A Privacy Preserving Upload Model for Crowd Sourced Health Care Monitoring System

Umar Khalid Farooqui, Ajay Kumar Bharti

Abstract**: *In recent years governments become more concern about health care monitoring which cannot be accomplished without enhancing trust on underlying system as various citizens hesitate to upload their sample because of privacy reasons and obviously the governmental decisions are based on the data collected by various PHCs and third party medical agencies. The accuracy and authenticity of this third party owned data is always doubtful.***

Crowd sourcing(a collaborative framework) make its sound presence in development of large scale health projects Scientist also impressed from crowd sourcing which is a faster and better alternative to traditional methods for predicting and monitoring infectious diseases. However the success of this type of crowd sourcing depends on the trust on underlying system as the user is always looking firm commitment to preserve their privacy and win a promise of not being re-identified later.

Here in this work we suggest a privacy protecting framework for upload process which could fulfill user's diverse privacy requirements while guaranteeing the quality of health care data.

Keywords **: Privacy preserving, Cloud computing,Health Monitoring.**

I. INTRODUCTION

In recent years governments become more concern about health care monitoring which cannot be accomplished without enhancing trust on underlying system as various citizens hesitate to upload their sample because of privacy reasons and obviously the governmental decisions are based on the data collected by various PHCs and third party medical agencies. The accuracy and authenticity of this third party owned data is always doubtful

The inception of crowd sourced technologies help a lot in this regard as various citizens may upload their health information on to cloud based web application or a centralized server in a hassle free manner, on the other hand Government gains much concerned citizens data as a whole which will further be exploited for various policy matters and research works.

Crowd sourcing helps scientist to collaborate on large scale health project such as 'pandemics'. Experts also attracted towards crowd sourcing as a faster and better alternative to traditional methods for predicting and monitoring infectious diseases.

However the success of this type of crowd sourcing is depending on the trust on underlying system. The user is always seeking firm commitment to preserve their privacy and win a promise of not being identified /exposed at later stage.

Crowd sourced healthcare monitoring system utilizes omnipresent smartphone users to upload there health data for investigation and experimentation of various diseases and medicines.

It results in bringing new treatments to faster and also bridges gap between patient and healthcare provider.

Here we suggest a privacy protecting framework for upload process that could fulfill user's diverse privacy requirements while guaranteeing the quality of health care data. The quality of health care data depends on the number of uploads by its citizens. The greater number of upload by citizens leads to great quality of collected healthcare data. The decision of uploading by user process is designed in a bilateral goal escalation task (citizen anonymity and health record quality) which is designed as a game model of inadequate information , player can independently take decision for uploading samples or not to balance healthcare data quality and individual privacy requirement.

II. RELATED WORK

In collaborative Health monitoring model, users may upload samples via their smartphones in an undisclosed manner just to safeguard their Medical privacy. But undisclosed techniques are not adequate for this objective [6–8]. Montjoye et al. [6] examine a 1 yea 3-month sample of mobility trace for 1.5 million person and conclude that using four points we can adequately identify majority of them.

Also unidentification generally hide various available identifiers, demographic restrain and spatiotemporal property of the gathered specimen from an obscure mobile could be used for re-identification. Several approaches are made to decline the spatio-temporal affiliation for the adversary attack have been suggested [8]. Such strategies have been either listed as hierarchical or distributed. The approaches to centralization [9–11],. The downside of hierarchical strategies is their dependency on the reliable privacy server and the privacy of all connected users is compromised if a server is compromised [12]. Distributed attempts[4, 13, 14] non dependency on a server hosted database and allow users of the smartphone to decide when and how to publish samples. Mix-zone hide user identity as a distributed solution by allowing a user set to invade, alter alias, and quit a mix-zone so that no one can relate between their old and new alias. Palanisamy et al. [13] Suggested a mixed-zone model to safeguard the privacy of users traveling on the road. Liu et al. [14] Focused on the efficient employment of multiple mix-zones.

Revised Manuscript Received on February 25, 2020.
* Correspondence Author

Umar Khalid Farooqui *, Research Scholar Department of CSE,MUIT, Lucknow, India. Email: uk.farooqui@gmail.com

Dr Ajay Kumar Bharti, Professor2, Department of CSE,MUIT, Lucknow, India

It can rarely support traffic monitoring as users are unable to publish their locations before quitting a mix-zone. In [4], Hoh et al. Designed a model which define geographic-markers to indicate at what location vehicles should publish its location updates. Use of such markers improves tracking uncertainty to its maximum value and to avoid specific locations that are sensitive to privacy. Nevertheless, the identifiers can hardly satisfy the complex security standards of all users. The suggested solution allows consumers to manage their own privacy and accomplish a dual objective of performance assurance of health care and confidentiality of users.

In order to inspect strategic decision making while many players are participating (game theory) with varying objectives, various efforts could be recognized in adopting game-theoretical methods to analyze security and privacy concern in the mobile network [15–18]. Freudiger et al. [16] Study of disengaging act of mobile entities in a known location with a gametheoretic model. Yang et al. [17] advocated a true sell-off-based opportunity for mobile users to encourage as anonymous collection in order to achieve k-anonymity. Shokri et al.[18] Using the Stackelberg Bayesian-games model, privacy for location attribute of mobile users in the services which depends on location (LBSs) was tested. We follow an incomplete data game in this research to examine mobile user’s behaviors with bilateral motives (Medical Privacy and Health monitoring quality of service) in a crowd-sourced health care monitoring system, and sharing privacy.

III. SYSTEM DESIGN

In proposed system, each citizen is asked to periodically update his/her health data which can be used to assess the real time health monitoring of the citizens by a server. on the other hand the citizen can get new treatment s to market faster also predicting & monitoring of infectious disease become easier using crowdsourcing

In reality the accuracy a health monitoring of citizen, i.e. , QoS of Q of the health care service for specified time has dependency on the total number of users equipped with smartphone,also they have regularly published their health data.

Let us suppose a set of smart phone user P = {1, 2,3,-----n } in a group of citizens ready to provide the health data because they want to have a better Q, Normally citizens have different privacy levels

' C ' is the symbol used for depreciation in privacy due to a sample upload.

The accuracy of health care monitoring Q has dependency on the total number (k) of users equipped with smartphone, who have regularly published their health data. large k tends to larger value of Q ,we categories citizens in to following

a) non adult

b) adult

c) old age

the samples uploaded by non-adult are either supervised or not a real sample because of childhood.

Also the adults between 18 to 30 are likely to participate with full zeal where as adults between 31 to 49 are more concerned about their privacy on the other hand old age people are either reluctant or scared towards privacy and security concern.

IV. DESCRIBING THE PROBLEM

4.1 Quality of health monitoring

The reliability of healthcare monitoring has dependency on the total number (k) of users equipped with smartphone .Let Si is the user's upload strategy ' I ' with two options, upload (yes) or upload (No) , let Qi denote the accuracy of health monitoring in a citizen's community ' I ' that can be interpreted logarithmically as:

$Q_i = \log_\alpha \mathbf{(1 + K_i\,\beta)}$ **………..(1)**

Here α , β are ' system variables ' further term logα (1 + Ki β) represents the reduction of Qi's return on Ki. From an empirical study on Q, we can obtain α and β by utilizing voluntary nonlinear minimization upon real-world samples.

The above can also be written as

$Q = (\mathbf{1 + \beta}\ \sum_{i=1}^{n} I(S_i, Y)$ ……………**(2).**

Here , I (x, y)=1 if x = y and 0 otherwise. The purpose is to assure the user's upload strategy profile (S1,S2 …………Sn) in such a way that Q ≥ Qmin Where Qmin is the criterion for minimum service performance.

4.2 Medical Privacy

Many people have a real sense of privacy with respect to their body's exposure to others, this is a feature of personal chastity. Medical privacy help in the practice of maintaining the confidentiality and security of patient records .

The inception of ‘electronic medical records’ (EMR). And ‘patient care management system’ (PCMS) have emerge new concern of privacy, balanced with effort to reduce duplication of service and errors.)

medical records may comprises following:-

- Demographic details such as address, age, sex, and race
- Name and account number and sometimes Aadhaar No./patient ID
- Diagnosis, medications, diagnostic test results and prescriptions, medical history,
- Payment or accounting details

There are also pharmacy benefit managers (PBMs) that oversee health plans pharmaceutical benefit programs. PBMs have your entire medication history dose and who prescribed it, as part of their role is to check your eligibility and get medication approval. They also offer DEIDENTIFIED INFORMATION to data miners (not covered by HIPPA due to the removal of personal identifiable information).

The client can be re-identified by monitoring the user's mistake and ambiguity of identity. The adversary's aim is to extract from many users a sub portion of samples produced by the person / device, provided a series of samples mixed.

Attacker shall be prone to utilize ‘Qasi-Identifiers’ available with the sample(s) to perform **re-identification** of the user.

The attacker relates an earlier upload sample based on its estimation, or the most likely result, with the next one. Below is the formulation.

$\mathbf{arg\,max\ p(t \mid t_{i-1})}$ **………(3),**

Here $\mathbf{p(t \mid t_{i-1})}$ is the conditional probability and the probability of the next upload sample at t_{i-1} are described. The error of the attack is defined as the difference between the real time ti and its estimated value based on

$\mathbf{p(t \mid t_{i-1})}$ that is to be calculated using below mentioned sum:

$\Sigma x \,\hat{}\;\, p(t \mid t_{i-1}) I_z(t, t_i)$……………..(4)

To calculate identity inference ambiguity using distribution entropy $\hat{}\;\, p(P = ID_i|t)$:

$H = \Sigma\, i \;\hat{}\;\, p(P = ID_i \mid t) \log_2 1 /(\hat{}\;\, p(P = ID_i \mid x))$……………….(5)

Entropy H shows how difficult it is to identify an outcome IDi out of P at t . With higher entropy, the volatility of the opponent is higher.

Through aggregating (4) and (5), we obtain the user's uniform medical privacy immediately before making a decision as to whether or not to upload.:

$MP-_i = ½(H/\log_2 n + \Sigma_{x \in R} [I_z(t, t_i) \{p(t \mid t_{i-1})\}]$………………..(6)

Here publishing samples is counterproductive towards privacy as the opponent can easily obtain more information pertaining to users and get more correct inference results. Let us suppose user i, 0 < ci<1 upload value, then the user strategy will determine the location privacy level

$MP_i(s_i) = \{ MP-_i - c_i\; s_i = Y;$ ……………………….(7)
$\{ MP-_i ,\; s_i = N;$

For higher value of privacy $MP-_i$, lower chance for being identified, and finally it turns to lower the cost c_i.

4.3 Optimizing process

Desired minimum value for service quality is Qmin and the MPi privacy level of every user in a group is given. The optimization method is for figure out S={S1, S2 - - - -Sn } upload strategy profile to optimize the overall privacy level $\Sigma_i MP_i$ such that $Q \geq Q_{min}$.

It must address following two key points.

1) User may be unaware about other’s privacy level and obviously hesitate to upload due to high risk of privacy compromise while upload samples.

2) Ways for assessing the (Q_{min}) minimum requirement for service quality

For the primary issue , we use an incomplete data game model[20], in this every user has given a category 0, the probability density function f(0) of which depicts the user's privacy rate distribution. Every client is only aware of the distribution of privacy and does not know the actual level of privacy. We also use the global view of the database (i.e., historical information on citizens ' health status in the given sample) to determine the minimum requirement for quality of service.

V. UPLOAD GAME PHASE

In order to model smartphone users ' upload (playing / decision) phase, we use the incomplete data game model In this game, each player (citizen) combines their health information privacy (medical privacy) and health monitoring accuracy to decide whether to upload or not. Let us suppose a group of participant P={1,2,3,4 ----n},

This belongs to a specific group of smartphone users. There are two possible movements for the player: upload (y) or not (n). User upload game / decision Bayesian Nash equilibrium (BNE) , calculated by adding the average participation of ' y ' to that of ' n. '

The optimal solution for user I strategy is based on the quality of the health monitoring system and the user's level of medical privacy, and the usefulness of user I is defined as:

$U_i(S_i(\Theta_i), S_{-i}(\Theta_{-i}) = w\; Q_i(S_i(\Theta_i), S_{-i}(\Theta_{-i})) + MP_i\, S_i(\Theta_i)$ …………….(8)

Where Qi(Si, S-i) is the performance of the health care monitoring service measured by user i's different samples and their opponent ‘-i ’ , MPi(Si) is the user i's ' personal confidentiality. '

W can also be taken as the degree of expectation of Q users.

ϴi is the degree of privacy just before the match.

VI. THE UPLOAD METHODOLOGY

Our aim is to equipped participant with a suitable degree of privacy protection and further gain an overall optimum performance by person's "Health care monitoring system" and "Medical Privacy."

6.1 Algorithm for upload

The proposed privacy algorithm for the protection of medical data collection is described here, which uses an incomplete information game and ensures the upload data's k-anonymity, which consists of three phases.

6.1.1 The ‘k’ estimation phase :

Firstly the server estimates required uploads in numbers by utilizing the historical health data of citizen.

Here we present the functional relationship between the asked quality of health monitoring in a population and the historical computation of average number of patients ‘n’.

$Q(n) =(P_s / (\sigma \sqrt{}\, 2\, \pi)) e^{-(n-\mu)2/ 2\, \sigma 2}$ …………………..(13)

Where PS is population size, μ is the mean and **σ** is the standard-deviation, and ‘n’ is estimate of average patients based on past data (for given population). Also P_S>30 is our system requirement.

Further $k= 1/ \beta\, [(\alpha^{Q(n)} - 1) P_s]$ ……………..(14)

Here ' k ' is the system's number of required uploads and Ps is the population size.

Upload user Selection Process : (optimizer)

A user compute w and for that Nash Equilibrium can be obtained and further decide for uploading or not (based on ‘w’).

If the participant(user) is knowing the oponents ' upload cost, i.e. c1 ≤ c2 ≤ c3 …. Then it's easy to get w as a value:

$W= [C_k / (\log_\alpha(1+ \beta\, (k+1))] \; 1 /(1+ \beta K).$

User is unaware about privacy level and privacy cost of others due to incomplete information model, we need to compute the value of C_k.

We assume that $MP_i = \lambda / C_i$, The level of confidentiality also has distribution $f(\theta_i)$,we get

$k/m = \int f\, (\theta i\,) d\theta i\; 10$ ……………………..(15)

The value of w could be given as-

$W =[\lambda / (F^{-1} (1- k/m) \log_\alpha (1+ \beta\, (k+1))] \; 1 /(1+ \beta K))$……………….(16)

At any time m>k is the necessary condition for the operation of proposed system, where ‘m’ is the number of smart phone users in the given population and ‘k’ is required number of upload .

6.2 Privacy preserving upload Model

We devise a privacy preserving distributed upload model which first compute number of required upload at server and then the user is empowered to make decision to upload or not based on the computed value of ‘w’ , which is further depending on the upload cost and given that the user is totally unaware about privacy level and privacy cost of others ,thus we use an incomplete information game model ,and based on NE final computed value of w is obtained. Further the value to be uploaded are generally medical records/ health information of citizens and the user never want to be re-identified from these uploaded record ,thus we use upload model which works on the basis of GenReq algorithm and finally achieve k-anonymity of health records. Following figure depicts a glimpse of proposed model. The user first interact with the interface and initiate request at this time the request is passed through optimizer and directed to the server while passing Q(n), i.e. the asked quality of health monitoring in terms of average number of patients ‘n’in a population ,now the server computes value of k (i.e, required number of upload users) and send it back to the optimizer next the optimizer compute value of w based on the received value of ‘k’.finally based on the computed value of w the user is empowered to take decision to upload or not.

As the upload records are health information and are interrelated so there exist a strong requirement of assuring k-anonymity property and for that we use a distributed collaborative model which has no single point of failure.The working of distributed collaborative model is given below.

Firstly the MaxGen node receives a request r with attribute a_i from a user i, next it returns maximum generalized value of the supplied attribute and stored it in ***temp***.

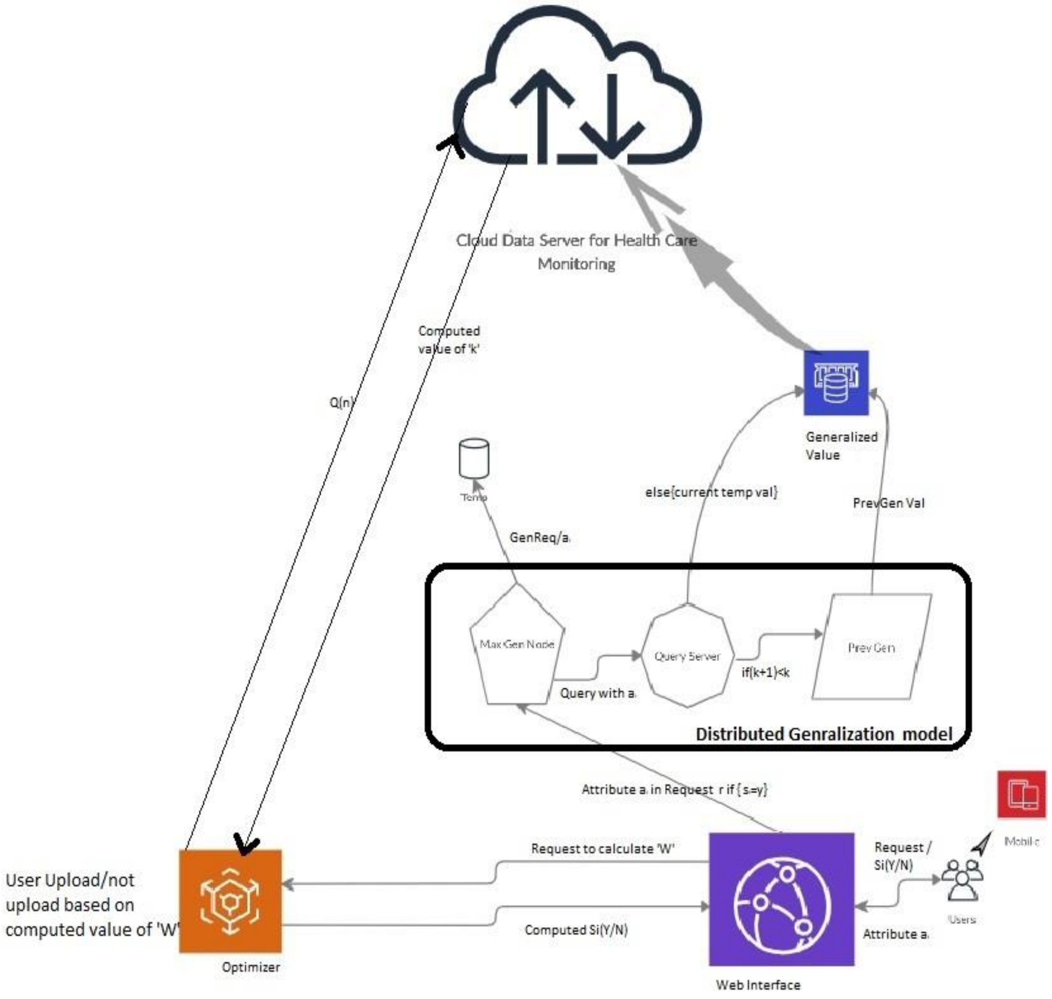


Fig 1: Optimal privacy preserving upload model

The Query Server queries peers of i with this value(a_i) to find the number of other cooperating users of i having the same attribute value in their database . Each peer simply responds with "yes" or "no" and finally Query Server aggregates the maximum number of positive responses and stores them in k. Further if k+1< k , it is an indication of not satisfying the k-anonymity requirement by the current value It would be reasonable enough to use the current time and the corresponding value.

The PrevGen Node response with the earlier generalized value from a generalized attribute sample.

The return value is the less generalized value that precedes the one passed as an argument, and if k+1<k then save it as a generalized query attribute.

Otherwise the value of ' current temp ' will be saved as the generalized query attribute.

For each attribute, it performs the loop operation in r and eventually forms the generalized request array r.

Finally these generalized requests are uploaded to the Heath care monitoring server generally hosted in cloud environment.

6.3 Algorithm for Suppression :

INPUT : request r = < a_1,a_2,………a_i>
OUTPUT : k-minimal suppressed values after eliminating Qasi Identifier r' = < a_1',a_2',………a_i'>

```
function findQasiName(fileName)
 [alldata]= xlsread(fileName);
//Read source file in to alldata
in={"name",”address”….etc};

//Qasi Identifiers
o=contains(alldata(1,:),in,'IgnoreCase',true);
item_FoundIndex=find(o);
length_Index=length(item_FoundIndex);
em='';
if(length_Index <=0)
display ('no name/address/PAN/Aadhar field found in database');
 return
end
 ls=length(alldata(:,1));
j=2;
for i=1: length_Index
   indx= item_FoundIndex (i);
 m=alldata(1,i);
 for u=j:ls
   if (isnan(alldata{u,indx}))
    display('blank cell found');
  else
  alldata(u,indx)= replace(cellstr(alldata(u,indx)),cellstr(alldata(u,indx)),'**');
            //Interchange data with a *.
   end
  end
end
n=m;
 xlswrite(file,alldata);     //write data into the file
end
```

6.4 Algorithm for Generalization :

INPUT : request r(set of attributes) = < a_1,a_2,………a_i>
OUTPUT : the k-minimal generalized values r' where r' = < a_1',a_2',………a_i'>

```
function listObject = linked(fileName)
[alldata]=xlsread(fileName);//Read database and store in alldata
 global l h;
in='age';
o=contains(alldata(1,:),in,'IgnoreCase',true)
item_FoundIndex =find(o);

     g=length(item_FoundIndex);

  ls=length(alldata(:,item_FoundIndex));
  j=2;
     values=(alldata(j:ls, item_FoundIndex));
   if( g <=0)
    disp('no age field found in database');
  return
  end

  data = reshape(values,[],1);
  j=2;
  values=alldata(j:ls, item_FoundIndex);
      if( g <=0)
    disp('no age field found in database');
  return
  end
   function k = quer(x,y) //function to find and count all other available //values
found in the range of x,y.
        k=0;
        for ( i=1:ls-1)
            m12 =data(i);
            if(data(i)>= x & data(i) <=  y)
               k=k+1;
                 end
                      end
        display(k);
    end
function [l,h]= max(x)         //function to find range from given age value
   m=1;

      s=x;
      y = (s/10);
    z=fix(y);
     disp(z);

     [a,b]=test(z);     //test is a function which returns range in a & b
     l=a;
     h=b;
     disp("function max ending");
     disp(l);disp(h);
  end
       end
  function  write(c,d)  //function to write generalize data

      val = string(int2str(c))+'-'+ string(int2str(d));
        v=str2num( val);

        firstRow = 0;
       lastRow = 0;
      firstCol = col(has);
      lastCol = firstCol;
      f=0;
       flag=0;

       for i=1:ls-1
                    if( data(i)>=c && data(i)<=d)
            flag=flag+i;
                              firstRow = i+1;

             disp(data(i));
            data(i)= str2num(val);
             disp(val);
              lastRow = i+1;
               cellRange =
[firstCol,num2str(firstRow),':',lastCol,num2str(lastRow)];
              xlswrite(file,cellstr(val),'Sheet1',cellRange); //write generalize
data

          end

       end
    end
end
```

VII. RESULT AND DISCUSSION

We propose a privacy preserving upload mechanism which provides a suitable degree of privacy protection of individual who participated in upload process and also gain an overall optimum performance by the “Health care Monitoring System”.

The solution utilizes benefits of Nash Equilibrium (uses an incomplete information game model) for this bilateral goal escalation process(i,e, Citizen anonymity and Health record Quality).

The user take upload decision based on the computed value of 'w' , which further depends on 'k'(number of required upload),the server comput it based on historical data of average number of patients in the given population & 'm'(number of smart phone users in the given population).

Also the upload decision is taken by user as follows:

```
If (w>1)
Then upload_decision='YES'
Else
Then upload_decision='NO'
```

The 'w' depends on the QoS(Q) , i.e,greater the quality greater the chance of upload decision ='YES'.

Further the upload samples are generally health records which may include individual's details along with Quasi Identifier(s), with which a person can easily re-identified by an adversary.To remove such quasi identifiers we use generalization and suppression algorithms which ensures k-anonymity of sample data.

The number of '*' introduced to the given data base represent the cost for K-anonymous solution. K-anonymity solution with a minimum cost suppresses the lesser number of cells needed to ensure K-anonymity. There is a strong possibility of a single point of failure to protect anonymity by using an anonymizer .We have shown that the end user privacy is preserved while quality of service(i.e,medical data/health monitoring service quality is also better,

We have shown that desired quality of service QoS(Q) depends on the number of upload(k) and further 'k' depends on population size(Ps).

We use suppression algorithm for eliminating quasi identifiers such as name,SSN,address DOB etc. and generalization for quasi identifiers like age.

For empirical study of proposed model we have implemented the shown algorithms in Matlab 2019 – a and applied 'Health care provider credential data' available for research work at following URL : https://healthdata.gov/dataset/health-care-provider-credential-data

This dataset carries thousands of records related to individuals along with multiple quasi identifiers. Following are the graph and snapshot of dataset before and after applying the solution.

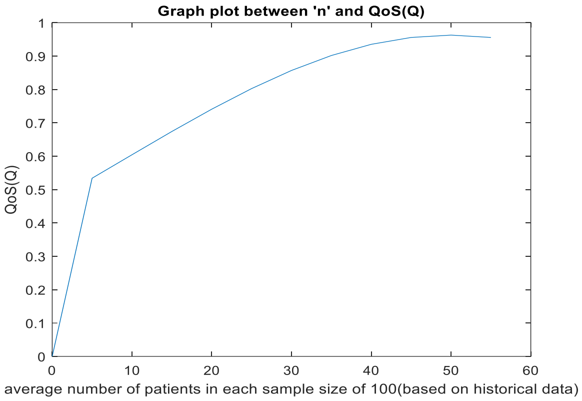


Fig2: Graph between QoS and 'n'

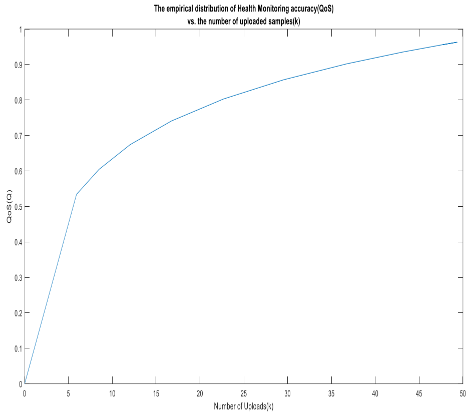


Fig3: Graph between QoS and 'k'

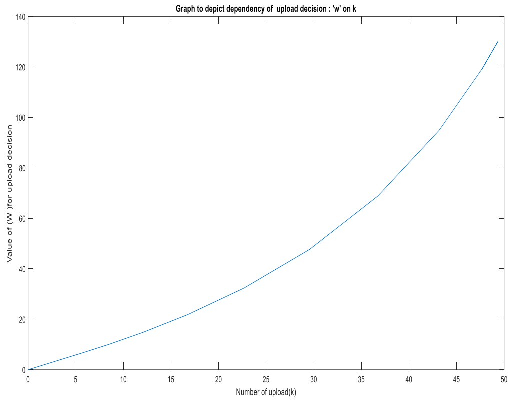


Fig 5: Graph between 'w' and 'k'

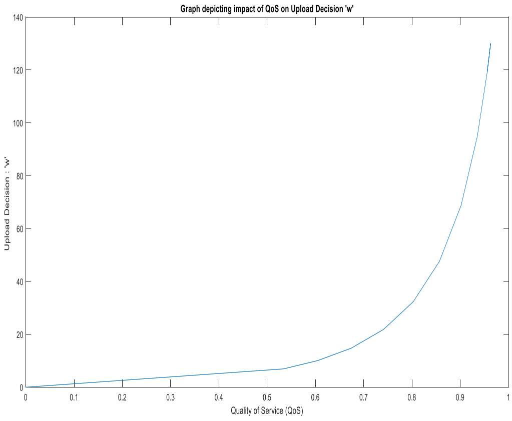
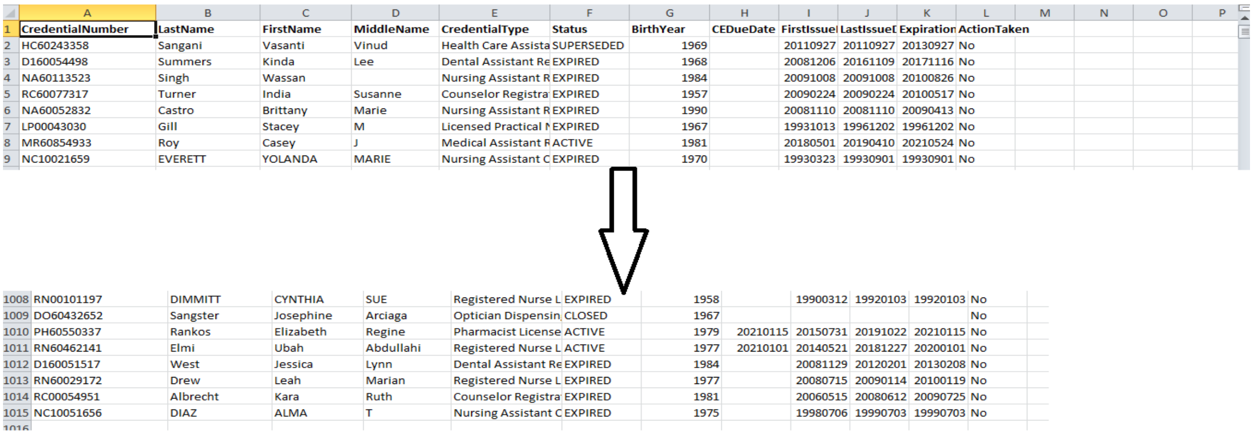
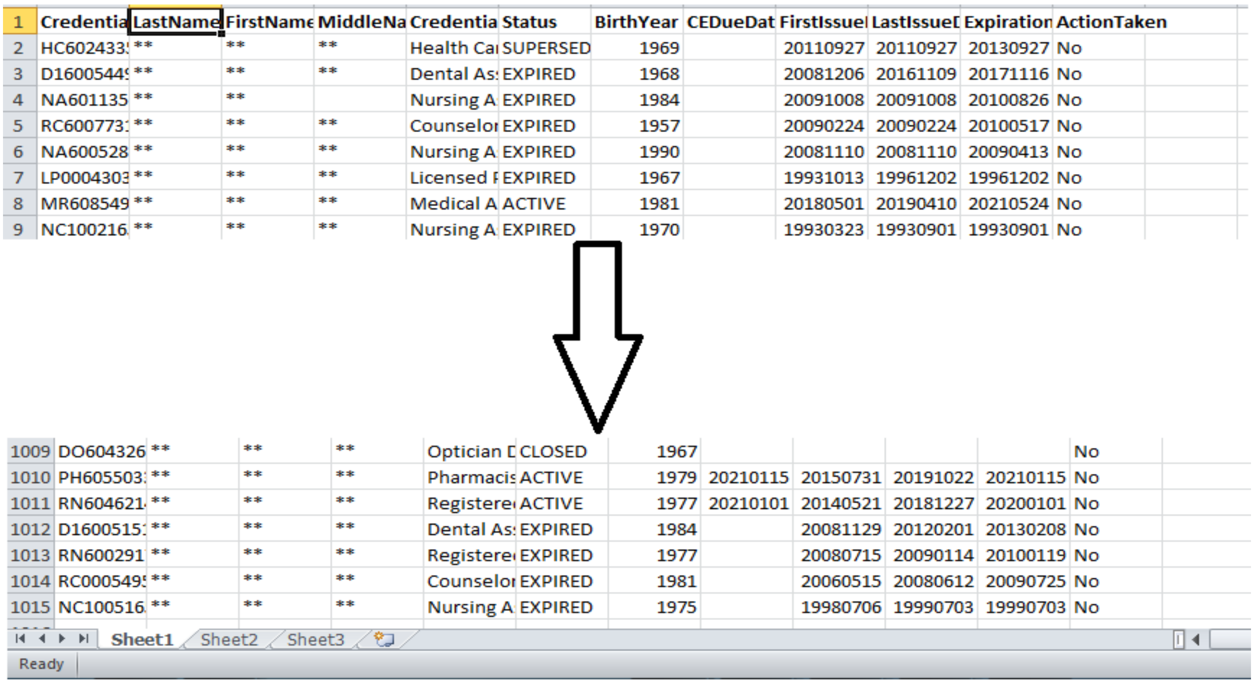


Fig4: Graph between 'w' and QoS



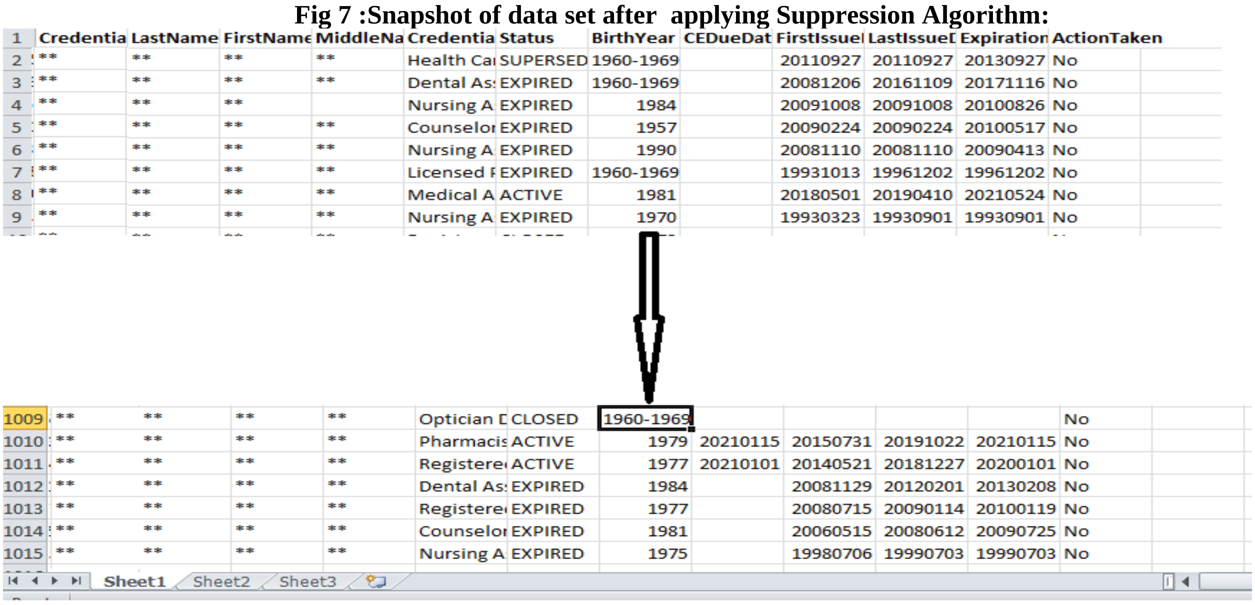
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	CredentialNumber	LastName	FirstName	MiddleName	CredentialType	Status	BirthYear	CEDueDate	FirstIssueI	LastIssueD	Expiration	ActionTaken				
2	HC60243358	Sangani	Vasanti	Vinud	Health Care Assista	SUPERSEDED	1969		20110927	20110927	20130927	No				
3	D160054498	Summers	Kinda	Lee	Dental Assistant Re	EXPIRED	1968		20081206	20161109	20171116	No				
4	NA60113523	Singh	Wassan		Nursing Assistant R	EXPIRED	1984		20091008	20091008	20100826	No				
5	RC60077317	Turner	India	Susanne	Counselor Registra	EXPIRED	1957		20090224	20090224	20100517	No				
6	NA60052832	Castro	Brittany	Marie	Nursing Assistant R	EXPIRED	1990		20081110	20081110	20090413	No				
7	LP00043030	Gill	Stacey	M	Licensed Practical N	EXPIRED	1967		19931013	19961202	19961202	No				
8	MR60854933	Roy	Casey	J	Medical Assistant R	ACTIVE	1981		20180501	20190410	20210524	No				
9	NC10021659	EVERETT	YOLANDA	MARIE	Nursing Assistant C	EXPIRED	1970		19930323	19930901	19930901	No				
1008	RN00101197	DIMMITT	CYNTHIA	SUE	Registered Nurse L	EXPIRED	1958		19900312	19920103	19920103	No				
1009	DO60432652	Sangster	Josephine	Arciaga	Optician Dispensin	CLOSED	1967					No				
1010	PH60550337	Rankos	Elizabeth	Regine	Pharmacist License	ACTIVE	1979	20210115	20150731	20191022	20210115	No				
1011	RN60462141	Elmi	Ubah	Abdullahi	Registered Nurse L	ACTIVE	1977	20210101	20140521	20181227	20200101	No				
1012	D160051517	West	Jessica	Lynn	Dental Assistant Re	EXPIRED	1984		20081129	20120201	20130208	No				
1013	RN60029172	Drew	Leah	Marian	Registered Nurse L	EXPIRED	1977		20080715	20090114	20100119	No				
1014	RC00054951	Albrecht	Kara	Ruth	Counselor Registra	EXPIRED	1981		20060515	20080612	20090725	No				
1015	NC10051656	DIAZ	ALMA	T	Nursing Assistant C	EXPIRED	1975		19980706	19990703	19990703	No				

. Fig 6 :Snapshot of data set before applying GenSup Algorithm:



1	Credentia	LastName	FirstName	MiddleNa	Credentia	Status	BirthYear	CEDueDat	FirstIssueI	LastIssueD	Expiration	ActionTaken
2	HC602433	**	**	**	Health Ca	SUPERSED	1969		20110927	20110927	20130927	No
3	D1600544	**	**	**	Dental As	EXPIRED	1968		20081206	20161109	20171116	No
4	NA601135	**	**		Nursing A	EXPIRED	1984		20091008	20091008	20100826	No
5	RC600773	**	**	**	Counselor	EXPIRED	1957		20090224	20090224	20100517	No
6	NA600528	**	**	**	Nursing A	EXPIRED	1990		20081110	20081110	20090413	No
7	LP0004303	**	**	**	Licensed P	EXPIRED	1967		19931013	19961202	19961202	No
8	MR608549	**	**	**	Medical A	ACTIVE	1981		20180501	20190410	20210524	No
9	NC100216	**	**	**	Nursing A	EXPIRED	1970		19930323	19930901	19930901	No
1009	DO604326	**	**	**	Optician D	CLOSED	1967					No
1010	PH605503	**	**	**	Pharmacis	ACTIVE	1979	20210115	20150731	20191022	20210115	No
1011	RN604621	**	**	**	Registere	ACTIVE	1977	20210101	20140521	20181227	20200101	No
1012	D1600515	**	**	**	Dental As	EXPIRED	1984		20081129	20120201	20130208	No
1013	RN600291	**	**	**	Registere	EXPIRED	1977		20080715	20090114	20100119	No
1014	RC000549	**	**	**	Counselor	EXPIRED	1981		20060515	20080612	20090725	No
1015	NC100516	**	**	**	Nursing A	EXPIRED	1975		19980706	19990703	19990703	No

Fig 7 :Snapshot of data set after applying Suppression Algorithm:



1	Credentia	LastName	FirstName	MiddleNa	Credentia	Status	BirthYear	CEDueDat	FirstIssueI	LastIssueD	Expiration	ActionTaken
2	**	**	**	**	Health Ca	SUPERSED	1960-1969		20110927	20110927	20130927	No
3	**	**	**	**	Dental As	EXPIRED	1960-1969		20081206	20161109	20171116	No
4	**	**	**		Nursing A	EXPIRED	1984		20091008	20091008	20100826	No
5	**	**	**	**	Counselor	EXPIRED	1957		20090224	20090224	20100517	No
6	**	**	**	**	Nursing A	EXPIRED	1990		20081110	20081110	20090413	No
7	**	**	**	**	Licensed P	EXPIRED	1960-1969		19931013	19961202	19961202	No
8	**	**	**	**	Medical A	ACTIVE	1981		20180501	20190410	20210524	No
9	**	**	**	**	Nursing A	EXPIRED	1970		19930323	19930901	19930901	No
1009	**	**	**	**	Optician D	CLOSED	1960-1969					No
1010	**	**	**	**	Pharmacis	ACTIVE	1979	20210115	20150731	20191022	20210115	No
1011	**	**	**	**	Registere	ACTIVE	1977	20210101	20140521	20181227	20200101	No
1012	**	**	**	**	Dental As	EXPIRED	1984		20081129	20120201	20130208	No
1013	**	**	**	**	Registere	EXPIRED	1977		20080715	20090114	20100119	No
1014	**	**	**	**	Counselor	EXPIRED	1981		20060515	20080612	20090725	No
1015	**	**	**	**	Nursing A	EXPIRED	1975		19980706	19990703	19990703	No

Fig 8 :Snapshot of data set after applying Generalization Algorithm:

VIII. CONCLUSION

Here we introduced an upload mechanism to preserve user's medical privacy in a crowd sourced health monitoring system. This approach is user centric and create an equilibrium between user's medical privacy and healthcare monitoring system .

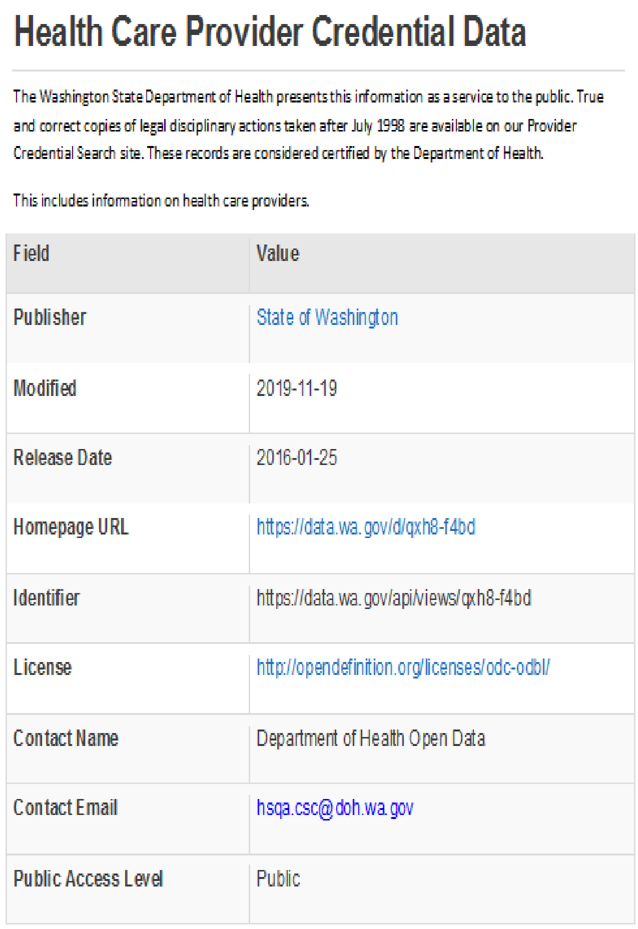
The Health care monitoring system quality depends on number of upload samples by various citizens/users in a population but the citizen hesitates in uploading due to high concern of privacy leak or re-identification. We address this issue by first calculating required number of upload which depends on number of patients in a population segment. Further the user is assured for required level of privacy by using user upload strategy based on game model which exploits benefits of nash equilibrium.

The health records are interrelated and sometimes subjective too ,also it has been learnt that the user may be re-identified at later stage by evaluating pair of values. We address the issue of k- anonymity of the uploaded health record by implementing PrevGen Algorithm and suppression algorithm in our model which results in generalized value saved into cloud database and thus assured k-anonymity of the health records .

In future we will look to implement the system in real environment and apply the real world data to assess its performance.

About Data Set:

URL : https://healthdata.gov/dataset/health-care-provider-credential-data



Health Care Provider Credential Data

The Washington State Department of Health presents this information as a service to the public. True and correct copies of legal disciplinary actions taken after July 1998 are available on our Provider Credential Search site. These records are considered certified by the Department of Health.

This includes information on health care providers.

Field	Value
Publisher	State of Washington
Modified	2019-11-19
Release Date	2016-01-25
Homepage URL	https://data.wa.gov/d/qxh8-f4bd
Identifier	https://data.wa.gov/api/views/qxh8-f4bd
License	http://opendefinition.org/licenses/odc-odbl/
Contact Name	Department of Health Open Data
Contact Email	hsqa.csc@doh.wa.gov
Public Access Level	Public

REFERENCES

1. Khalid Farooqui, Umar. (2018). A Comparative Study on Privacy Preserving Schemes based on Encryption Proxy and Cloud Mask. International Journal for Research in Applied Science and Engineering Technology. 6. 401-405. 10.22214/ijraset.2018.3064.
2. He, Yunhua & Sun, Limin & Li, Zhi & Li, Hong & Cheng, Xiuzhen. (2014). An optimal privacy-preserving mechanism for crowdsourced traffic monitoring. Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc). 2014. 10.1145/2634274.2634275.
3. Farzana Rahman,Hoque,Sheikh, ProQuPri, Proceedings of the 2011 ACM Symposium on Applied Computing,2011
4. Baik Hoh, Toch Iwuchukwu, Quinn Jacobson, DanielWork, Alexandre M Bayen, Ryan Herring, J-C Herrera, Marco Gruteser, Murali Annavaram, and Jeff Ban. Enhancing privacy and accuracy in probe vehicle-based traffic monitoring via virtual trip lines.Mobile Computing, IEEE Transactions on,11(5):849–864, 2012.
5. Yves-Alexandre de Montjoye, C´esar A Hidalgo, MichelVerleysen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. Nature srep., 3, 2013.
6. Chris Y.T. Ma, David K.Y. Yau, Nung Kwan Yip, and Nageswara S.V. Rao. Privacy vulnerability of published anonymous mobility traces. In Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking, MobiCom '10.ACM, 2010.
7. Laurent Bindschaedler, Murtuza Jadliwala, Igor Bilogrevic, Imad Aad, Philip Ginzboorg, Valtteri Niemi, and Jean-Pierre Hubaux. Track me if you can:On the effectiveness of context-based identifierchanges in deployed mobile networks. In NDSS. TheInternet Society, 2012.
8. Baik Hoh, Marco Gruteser, Hui Xiong, and Ansaf Alrabady. Achieving guaranteed anonymity in gps traces via uncertainty- aware path cloaking. Mobile Computing, IEEE Transactions on, 9(8):1089–1107,2010.
9. Mehmet Ercan Nergiz, Maurizio Atzori, Yucel Saygin,and Bar Guc. Towards trajectory anonymization: A generalization- based approach. Trans. Data Privacy, 2009.
10. Karen P Tang, Pedram Keyani, James Fogarty, and Jason I Hong. Putting people in their place: an anonymous and privacy- sensitive approach to collecting sensed data in location-based applications.In Proceedings of the SIGCHI conference on Human Factors in computing systems, pages 93–102. ACM,2006.
11. Elaine Shi, Richard Chow, T h. Hubert Chan, Dawn Song, and Eleanor Rieffel. Privacy-preserving aggregation of time-series data. In In NDSS, 2011.
12. Balaji Palanisamy and Ling Liu. Mobimix: Protecting location privacy with mix-zones over road networks. In Data Engineering (ICDE), 2011 IEEE 27th International Conference on, pages 494–505. IEEE, 2011.
13. Xinxin Liu, Han Zhao, Miao Pan, Hao Yue, Xiaolin Li, and Yuguang Fang. Traffic-aware multiple mix zone placement for protecting location privacy. In The 31st IEEE International Conference on Computer Communications (INFOCOM 2012), 2012.
14. Tansu Alpcan and Sonja Buchegger. Security games for vehicular networks. Mobile Computing, IEEE Transactions on, 10(2):280–290, 2011.
15. Julien Freudiger, Mohammad Hossein Manshaei,Jean-Pierre Hubaux, and David C Parkes. On non-cooperative location privacy: a game-theoretic analysis. In Proceedings of the 16th ACM conference on Computer and communications security, pages 324–337. ACM, 2009.
16. Dejun Yang, Xi Fang, and Guoliang Xue. Truthful incentive mechanisms for k-anonymity location privacy. In INFOCOM, IEEE International Conference on Computer Communications, pages 3094–3102, 2013.
17. Reza Shokri, George Theodorakopoulos, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. Protecting ocation privacy: Optimal strategy against localization attacks. In Proceedings of the ACM conference on Computer and Communications Security (CCS), pages 617–627.ACM, 2012.
18. Mudhakar Srivatsa and Mike Hicks. Deanonymizing mobility traces: Using social network as a side-channel. In Proceedings of the 2012 ACM conference on Computer and communications security, pages 628–637. ACM, 2012.
19. John C. Harsanyi. Games with incomplete information played by "bayesian" players, i-iii. Manage. Sci., 50(12
20. Supplement):1804–1817, 2004.

AUTHORS PROFILE

Mr. Umar Khalid Farooqui pursed Bachelor of Science from KNIPSS, UP in 1998 and Master of Computer Application from Agra University in year 2002. He is currently pursuing Ph.D.from MUIT,Lucknow since 2015. He has published more than 10 research papers in reputed international journals and National/International conferences and it's also available online. His main research work focuses on Privacy preservation schemes and techniques, Cloud Security and Privacy, Cloud Architecture,. He has 15 years of teaching experience and 4 years of Research Experience.

Mr Ajay kumar Bharti is professor and dean in the faculty of computer science,MUIT,Lucknow ,u.p,India.His research interest is in SOA,ICT and E-Governance,cloud computing.He has published numberof research papers in International journal and confrences. He has also reviewed various reputed journals.He has more than 15 years of teaching and 6+ years of research experience.