International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-9, July 2020

# A report on Privacy Preservation Methods on Big Data, Social Network Data, Medical Data.

**Guna Sekhar Tirumalasetti, Bhargava Sai Sathvik Gontla, Vijayakumar Kuppusamy**

***Abstract:*** ***Now a days Privacy-preserving of data is an essential philosophy in the digital computer era. As the leakage of preserved data has been increasing in the daily life. In this survey paper we have collected the different types of privateness maintenance in the view of big records of data , clinical and social networking data. In this paper we had explained deeply about the privacy preserving methods on these three types of data. We also gathered the various implementations which are implemented by various researchers until the present trend. In this survey paper we have given a clean view on the kind of arts of the secure and privateness maintaining techniques of big analytic data, in medical field data, and in social networking data. We also mentioned about the future works to be done for preserving the data. In now a days the attackers who stole the preserved data has been increased so we should increase the security levels of privacy gradually. In this survey paper we had mentioned what are the infrastructures that are preserving based on the type of data.***

***Keywords:*** ***Sensitivity, touchy degree, clustering, ppdp, bottom-up generalization, social network facts, privateness assaults, anonymized graphs, privacy retaining, records privacy, access manipulate, blockchain, encryption, medical data, privacy, security.***

## I. INTRODUCTION

As a piece of information sharing through web each association circulates the individual information from the data accumulated from various clients. The circulated information might uncover individual personal data. This information gathered from various organizations, governing bodies and individuals will make high possibilities of singular data dynamic. With regards to the common points of interest or by some rules that be required to convey the information, there is interest in exchange or circulation of information between various gatherings. Individual information is in the genuine structure, be that as it may, usually contains singular sensitive information besides, If the information distributed as it is then that sort of information will abuse the people protection. Expelling personality information may not be adequate to guarantee singular security. While dispersing the information, associations ought to know about other data sources. To strengthen clinical information administration, Privacy assurance rules, for example, The Health Insurance Portability and Accountability Act (HIPAA) in the US.

**Revised Manuscript Received on July 30, 2020.**
***** Correspondence Author
**Guna Sekhar Tirumalasetti*,** School of Computer Science and Engineering, Vellore Institute of Technology, Vellore
**Bhargava Sai Sathvik Gontla,** School of Computer Science and Engineering, Vellore Institute of Technology, Vellore
**Vijayakumar Kuppusamy,** School of Computer Science and Engineering, Vellore Institute of Technology, Vellore



Disciplines for events of human services information penetrate. Then again, the hour of conveyed figuring and enormous data necessitates that scientific statistics to be exchanged between different customers including their associates to permit examination, so higher human administrations company and, many new remedy plans can be given. Social networks destinations, for example, Facebook, LinkedIn are altogether changing the way in which people participate. They have become an overwhelming strategy for interfacing, associating, passing on and exchanging information on the web. People use relational association districts for finding elderly individuals, finding new individuals, or searching individuals who has comparable interests or issues in politics, monetary, including geographic edges. By making individual data outlines that consists of data, for instance, images, sound records, social network areas license customers for relating, post textes, send messages and textes to each other. As a social network regions is going on creating in number, the holder hoard striking measure of data about OSN CLIENTS

As the social network destinations gather and store information from administration. These social network information offer the private information to different outsiders. As the accumulated data much of the time contains senitive data, network administrators may leak anonymized and sterilized types of absolute social network outline or the subgraph to outcast clients, for example, advertisers, publicists, sociologists, sickness transmission pros, and restorative administrations specialists. The web social network clients make outlines and share texts using worked in gmail or texting. The outline consists lot rich individual data which could remarkably recognize them (e.g.,name, gmail, spot of current work and occupation )Quasi recognizable data (e.x., individual house, qualification or past working spot and title), segment data(e.x.,year of birth and sex), similarly as delicate data that might be stayed away from open perceivability (e.x., pay, marital status, religion, political view). There is some similarly good game plan of information made up of the correspondences among various clients. The web social network clients make outlines and texts using worked in gmail or texting. The outline has lot of good individual data that can exceptionally recognize them Quasi recognizable data (e.x., individual habitation, enlightening establishment or past work spot), segment data (e.x., date of birth and sex), similarly as delicate data that might be kept away from the open perceivability (e.x., compensation, marital status, religion, political life).

There are similarly a good course of action of data made from the interchanges between various client

## II. PRIVACY PRESERVING IN BIG DATA

Organization as a piece of data sharing through web distributes individual information.

This data may disclose personal data. It is the responsibility of the information supplier to create strategies and devices for distributing information in progressively Antagonistic condition this provides individual privacy and data will be completely utilized. This process is called **Privacy Preserving Data Publishing (PPDP).**In the process of PPDP removing the identity information is not enough for individual privacy .Organization should be the aware of publicly available datasources. As the attackers may not just focus on the distributed table alone they focus on connecting more tables to uncover individual from the distributed information.

### 2.1 Related Work

A lot of algorithms were introduced for PPDP to ensure individual privacy. Of these algorithms one is the K-anonymity which makes sure that published data consists of attributes with least of k comparable records. In any case, it has no significance for the affectability of the qualities which brings about the trading off of the protection in the majority of the scenarios. K anonymity is base for some explores. A considerable lot of the calculations were distributed for PPDP. Fung at-el.Introduced an algorithm for cluster analysis by extending K-anonymity . Privacy is gained by partioning the original data into cluster and class labels and K-anonymity is done by encoding these cluster information. Incognito is another algorithm proposed by Lefevre et al is a set of multiple bottom-up-generalization algorithms. In this strategy each conceivable k-anonymous full space speculations are created. In speculation parent esteems are subbed with child esteems to give protection. Wang et al .To address the productivity issue of K-anonymity proposed Bottom up generalization. Manchajjhala et al.proposed l-diversity. Mohd et al.For online e-learning activities to be trusty for the individuals. Learners privacy is protected by **Identity Management (IM)**.IM provides privacy with some level of member anonymity and pseudonymity. A members can have various personalities or may receive pseudomonas people, a notoriety transfer(RT) starting with one individual then onto the next individual is required .A notoriety move model must safeguard protection and furthermore forestall linkability of students character and people .This is **Privacy Preserving Reputation Management(PPRM)** permits a safe Reputation move. Emiliano Introduced idea called humming bird creature secures tweeter contents, fans interests, and furthermore hash tags(#) from assulters out the brought together server. The idea of protection is wide spread over various zones like **Wireless Sensor Networks(WSN)** In which security is fundamental concern.No wire itself is cannot be trusted where anonumous hubs can likewise be associated. A made sure about technique called Three-factor client verification is proposed for conveyed WSNs. To prevent impacts in correspondence with Privacy Preserving Data Mining. Drushina. Introduced a system coding strategies by expelling the factual reliance among approaching and active messages so to forestall following. Valeria presented edge relapse calculation for AI activity. This calculation takes an enormous no of information focuses as information and serches the best fit direct bend through these focuses. Xiaokui presented a Privacy preserving **data leak detection.(DLD).** In this an uncommon arrangement of sensitive information digest is utilized in detection. It empowers the information proprietor to securely designate the recognition activity to less genuine supplier by not releasing the sensitive information to supplier. Huang. introduced security model known as (v, l)- anonymity (namelessness), this focuses for the most part on the vulnerabilities of affectability. It follows already existing security models and gives the other method of protection. It is proposed another strategy for doling out the sensitive levels to the delicate qualities. They characterize an affectability characterization and derived a measure which is called as levels of sensitive values (LSV).Which is a measure to figure the sensitive levels. This proposal can likewise work effectively with various qualities. Qinghai introduced a protection maintaining information distributing strategy, which is MNSACM, for distributing smaller scale information with different numerical delicate properties with the thoughts of bunching and Multi-Sensitive Bucketization (MSB). Sweeney tested utilizing k-obscurity for recognizing the different potential assaults by a considered staggered datawarehouses. Tsai led concentrates on the information examination from the conventional information investigation to the recent large information investigation. Zhang gave a review of large information handling frameworks, for example, stream, chart, and AI preparing and furthermore some conceivable future work directions were examined.

### 2.2 Methodology

Privacy preserving of big data contains three modules such as

- Sensitive level calculation
- Nearest Similarity based clustering
- Bottam up generalization

Over all process contains five-steps (fig.1.1)

As appeared in Fig 1.1 procedure comprises of 5 stages. In the initial step miniaturized scale information is input. In sync 2 affectability levels will be determined and refreshed in small scale information table. And afterward Bottom-up generalization be applied on this refreshed data. In sync 3 (v, l)- anonymity condition will be tested in every identicalness class. At that point in the accompanying advance. On off chance that this condition is fulfilled, at that point anonymized information tables will be distributed in any case the data again goes under goes speculation process for the following level and the condition for anonymity will be tested.

*2.2.1 Sensitivity level calculation*

Here in the segment w'll talk about how to decide the degree of affectability and to pick elements to ascertain the sensitive level. We have considered the patient informational index, which has the data about the particular patient information containing the illnesses they suffered.in this, 'illness' is the touchy trait (sensitive attribute), so the classifcation is on diferent sicknesses dependent upon seriousness. For instance, HIV and temperature are the two diseases, the seriousness of HIV is more. Trough this classification which depends on seriousness, one can characterize and arrange HIV and temperature.

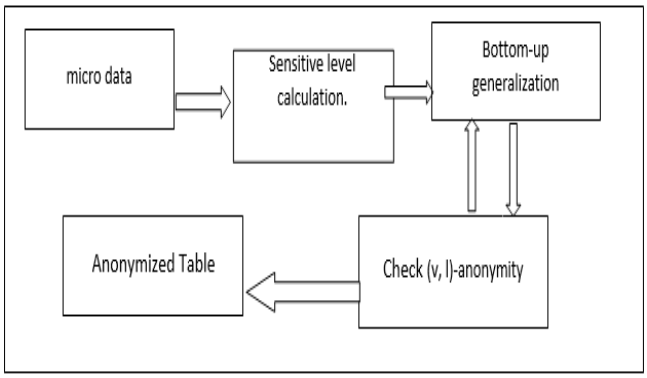


**Fig 1.1 (flow diagram for privacy preserving big data)**

*2.2.2 Nearest similarity based clustering*

Clustering is a procedure of apportioning more information into sub-classes, called clusters. A group of information

articles could be treated in one group. While examining the clusters, one always frst parcel the arrangement
information into clusters depending on the data including afterward dole out names into clusters. The principle bit of leeway of grouping upon classifcation is it is more versatile to changes and assists single with excursion valuable highlights that recognize diferent groups.

*2.2.3 Bottom up generalization generalization*

restores the most specifc esteem with the very summed up esteem. generalization is very famous technique to make the statistics unknown to present the protection. This technique which is summing up statistics is based upon the concept of the information and programs. it's far a various leveled system, which is spoken to in a type of 3. the node on the pinnacle level can be known as determine. all the relaxation of the hubs are baby nodes. generalization replaces baby node esteems with their figure node esteems in a scientific class tree. This turn around procedure of generalization is called as specialization. We go with the bottam up generalization technique, where the generalization procedure begins at base and kept on fixing this scientific categorization tree.This implies the last nodes be supplanted with the prior hub, in light of degree of generalization. This generalization procedure may halted after frst level, on the off chance that the level satisfes the mysterious conditions to recieve the necessary degree of protection. Generalization procedure will be performed on the quasi identifer set, which will leads to re-identifcation else connecting attacks. Not so much as generalization yet any anonymization strategies apply on these identifer traits. In the informational collection date of birth, PinCode, Race etc are some of the semi qualities.

*2.2.4 Testing for (v,l)- anonymity*

(v,l)- anonymity is most ideal approach to manage sensitive vulnerabilities whilst distributing information.In this 'v' speaks to the sesitive value, 'l' speaks to the delicate level. Each and every cluster contain 'v' number of wonderful properly represented sensitive ranges and 'l' exceptional nicely represented touchy stages. for example, a desk can be referred to as as (4,2)- nameless,if and simply if every organization or equivalence magnificence consists '4' different sensitive qualities and '2' particular sensitive levels. In patient informational index, illnesses are the touchy qualities and sensitive traits are calculated inside the past segments. So that sickness trait must contain '4' distinctive very much spoke to diseases and '2' diverse all around spoke to sensitive qualities these sensitive levels (SLs) so as to get the (4,2)- anonymization. Above all else the cluster must have least of 4 qualities, yet this table become failed there itself, and furthermore the subsequent cluster doesn't contain two diverse sensitive values, and this if of fundamental concern.

So the generalization procedure proceeds with straightaway to further level of speculation. Let the A1,A2,A3,…Am clusters frames after level one speculation utilizing Nearest similarity Based clustering.

'Am' is the last group. 'v' speaks to the delicate qualities, 'l' speaks to the sensitive levels. 'k' speaks to present value of the cluster or group. countf, Countg, Counth, Countc, Counthi, Counta speak to recurrence of sicknesses Temperature, influenza, gas, coronary illness, malignancy, Human Imuno Deficiency Virus separately, and also be increased each time they experiences the particular ailments. Countercheck is utilized to include the general various ailments in singular group, and also be increased every time when there is a general tally of an illness more prominent than 0 of every a bunch. count1, Count2, Count3, Count4 speaks to the frequencies of sensitive level 1, 2, 3, 4 individually. Countcheck1 is utilized as comparable as count check for the sensitive levels.

## III. PRIVACY PRESERVING IN NETWORK DATA MODEL

The presentation of online social networks (OSN) have changed the manner in which individuals interface and cooperate with one another just as offer data. OSN have prompted a gigantic blast of the network-driven information that can be collected for good comprehension which is intriguing wonders, for example, sociological and conduct parts of people or gatherings. Subsequently, online social network administration administrators are constrained to distribute the social network data for use by outsider purchasers, for example, scientists and publicists. As social network data distribution is defenseless against a wide variety of re-identification proof and exposure assaults (disclosure), creating protection saving components is a functioning examination zone. This paper presents a thorough review of the ongoing improvements in social networks data distributing protection dangers, assaults and security safeguarding procedures. We overview and present different sorts of security assaults and data misused by foes to execute protection assaults on anonymized social network data. We present an out and out overview of the best in class security protecting strategies for social network data distributing, measurements for evaluating the obscurity level gave and data misfortune just as difficulties and new research bearings. The overview assists perusers with understanding the dangers, different security protecting components and their vulnerabilities to protection penetrate in social network data distributing just as watch basic topics and future bearings.

### 3.1 SOCIAL NETWORK DATA PUBLISHING

Here we present social network information and survey a portion of striking properties of diagrams generally utilized by a foe in submitting security based assults on published social network data graphs. All through the paper, we consider chart and network, vertex and hub just as connection and edge conversely.

### 3.2. Social community information platform

The online social network clients make profiles include trade information utilizing builtin gmail or texting. This outline has a plenty of rich individual data which could be remarkably identify them(e.x., names, gmail-ids, spot of work and employment),known as quasi(semi)identifiers data (e.x., street number, instructive foundation or previous work spot and title), segment data (ex.date of birth and sex), just as touchy information is avoided the general visibility [1]. This is likewise a good arrangement of data produced from associations among various clients. This info might likewise incorporate sensitive data, for example, client shopping propensities. OSN clients are progressively utilizing cell phones to interact with social network in near future intensifying protection and concerns of security[2].

**High degree risk analysis framework**

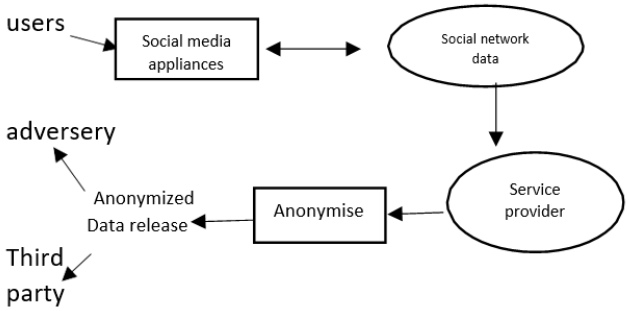


**Fig.2.1 (danger analysis frame work)**

The social networks administrators gather and store data from the administration clients to impart it to a wide assortment of outsider shoppers, for example, scientists to contemplate illness proliferation. A small suevey on social network information sharing is available in[3]. As the gathered information frequently contains sensitive data, administrators of the network may release anonymized and cleaned variants of total network chart or a subchart to the outsider clients, for example, publicists, advertisers, sociologists, disease transmission specialists, and medicinal services experts. The foes are likewise accepted to approach the distributed social networks data.

### 3.3 Social community statistics version

most of the component, social network records is inside the model non reflexive graph $G$= $(V, E)$, where $V$ = {$v1$,$v2$, … , $vn$} is a set of $n$ vertices and $E$ = {$e_{ij}$ = ($v_i$, $v_j$) | $v_i$ ,$v_j$ ∈ $V$, $i \neq j$} is the orders of edges between the $n$ vertices(nodes) of the graph$G$. The two Vertices(nodes) $v_i$ ∈ $V$and $v_j$∈ $V$are told to be adjoint(adjecent) if they have an edge in comman. The edge connecting two nodes $vi$∈ $V$and $v_j$ ∈ $V$ is called as an incident to $v_i$∈ $V$ and $v_j$ ∈ $V$. An edge which is pointed to the same node or vertex on either sides is called as loop . The edges(two or more) $e_i$ ∈ $E$ and $e_j$∈$E$ connected to the same set of hubs (and pointing within the identical course if the graph is directed) are called asparallel edges.

Example of a graph is:

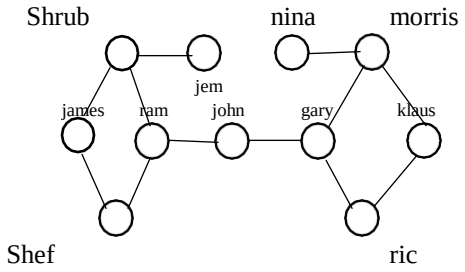


**fig.2.2(network graph)**

Now the anonymized graph for this graph is

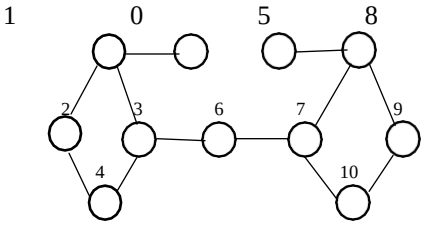


**Fig.2.3 (anonymized network graph)**

This is social network anonymization.

### 3.4 Degree based attacks

In the diploma-primarily based assault, the attacker queries the anonymized graph $G$ for a vertex $v$∈ $V$( $\bar{G}$) with the $deg(v)$ with the intention to re-identify the target character. to demonstrate the diploma- primarily based assault proposed[4], bear in mind the anonymized graph $\bar{G}$= $(V, E)$. allow the goal man or woman be gary inside the original graph shown in fig. 2.2 (vertex #7 in fig). the adversary has previous expertise that ′deg($G$ary) = four′. the adversary queries the anonymized graph $\bar{G}$ for a vertex $v$ $V$( $\bar{G}$) with $deg(\bar{v})$ = four response to the question can be $v$ because the handiest. vertex with degree $deg(v)$= 4 in graph $\bar{G}$ of entry to adversary is aware of that gary has 4 pals (degree) in authentic graph and have to correspond to gary within the authentic graph. for this reason the adversary can uniquely re-identify gary from the anonymized graph

**Definition:**

an enemy with vertex degree foundation statistics is stated to have re-recognized an objective vertex $v$ ∈ $\boldsymbol{V}$ from a allotted interpersonal enterprise diagram iff a foe can understand a verteX $u$∈ $V$( $\bar{G}$) with excessive degree of likelihood from the anonymized chart $G$ to such an volume that the diploma records of vertex $u$ exactly coordinates that of vertex $v$.

### 3.5 Neighbourhood assault fashions

The attacks attacks[5,6,7,8] abuse the pals statistics of an objective vertex $v$ ∈ $V$ ($G$) in a diagram interior $r$-jumps in which $r \geq 1$. in this manner, it's miles typically expected that a foe have earlier statistics at the nearby statistics of a few objective people $t$ ∈ $V$ inside the first diagram $G$ = $(V, E)$. the enemy mounts a re-identifing evidence assault on an anonymized diagram $G$ = $(V, E)$.

**Definition:**

An attacker with the sooner statistics on any vertex $v \in V$ community records is stated to have re-identified a vertex $v \in V$ $(G)$ iff a foe can entirely distinguish a vertex $ui \in V(\bar{G})$ with the give up goal that the local records of vertex $uj \in V(\bar{G})$ exactly coordinates that of vertex $v \in V$ $(G)$ in the first chart $G = (V,E)$.

### 3.6 Network topology assault models

The gadget topology assault models[9,10,11] speak to the class of assaults that utilization an assistant interpersonal employer datagraph and a couple of beginning mappings (seed vertices) to mount an assault on the anonymized social community graph. This method misuses the way that a solitary individual the has club on more than one social network sites to mount re-identification attacks within the anonymized community. An adversary has get right.to $G = (V, E)$ (reference or auxiliary) graph and $G = (V, E)$ (anonymized aim) graph. generally, these attacks start with regarded mutual vertices within the two networks after which expand the wide variety of re- recognized vertices with the aid of evaluating and matching in every network.

### 3.7 Anonymity stage and facts

loss maximizing privateness of the social network users even as on the equal time minimizing community statistics loss is the main objective in privacy keeping facts publishing. a ramification of metrics to quantify the privacy stage and the extent of statistics loss after a graph is anonymized had been used in the literature. in this section, we overview a number of the not unusual metrics used and gift a coherent classification of numerous metrics used to measure the anonymity stage and statistics loss because of anonymization.

### 3.8 Data loss metrics

on this segment, we gift various metrics used to assess the volume of records loss a graph stories due to an anonymization method. the graph structural facts loss difficulty can be informally stated as follows: trouble 2: given an unique graph $G = (V, E)$ and an anonymized model $G = (V, E)$, the assignment is to determine: (i) the extent to which the structure of the authentic graph remains intact in $G = (V, E)$; or (ii) the amount of efforts had to reconstruct the structural properties of the original graph from the anonymized graph. normally, the application of the anonymized graph is measured in phrases of the way plenty the structural attributes of the authentic graph is maintained in the anonymized graph or how a good deal effort is wanted to reconstruct the structural homes of the authentic graph from the anonymized graph. the same old approach to cope with the latter is through two step process[12,13].

(i) sampling to achieve close versions of the authentic graph from the anonymized graph, and (ii) investigate estimated values of the primary properties of the original graph from the pattern graphs

### 3.9 Privacy Preserving Mechanisms

publishing naively anonymized network graph is vulnerable to a giant privateness risks. in this segment, we gift a evaluation of the art of privacy keeping approaches.

### 3.10 Non-perturbation privacy upkeep models

On this section, we gift a -stage taxonomy of privateness retaining tactics. on the top degree, we categorize the anonymization strategies into 4 instructions: random graph editing techniques, probabilistic graph enhancing strategies, k- anonymization techniques, and generalization through clustering of vertices. given that privateness retaining techniques depend upon the type of attacks they are intended to mitigate, the second level of the taxonomy consists of numerous anonymization techniques classified along the type of attacks they intend to mitigate.

### 3.11 Neighbourhood anonymization strategies

the motive of the neighbourhood anonymization methods is to save you an adversary with prior knowledge of neighbourhood facts from mounting the vertex re-identity attacks. several techniques have been evolved to mitigate neighbourhood-based totally assaults on social network statistics publishing. the following definition captures the privacy necessities beneath the $k$-neighbourhood anonymization.

## IV. SURVEY ON MEDICAL DATABASE: BACKGROUND

### 4.1 COMMAND IN REGULATORY REQIREMENTS ON PRIVACY IN DATABASE

According to Act HIPAA[14] and HITECH there should be extension of security and privacy essentials to business partners. These Acts states that every vital measure are to be activated to store the information of patients secured at whatever point it is gotten to or shared. Due to the absence of command in HIPAA security principles could prompt the loss of medical licenses.

Almost all the aspects of security are covered by HIPAA regulation. The basic requirements of security such as integrity, confidentiality and authentication are replaced by identity tracking, access control and emergency get entry to and interest auditing. It shows there is a hybrid security management in healthcare data. Which leads to means of various technical and mechanisms to incorporated to meet these targets of security and privacy.

### 4.2 BLOCKCHAIN

In decentralized computing due to emergence of Bitcoin blockchain has garnered a wide reputation. In general blockchain is classified into two types permission less and permissioned.

Permission less blockchain is also known as public blockchains which gives permissions for performing several actions on ledger such as verifying and creating transactions and adding blocks.

Ethereum is an example for permission less blockchain. The mechanism in this Ethereum is the mix of Proof of Stake and Proof of Work [15]. These two involve in adding blocks at any cost.

Permissioned blockchain is also known as an association of blockchains which act a progressively like a closed biological system: These blockchains keep up a layer of access control which permit certain kinds of nodes which performs certain actions. The algorithms such as [16] state machine reduplication, a variation of [17] are the examples of permissioned blockchains.

### 4.3 SMART CONTRACT

It is hard to broaden Bitcoin[18] to help different applications. Because the script programming language implemented inside the bitcoin is not Turing completely. Later Ethereum develops "SMART CONTRACT" by implementing a high level script programming language which leads to a complete Turing. This makes to construct a numerous decentralized programs.It is also known as tiny-size user- defined computer programs which specifies rules governing transactions. In daily life all insurance companies use these smart contracts. Chain code and Solidity are the famous platforms which uses the smart contracts.

### 4.4 OVERVIEW ON HEALTHCARE DATA MANAGEMENT

#### *4.4.1 TWISTERS OF CLOUD-BASED MEDICAL DATA MANAGEMENT*

Sharing of data in cloud-based depends on the trust that users can place on it. The security and privacy fluctuation of HIPAA, cross-institutional sharing of medical information turns out to be significantly more confused. Existing IT infrastructure collaborated with the medical organization is usually private cloud architectures. These private architectures have limitations for collaborations who reside from outside of domain area. These limitations prevents the medical organization from sharing further medical information with bigdata. Besides the public clouds gives the accessibility for sharing any data present in the database. It leads to several types of attacks. In these data encryption which provides both security and privacy for the information there is a key management problem which leads to the weaken of privacy of data. Granting users to maintain their own encryption keys leads to the increase in risk of leakage in the health care information present in the cloud.

### 4. 5 INTRODUCING CRYPTOGRAPHY IN MEDICAL DATA SHARING

In medical data sharing there is an evolution of different number of implementations for solving the problems with the cloud based data sharing .

**FAMOUS INNOVATIONS TAKEN PLACE IN THE PRIVACY PRESERVING OF MEDICAL DATA**

[19]. utilized Attribute-based Encryption(ABE) which leads to the secure data sharing of individual medical records that are put away in a cloud server.[20]. Introduced the use of ciphertext- strategy attribute-primarily based encryption to impart excellent-grained get to govern and comfy sharing of Personal Health Records(PHRs)l[21]. set ahead a cooperative architecture between authoritative distrubuting of medical information in nominal-believed distributed computing.[22]. mixed blockchain innovation with a various-authority attribute primarily based basic plan for making sure about the capacity includung passageway of electronic healthcare data.[23]. suggested a patient-driven system to give chance to patients to choice share segments of their healthcare information put away in cloud.[24]. suggested a powerful and safe patient-centric access control(ESPAC) conspire on the foundation of the ciphertext-policy attribute-based encryption to accept patient-centric access control.[25] introduced a cloud-based privacy-aware role-based access control(CPRBAC) model for data controllability including detectability, including authorized access for medicinal services cloud assets.

### 4.6 DATA ANONYMIZATION MODELS

Scholars have devised different data Anonymization calculations, for example, speculation, concealment, and decent variety slicing to secure from attackers.

Usually there are triple sorts of privateness-keeping models. They are k-anonymity, l-diversity, and t-closeness. K-anonymity is to permit each mix of quasi identifiers be vaguely matched to at minute k individuals.

L-diversity is a stronger privacy protection model. It requires every sensitive attribute comprise in any value of l well- constitute values in the distributed dataset in addition to involving k-anonymity.

t- closeness is an additional depuration of l-diversity model which stores privacy by decreasing poor quality of information portrayal, it will handle the different values of an attribute differently in the basis of captivating into account that allocation of estimations of the attribute.

Differential privacy is also a method to supply data anonymization by introducing noises to the dataset. Then the attacker gets confused whether the data portion is included or excluded. This method is proposed by [26].

### 4.7 DATA ANONYIMIZATIOIN HEALTHCARE MANAGEMENT

In general both structured and unstructured data present in the advanced healthcare database is de- identified by a system known as HIDE[27].It works to de-distinguish information while consisting maximum information usage in k-anonymity identifiable attributes.[28]. introduced optical lattice anonymization (OLA) program dependent on the properties of k- anonymization.[29]. Introduced a grouping based secrecy conspire in wireless sensor monitoring systems for sensor information assortment and conglomeration.[30]. Introduced a method to permit data proprietors to share individual healthcare records without acquiring unreasonable Loss of information, disclosing identities, or usefulness data harming.

### 4.8 SOFTWARE-DEFINED INFRASTRUCTURES FOR HEALTHCARE

As far as throughput and inertness software defined infrastructures(SDI) conditioned at the last point of of network support applications with notable performance necessities. Many of the home made medical applications are proclaimed by the software- defined infrastructures(SDI). [31].

Introduced Care Net, an adjustable fault and framework to the domestic dependent healthcare applications.[32].
Introduced a brilliant healthcare checking strategy based on software characterized networking. In this networking the processing and transmission is taken care.

### 4.9 BROADCAST ENCRYPTION

In this broadcast encryption owner can encode a bit of information into a subsets of users. Just the users present in the subset can decode the bit of information that is sent by the owner. In the cloud storage in the place of whole encryption the keys present in the encryption are encrypted in the broadcast encryption. In this methods the attackers could not attack our private data.

### 4.10 IDENTITY BASED ENCRYPTION

[33] proposed the theory of personality based encryption he advised to keep the public key as an arbitrary string. [34] utilizing Weil pairing on elliptic curves updated identity primarily based encryption. In identity primarily based encryption a trusted third party produces a private key pair for every identity string. In IBE it eliminates the use of the public key for the encryption foe each identity data there present an private key pair which acts as a main key for the data encryption. This private key is also known as the master private key which is connected in the back layer of the data base with the primary key(Identity ID). With the help of that unique private key the data can be encrypted

### 4.11 ATTRIBUTE BASED ENCRYPTION

Generally in numerous programs there's want of sharing information without any basic knowledge about the receiver. In attribute based encryption[35] data can be shared with the mentioned attributes only. For example a patient want to share his data only to the doctor who is approved by medical organization. Then while sharing data patient should mention attribute(“DOCTOR”) so that the data will be sent only to the doctor. These are classified into two types (a) key- strategy [36] attribute based encryption. (b) ciphertext- strategy attribute based encryption[37]. In this attribute based encryption no focal authority is required and there is a guarantee for collision resistance.

### 4.12 PROXY RE-ENCRYPTION

It is introduced by Blaze et al[38] and later updated by Ateniese et al[39]. The main phenomena in proxy re-encryption is It is a cryptosystem which permits third party(proxy) to change the cipher text so that it can be encrypted by the authorised users only. The third party shares the proxy key with the required user at the receiver side. So that the receiver can easily encrypt the cipher text and can access the required data. Proxy re-encryption is appropriate for information sharing over various domains.

### 4.13 SEARCH ON ENCRYPTED DATA

Accessible symmetric encryption[40] can avoid the decryption process and can prevent the secured data from any risk of leakage. There are two approaches for search on encrypted data (a)Owners of the data can either download the decrypted data locally and perform a query .(b) sending the keys for data decryption to the service providers before executing a query. Because of the security reasons these two approaches are unacceptable. For performing a search a token is supplied to the user and then this token is passed through the server and then it provides the matched encrypted data to the user.

## V. FUTURE WORK

We can likewise Ascertain the affectability levels with various list esteems. Researchers must be carried out on the devices to manage big data mining, so the examinations may be finished in real fact.. Future experiments on structuring blockchain based methodologies for safe clinical information splitting can concentrate on

The subsequent regions:

1) Cryptography-primarily based get right of entry to and privacy control to guarantee the privacy and protection need via HIPAA guidelines, cryptography should be installed inside the structure to implement severe access management and privateness protection. The. The cutting edge plans[41] in medical data areas depend pretty much in reception of some particular cryptographic natives to actualize conformation, access control, key administration, including security assurance to annoy clinical data.

2)Tingle Contract-Driven Business rationale Smart agreements, like progression to accomplish-executing legally binding states excluding outsiders, are the fundamental component for executing the business rationale of blockchain based medical information distrubuting. By planning savvy contracts precise to some particular prerequisites, the making of clinical information, authorization including repudiation to get to authorizations, including reviewing including contracts explicit to sure of clinical records, approval and provenance of get admission to permissions could be executed on the blockchain.

3)QUERY ON SCATTERED MEDICAL DATABASES

Most existing plans decide to store encoded metadata showing information areas on chain. At the point when a customer needs to play out a worldwide inquiry on totally associated databases, a further test would be in how to productively play out the inquiry on every autonomously overseen databases at the same time and also should get an accumulated inquiry result. This issue stays unaddressed in existing plans. A potential arrangement is to let a few mostly brought together servers appropriated in the system to gather and total similarly figured inquiries

what's more, return the complete outcome to the querier. In any case, solid security and recoup systems should be cautiously conveyed on these servers to shield them from denial of service (DOS) attacks.

4)FINER-GRAINED ACCESS AND PRIVACY CONTROL

At present, there left only a few techniques that have received progressed cryptographic natives (e.g., property based encoding) to implement severe and flexibleAccess control for clinical information permissions.

Specially, these concentrate on the development of access arrangements including high level semantics, these, obviously, is important in get to approach customization. Nonetheless, the separation of different EMR fields in affectability is too a basic significance for the protection control.

An innocent methodology is used to section a record into various divisions as per sensitivities and scramble every division with an alternate key,

Nonetheless, it convolutes the errand of key administration whilst the partition is one grained. To address this trouble, a few key inference instruments will also incorporated with access control arrangements to encourage key administration.

5)SIMILARITY OF SECURITY MECHANISMS AMONG HEALTH DOMAINS

Geneally every social health organization can be regarded as an free space furnished with its own security and protection component, it is hard to foresee how much these systems good with

one another. Besides, one ought to likewise think about how to address the similarity

issue brought about by various or even conflicting data privacy rules of variant regions or countries.

6) software program-described NETWORKING is wanted TO domain control

Software defined network gives an essential issue of control the disseminate arrangement data. Be that as it may, brought together control by one substance has a weakness of making an essential issue of attack.

In addition, the programmability related with the SDN stage incorporates security dangers. In this manner, appropriately and safely actualizing a SDN controller to participate including the blockchain and encourage the administration and coordinated effort among different medical areas is critical.

It should disentangle the administration of remaining heritage human services frameworks to leave them alone effectively combined with the new blockchain based designs. Current researches make certain presumption that the system is static. Practically, social network condition is powerful, which infers that a few basic features will alternate powerfully. In this manner, a methodology that considers the power of the social network condition, particularly in the space of the foe foundation information is an open issue. Also, the foundation information utilized by a promotion foe is verifiably thought to be exact.

In practice, the data accessible to a foe might be disturbed and ambiguous. Which brings up an open issue of the way viable an adversary attack might utilize uproarious in addition with ambiguous information.

Further examination is expected to deliver the provoking issue to dissect the attack models and the anonymization models for huge scope pragmatic datasets. The present systems has been corrected on little, recreated systems whose qualities are distinguished from true social networks.

Last but not least, the best approach to detail the thought of differential security with regards to interpersonal organization is as yet an open issue.

The hub differential security gives an exceptionally solid assurance. Anyway, it may not give a response to every apprehensions of protection break in an social network.

This difficulty should be routed to propel differential protection to deal with this concern. since various kinds of examinations are sensitive to various degree of blunders, a different line of studies at the differential protection is the way to decrease the measure of commotion introduced to question and consequences to consent to necessities of the differential privateness.

It is supported via Kifer and Machanavajjhala, the effectiveness of differential privateness for exceptionally associated facts is constrained, recommending a more grounded privateness safety version for datasets with profoundly connected information is required.

Majority of existing research on differential security focuses in the intuitive fixing meanwhile the non-intelligent fixing had received tiny consideration.

## VI. COMPARITIVE ANALYSIS

### SOCIAL NETWORK DATA MODEL

The, center to the privacy issue is issues of: I) relationship space, to be specific limit which is two- crease: deliberation and granularity; and ii) connection space, to be specific course, multiplex relationship and association degree. Then again, assent and the ability to erase data are significant issues depend on the specialist co-op's aims.

Scientific models to help programmed search and inquiry noting usage need to suit all the distinguished factors as follows:

Simplex relationship vs multiplex relationship:

In the offline world, human connections are undeniably more complicated than just friends.

Different relationship types exist in the human culture. The online social condition ought to reenact the disconnected world to help multiplex connections. Giving such a relationship space to social collaboration can encourage the control of access level and data stream, prompting better conservation of data privacy.

Individual level vs network/group level:

Interaction on some deliberation levels, e.g., individual-to-group, group to-group, group to- network, and so on., can keep away from certain detail of data being revealed. Along these lines, pleasing connections at different deliberation level for cooperation is required.

This issue can possibly prompt the issue of relationship structure, i.e., the capacity to oversee various leveled connections and use them for security conservation.

Symmetric vs asymmetric:

Relationships are frequently asymmetric. Better decision-making on privacy conservation requires a comprehension of properties of the relationships involved.

**VI.2 COMPARISION OF PRIVACY MODELS IN BIG DATA**

| MODEL | MERITS | DEMERITS |
|---|---|---|
| Randomization | It is a simple method which can be easily implemented. | In the time of categorical attributes and multiple attributes it faces difficulty. |
| K-anonymity | Implementation is very easy. At the point when the estimation of k-anonymity is high there will be an opportunity of re-distinguishing proof. | It slumps in hindering the background information include homogeneity attacks, affects from attribute linkage and file linkage, lengthy handling time, application can be subverted by some question which gives least of k fits. |
| I- diversity | Minimize the information rundown structure. sensitive attribute which have nearly a similar recurrence. | Relies on the scope of sensitive attributes.l-diversity, there ought to l various estimations of sensitive attributes. I-diversity is slanted to imbalance furthermore, comparability attack and may not forestall disclosure attribute. helplessness in opposition to homogeneity append and back ground information attack. |
| t- closeness | It prevents the information from skewness attack. | complicated computational strategy to implement t- closeness. tcloseness loses the co connection among various attributes considering the fact that each attribute is summed up independently. usage is harmed on the factor whilst t could be very little. |
| Differential Privacy | It is the most appropriate model for Big information. It gives the most grounded privacy assurance. | Data usage might be diminished. Data miner is as it were permitted to present total queries. Likelihood of attacking each the databases by means of foe isn't always taken into . |
| Slicing | Prevents attribute disclosure. Unsystematic on sensitive attributes. | Usage and hazard measure aren't coordinated. Slicing might break relationship between attributes. |

## VII. CONCLUSION

In this paper the procedure was done with regards to privacy Preserving pesrevation methods in Big Data,Medical Data,Social Network Data.In the aspect of Big Data, privacy is achieved through (v,l)-anonymity by using the flavours of clustering and bottom-up generalization. clustering and bottom-up generalization. The introdused privacy model deals with sensitivity vulnerabilities and overcome the disadvantages of existing privacy models. the entire process was wiped out the context of massive Data, which is that the results of increase within the communication means and knowledge sharing. Traditional sys- tems might not efficiently deals with this huge amount of knowledge and should results in the sys- tems run very slowly. But hive is an environment where we will handle Big Data very efficiently with none changes to the prevailing procedures. during this paper we only consid- ered one sort of index value, which is deathrate . we will also calculate the sensitiv- ity levels with different index values. The more researches need to be done on the tools to handle Big data processing , in order that the experiments are often administered within the world context. Medical data distribution without damaging security and privacy rules has for some time been a difficult point.

In this paper we collected the audits related arrangements right now, cloud based methodologies, blockchain dependent methodologies, including software defined network dependent methodologies. We saw that security including assurance of clinical data secrecy, respectability,

including validness of information in movement including very still, access including protection control and so on. So , a pragmatic methodology to clinical information distrubuting might need for coordinating a wide range of systems to gain structure targets.

Updated computing worldview, blockchain also contains points of interest on customary advancements.Besides,few issues calling for additional examination and investigation in blockchain dependent clinical information administration. We mentioned about difficulties by bringing up potential studies bearings including strategies which might additionally make sure about and encourage the sharing of healthcare information.

Social networks administrators are progressively distributing and sharing social network information for proxy consumers.Distributed social network information includes sensitive information of clients inclu. This has provoked security concerns and dynamic research in protection safeguarding components. Right now, proposed a significant level structure for social network publishing risk examination. We likewise introduced the risk model and measured and arranged the foundation information that is possibly utilized by attackers to rupture privacy of the distributed interpersonal organization information. We likewise introduced various techniques, approaches, procedures and methods in privacy preserving publishing of social networks. In conclusion, privacy preserving publishing stays a difficult issue

## VIII. AKNOWLEDGEMENT

We would like to thank the Vellore Institute of Technology, (VIT)for giving us this opportunity and support throughout the course of this work. We would like to extend our sincere gratitude to our professor Prof. Vijaya Kumar K and all other faculties for the knowledge and information that was taught and shared in class

## REFERENCES


1. M. Fire, R. Goldschmidt, Y. Elovici, Online Social Networks: Threats and Solutions, IEEE Communications Surveys & Tutorials, Volume: 16, Issue: 4, 2019 - 2036, 2014.
2. N. Vastardis and K. Yang, Mobile Social Networks: Architectures, Social Properties, and Key Research Challenges, IEEE Communications Surveys & Tutorials, Volume: 15, Issue: 3, Pages: 1355 - 1371, 2014.
3. A. Narayanan and V. Shmatikov, "De-anonymizing social networks," In Proceedings of IEEE Symposiums on Security and Privacy, pp. 173–187, 2009.
4. K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," In Proceedings of the 2008 ACM Proceedings of the 2010 ACM International Conference on Management of Data (SIGMOD 2008), pp. 93–106, 2008.
5. N. Li and S.K. Das, "Applications of k-Anonymity and ℓ-Diversity in Publishing Online Social Networks, in Y. Altshuler et al. (eds.): Security and Privacy in Social Networks, Springer, pp. 153–179, 2013.
6. J. Cheng, A. W. Fu, and J. Liu, "K-isomorphism: privacy preserving network publication against structural attacks," in Proceedings of the 2010 ACM International Conference on Management of Data (SIGMOD 2010), pp. 459–470.
7. B. Zhou and J. Pei, "Preserving privacy in social networks against neighborhood attacks," in Proceedings of IEEE 24th International Conference on Data Engineering, 2008 (ICDE2008) IEEE Press, pp. 506–515, 2008.
8. B. Zhou and J. Pei, "The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighbourhood attacks," Knowledge and Information Systems, Volume 28, pp. 47-77, 2011.
9. W. Peng, F. Li, X. Zou, and J. Wu, "A Two-Stage De- anonymization Attack against Anonymized Social Networks", IEEE Transactions on Computers, Volume 63, No 2, pp. 290–303, 2014.
10. A. Narayanan and V. Shmatikov, "De-anonymizing social networks," In Proceedings of IEEE Symposiums on Security and Privacy, pp. 173–187, 2009.
11. K. Bringmann, T. Friedrich, and A. Krohmer, "De- anonymization of Heterogeneous Random Graphs in
12. Quasilinear Time" in A. Schulz and D. Wagner (Eds.): ESA2014, Lecture Notes in Computer Science 8737, pp. 197–208, 2014.
13. P. Boldi, F. Bonchi, A. Gionis, and T. Tassa, "Injecting uncertainty in graphs for identity obfuscation," In Proceedings of VLDB Endowment, Volume 5, No 11, 1376–1387, 2012.
14. Y. Li, Y. Li, Q. Yan, R. H. Deng, Privacy leakage analysis in online social networks, Computers & Security, Volume 49, Pages 239-254, 2015.
15. (2017). Summary of the HIPAA Security Rule. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/
16. A. Gervais, G. O. Karame, K. Wüst, V. Glykantzis, H.
17. Ritzdorf, and S. Capkun, ''On the security and
18. performance of proof of work blockchains,'' in Proc. ACM SIGSAC Conf. Comput. Commun. Secur. New York, NY, USA: ACM, 2016, pp. 3–16.
19. J. Sousa, E. Alchieri, and A. Bessani, ''State machine replication for the masses with BFT-SMaRt,'' in Proc. 44th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw., Atlanta, GA, USA, 2014, pp. 355–362.
20. M. Castro and B. Liskov, ''Practical Byzantine fault tolerance,'' in Proc. OSDI, vol. 99. 1999, pp. 173–186.
21. S. Nakamoto. (2009). Bitcoin: A Peer-to-Peer Electronic Cash system. [Online]. Available: http://bitcoin.org/bitcoin.pdf
22. M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, ''Scalable and secure sharing of personal health records in cloud computing using attributebased encryption,'' IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 1, pp. 131–143, Jan. 2013.
23. J. Liu, X. Huang, and J. K. Liu, ''Secure sharing of personal health records in cloud computing: Ciphertext- policy attribute-based signcryption,'' Future Generat. Comput. Syst., vol. 52, pp. 67–76, Nov. 2015.
24. B. Fabian, T. Ermakova, and P. Junghanns,''Collaborative and secure sharing of health caredatainmulti-clouds,''Inf.Syst.,vol.48,pp. 132–150, Mar.2015.
25. R. Guo, H. Shi, Q. Zhao, and D. Zheng, ''Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems,'' IEEE Access, vol. 6, pp. 11676–11686, 2018.
26. S. Narayan and M. Gagné, and R. Safavi-Naini,
27. ''Privacy preserving EHR system using attribute-based infrastructure,''inProc.ACMWorkshop Cloud Comput. Secur. Workshop. New York, NY, USA: ACM, 2010, pp. 47–52.
28. M. Barua, X. Liang, R. Lu, and X. Shen, ''ESPAC: Enabling security and patient-centric access control for ehealth in cloudcomputing,''Int.J. Secur. Netw., vol. 6, nos. 2–3, pp. 67–76, 2011.
29. L. Chen and D. B. Hoang, ''Novel data protection model in healthcare cloud,'' in Proc. IEEE Int. Conf. High Perform. Comput. Commun., Sep. 2011, pp. 550–555.
30. J. Soria-Comas, J. Domingo-Ferrer, and D. Sánchez, and S. Martínez, ''Enhancing data utility in differential privacy via microaggregation-based k-anonymity,'' VLDB J. Int. J. Very Large Data Bases, vol. 23, no. 5, pp. 771–794, 2014.
31. J. Gardner and L. Xiong, ''Hide: An integrated system for health information DE-identification,'' in Proc. 21st IEEE Int. Symp. Comput.-Based Med. Syst., Jun. 2008, pp. 254–259.
32. K. El Emam et al., ''A globally optimal k-anonymity method for the deidentification of health data,'' J. Amer. Med. Inf. Assoc., vol. 16, no. 5, pp. 670–682, 2009.
33. P. Belsis and G. Pantziou, ''A k-anonymity privacy- preserving approach in wireless medical monitoring
34. environments,''Pers.UbiquitousComput., vol. 18, no. 1,pp. 61–74, 2014.
35. G. Loukides, J. Liagouris, A. Gkoulalas-Divanis, andM. Terrovitis, ''Disassociation for electronic health record privacy,'' J. Biomed. Informat., vol. 50, pp. 46–61, Aug. 2014.

36. P. Li, C. Xu, Y. Luo, Y. Cao, J. Mathew, and Y. Ma, ‘‘CareNet: Building regulation-compliant home-based healthcare services with software-defined
37. infrastructure,’’ in Proc. IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Eng. Technol. (CHASE), Jul. 2017, pp. 373–382.
38. L. Hu et al., ‘‘Software defined healthcarenetworks,’’ IEEE Wireless Commun. Mag., vol. 22, no. 6,pp. 67–75, Jun. 2015.
39. A. Shamir, ‘‘Identity-based cryptosystems and signature schemes,’’ in Proc.WorkshopTheoryAppl. Cryptograph.Techn.Paris, France:Springer, 1984, pp. 47– 53.
40. D. Boneh and M. Franklin, ‘‘Identity-based
41. encryption from the weil pairing,’’ in Proc. Annu. Int. Cryptol. Conf. Santa Barbara, CA, USA: Springer, 2001, pp. 213–229.
42. A. SahaiandB.Waters, ‘‘Fuzzyidentity-
43. basedencryption,’’inProc.Annu. Int. Conf. Theory Appl. Cryptograph. Techn. Aarhus, Denmark: Springer, 2005, pp. 457–473.
44. V. Goyal, O. Pandey, A. Sahai, and B. Waters,
45. ‘‘Attribute-based encryption for fine-grained access control of encrypted data,’’ in Proc. 13th ACM Conf. Comput. Commun. Secur. New York, NY, USA: ACM, 2006, pp. 89–98.
46. M. Chase, ‘‘Multi-authority attribute-basedencryption,’’ in Proc. Theory Cryptogr. Conf. Amsterdam,The Netherlands: Springer, 2007, pp. 515–534.
47. M. Blaze, G. Bleumer, and M. Strauss, ‘‘Divertible protocols and atomic proxycryptography,’’inProc.Int. Conf.TheoryAppl.Cryptograph.Techn. Espoo, Finland: Springer, 1998, pp. 127–144.
48. G. Ateniese, K. Fu, M. Green, and S. Hohenberger,‘‘Improved proxy re-encryption schemes withapplications to secure distributed storage,’’ ACM Trans.Inf. Syst. Secur., vol. 9, no. 1, pp. 1–30, 2006.
49. G. S. Poh, J.-J. Chin, W.-C. Yau, K.-K. R. Choo, and M.S. Mohamad, ‘‘Searchable Symmetric Encryption: Designsand Challenges,’’ACMComput. Surv., vol. 50, no. 3, pp. 40:1–40:37, 2017.
50. R. Guo, H. Shi, Q. Zhao, and D. Zheng, ‘‘Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems,’’ IEEE Access, vol. 6, pp. 11676–11686, 2018

## AUTHORS PROFILE

**Guna Sekhar Tirumalasetti,** Curently pursuing Btech in Vellore Institute of technology,VELLORE,Fields instrested in Security and Data privacy

**Bhargava Sai Sathvik Gontla** Currently pursuing Btech in Vellore Institute of technology,VELLORE,Fields instrested in Security and Data privacy

**Vijayakumar Kuppusamy** received his M.Tech. and Ph.D. Degree in 2006 and 2018 from VIT, Vellore, India. Research interests are Data Mining and Missing data. Working as Associate Professor, in Department of Information Security, at School of Computer Science and Engineering (SCOPE), in Vellore Institute of Technology, Vellore632014.