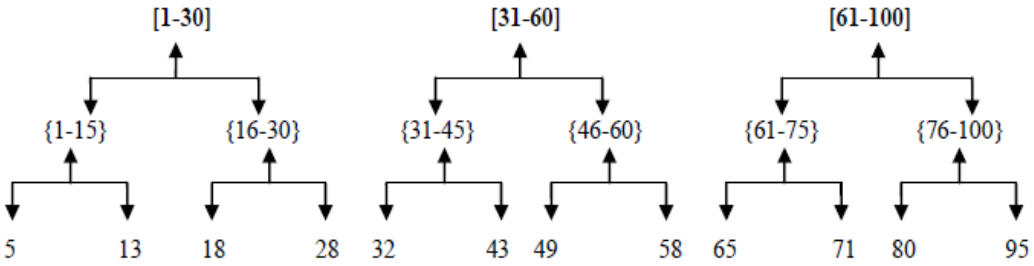
Generalization replaces a value with a "less-specific but semantically consistent" value. In a VGH, leaf nodes correspond to actual attribute values, and internal nodes represent less-specific values. For understanding Figure 3 contains VGHs for Quasi identifier or attributes DISEASE, GENDER and AGE. Generalization schemes can be defined based on the VGH that specify how the data will be generalized [7]. According to the VGH of DISEASE, we say that the value of disease is generalized according to the disease causes. Like "HIV" cause by virus so it can be generalized to "Diseases Caused by virus". The Gender hierarchy in the figure is generalized based on Male and Female category. The attribute Age is generalized to the interval (1-15) and (16-30), then to the interval (1-30) [5].



(a) VGH of Disease



(b) VGH of Gender



(c) VGH of Age

V. EXPERIMENTAL RESULTS

In this section, all the experiments were conducted on a system running with Windows 7 or higher Professional Edition operating system, with a 3:0 GHz Intel core 2 duo CPU, 2:0 GB main memory, and a 160 GB hard disk. The programs were implemented in JAVA as front end MS SQL server 2008 as backend. We implemented the proposed system for k-anonymous database has been anonymized using both suppression and generalization-based approaches, when we actually insert or update the database at that time carefully observe the value of parameter k because insertion or updation will depend on this k parameter. The data owner will able to see the inserted tuple in original datset by using MS SQL server 2008. We are tested several times the insertion of a new tuple in such anonymized databases with the values of parameter k equal to 2, 3,4,5, and 10.

We observe and perform actually insertion of record into database at that time our system calculates the average execution times means how many time spent in insertion or updation of database (calculated in milliseconds) of both suppression and generalization-based approaches. Our system also has a functionality to generate the graph of calculated average execution times in different chart style such as column, Line and Bar Figures 3 and 4 shows Execution times in suppression and Generalization based techniques respectively. In the figure it is noted that time spent by both techniques in testing whether the tuple can be safely inserted in the anonymized database decreases as the value of k increases.

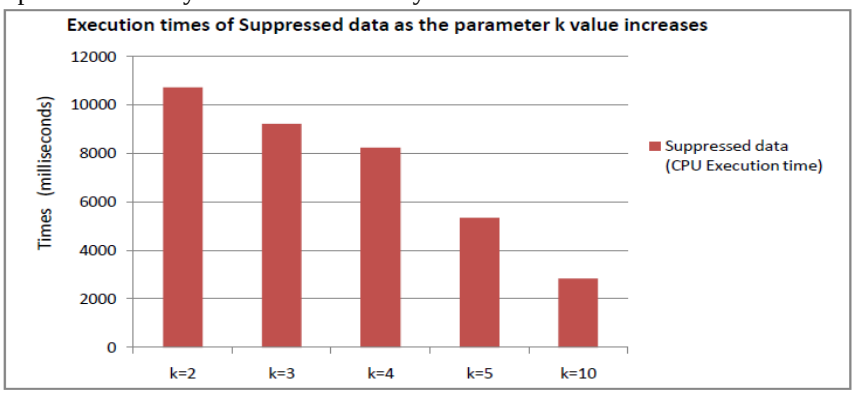


Figure 3: Execution times in suppression based techniques

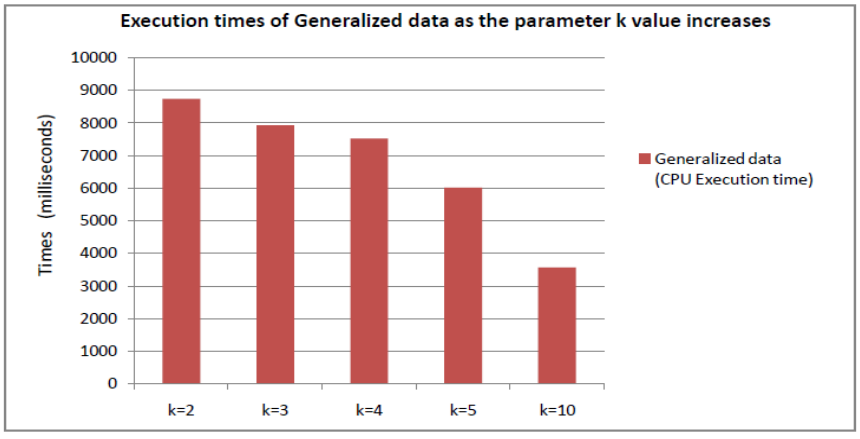


Figure 4: Execution times in Generalization based techniques

In a general view, the insertion of record having two separate background process of database k-anonymity checking and the actual update into two different phases, in the first phase, when a tuple is inserted into database then it check inserted tuple with existing data and whether the updated database is remains k-anonymous, then updation will occurred. In this checking process checker doont know the content of user's tuple. In second phase, system actually updates the database depends on the result of the anonymity checker. In some cases the insertion or updation failed in k-anonymous database then it waits until k-1 value becomes positive and other tuples fail the insertion.

VI. CONCLUSION

In this paper, we focus on a privacy preserving of *k*-anonymous database. We have presented two secure protocols suppression based and generalization-based for database anonymization techniques for protecting individual's privacy. In a anonymous database when new tuple is inserted then database owner privately check that whether anonymous database remains anonymity. Since the proposed protocols work perfect on updated database will definitely k-anonymous. Thus by using proposed protocol the database is updated properly. System every time check when new tuple inserted into database and if it satisfies k-anonymity then tuple is accepted for insertions otherwise it will prohibited. Suppressed the value of attribute by replacing '*' and Generalized the value with related possible general value to maintain the k-anonymity in database. Thus by making such k-anonymity in table it becomes complicated for third party to identify the record.

In future work, the following issues are noted.

- In private updation for database that supports better performance of anonymity different than k-anonymity.
- In case of unauthorized user, non-colluding third party, implementing a real-world anonymous database system.
- Increase the efficiency of communication protocols, in case of lots of communicating messages exchanged between the data provider and in terms of their sizes, as well.
- How to increase the efficiency of implementation and quality of the released output data in such a way to get the various requirements.

REFERENCES

[1] Mr. Mahesh T. Dhande, Mrs. N.A.Nemade, and Mr. Yogesh V. Kolhe, *"Privacy Preserving in K- Anonymization Databases Using AES Technique"*, International Journal of Emerging Technology and Advanced Engineering(IJEATE), ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 3, March 2013.

[2] Alberto Trombetta, Wei Jiang, Member, IEEE, Elisa Bertino, Fellow, IEEE, and Lorenzo Bossi, *"Privacy-Preserving Updates to Anonymous and Confidential Databases"*,. IEEE Transactions On Dependable And Secure Computing, Vol. 8, No. 4, July/August 2011.

[3] Benjamin C. M.,Fung, Ke Wang, Rui Chen, and Philip S. Yu. 2010. *"Privacy-Preserving Data Publishing: A Survey of Recent Developments"*, ACM Computing Surveys, Vol. 42, No. 4, Article 14.

[4] Neha Gosai, and S. H. Patil *"Security Preservation Methods to Confidential Databases"*, International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-1, Issue-4, April 2012.

[5] Xiaoxun Sun, Min Li Hua, and Wang Ashley Plank. 2008. *"An efficient hash-based algorithm for minimal k-anonymity* ", Australasian Computer Science Conference (ACSC2008), Wollongong, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 74.

[6] Benjamin C. M. Fung, "*Privacy-Preserving Data Publishing"*, Simon Fraser University 2007.

[7] Tiancheng Li, *"Privacy Preservation in Data Publishing and Sharing"*, Center for Education and Research Information Assurance and Security, Purdue University West Lafayette, August 2010.

[8] Behrouz Forouzan, *"Data Communications and Networking"*, TMH, fourth Edition.

AUTHOR'S PROFILE

Mahesh Dhande is a scholar of M.E. (Computer Science Engineering), at S.S.G.B.C.O.E.T. Bhusawal, under North Maharastra University Jalgaon (India). He is working as a lecturer in J.T. Mahajan College of Engineering, Faizpur.

Mrs. Neeta A. Nemade received her B.E. degree in Electronics and Telecommunication from J.T. Mahajan College of Engineering, Faizpur. And Master's degree in Digital Electronics at S. S. G. M. College of Engineering, Shegaon. She is currently working as Assistant professor in Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal.

