AMSE Association for the Advancement of Modelling & Simulation Techniques in Enterprises
IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Advances in Modelling and Analysis B
Vol. 61, No. 1, March, 2018, pp. 9-14
Journal homepage: http://iieta.org/Journals/AMA/AMA_B

Maintening multi-level confidentiality on big data using Pk-anonymization methods and cryptographic techniques

Ahmed Mohammed[1*], G. Rama Mohan Babu[2]

[1] University College of Engg. & Technology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur 522510, India
[2] Department of IT, RVR&JC College of Engineering, Guntur 522510, India

Corresponding Author Email: ahmedsdbasha@gmail.com

https://doi.org/10.18280/ama_b.610103

Received: 9 February 2018
Accepted: 26 March 2018

Keywords:
k-anonymization, pk-anonymization, multilevel-trust(MLT), MLTPPDM, perturbation, data mining, cryptography, steganography, diversity attack, secure transformation

ABSTRACT

Anonymization innovation is basic for accomplishing assurance on security when utilizing individual data. In the time of bigdata a lot of data has been aggregated by different data repositories. Manu problems are arised in personal informations are recognized by coordinating through other information. Anonymization process in bigdata being a test to change over individual information into non individual information. With the assistance of the guide reduction structure the huge number of organizations and associations use anonymization methods to process massive volume informational collections. Security safeguarding and elevated function of Data is conceivable because of the guide diminishing structure and k-Anonymization expertise which can be effectively used with big data.In this manuscript, probabilistic k-anonymization process is used for the information conversion. Here we give multilevel security to our information to make framework more secure by using Pk-anonymization.by using this technique the personal data of a individual is converted into a un-identified format which is highly secured. One of the primary destinations of these framework is to prevent from straight variety assault and other non-straight assaults. To give multilevel security we consolidate cryptography and steganography approaches also along with Pk-Anonymization method. The advantage of these plan is that steganography can work on encoded content and thus it offers a twofold layer information assurance and heartiness for secure information transmission over an open channel. The cryptographic mechanisms are applied on the data which is modified using Pk-Anonymization technique and secure information transferring can be achieved over the cloud.

1. INTRODUCTION

With the vast volume of information, manual examination is not any more practical. Rather, programmed or self-loader instruments that utilize information mining procedures have been generally used to help information investigation [2]. More or less, information mining is the way toward finding designs in information, and the nature of information [11] is vital for the accomplishment of an information mining process[1, 25]. Be that as it may, as the general population turns out to be more worried about protection, an ever increasing number of individuals are unwilling to give genuine individual information when made a request to do as such. In addition, organizations that desire to utilize their clients' information for information mining can't undoubtedly do as such without bargaining their clients' protection [6, 22].

This is a predicament between information quality and information security. The k-anonymization show [9-10] is a way to deal with shield information from singular distinguishing proof. It results in guaranteeing that each and every testimony of table is indistinguishable to any rate k_m1 different records as for an arrangement of protection related traits[12, 18], called semi identifiers, that could be possibly used to distinguish people by connecting these credits to outside informational collections. For instance, consider the clinic information in Table 1, Corresponding creator. where the traits ZipCode, Gender and Age are viewed as semi identifiers.

Table 1. Clinical information

ZipCode	Gender	Age	Disease	Expense
75275	Male	22	Flu	100
75277	Male	23	Cancer	3000
75278	Male	24	HIV+	5000
75275	Male	33	Diabetes	2500
75275	Female	38	Diabetes	2800
75275	Female	36	Diabetes	2600

The proposed work is focus on multilevel trust. Presently a large portion of people groups utilizes public or private clouds and individuals utilize them to store their vital information like Mastercard number, ledger number,properties details, email locations and passwords or some business related information

or bank details and so forth. As some mobiles give office to store their information on google cloud on the web or other service providers many individuals are interested in using them. Under this administration, noxious information excavator may approach that information while information transmission occurred. Different suspicious assaults like assorted variety and nonlinear assault may focused on this information. The answer for this current approach are restricted in their inferred presumption of single-level trust on information diggers [24]. The answer for these issue is to served multilevel security while exchanging information.

In proposed framework, we are going to served to multilevel security to information. To guarantee multilevel security we join Pk-Anonymization strategies and cryptography, steganography approach [20]. At the point when information is exchanged from sender to beneficiary in cloud, In first level,Pk-Anonymization techniques are applied to convert the sensitive information to unreadable format and then cryptography is utilized on the modified data [15, 24]. In the wake of applying cryptography the discharge information [27] can be changed over in encoded design, that is ambiguous to outsider. In second level, picture steganography is utilized. This enables the clients to safely conveying the information. The main undertaking of the Steganography [19] is to nourish client flexibility of transient the data, actualizing the process of encryption benchmarks according to the particular and computations proposed and stores the data which is in imperceptible frame into the cloud safely and securely. Utilizing steganography the encrusted information is holed up behind a picture [8]. The key based recovery office is use to recover the concealed information at recipient side. In proposed framework, multilayer security [23] is given to information and in this way, nobody separated from sender and expected recipient distinguish the presence of message. The advantage of these approach is that steganography can connected on encoded writings and thus it bolsters a double layer information insurance [4, 26]. This framework improves security of client's touchy data or information or some essential documents by consolidated science of steganography [3, 13] and cryptography to fulfill necessity of security and vigor for secure information transmission over a station.

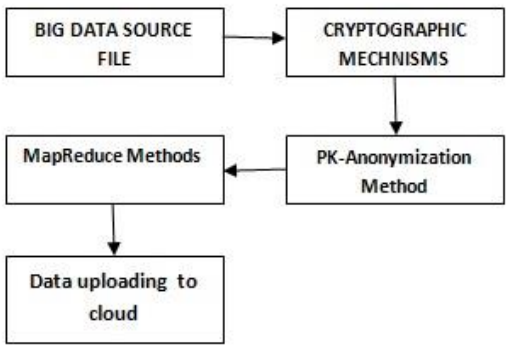

Figure 1. System setup

The above figure illustrates system arrangement for assessment of big data with anonymization techniques [21, 24].The framework covers a different put stock in levels for information to give security. , In this process, first information is send from sender side to planned client. Pk-anonymity performs the override operation for replacing the personal data and then utilizing Cryptography this information gets changed over into encoded information [5, 7]. On this scrambled information picture steganography is connected. Utilizing this the information is covered up by picture and send to cloud. This framework has an inversion procedure, which is utilized to deembed the information from picture document and then unscramble the information to its unique organization as per the demand by the client [7, 10]. To retrive the concealed subtle elements of information the key based recovery technique is utilized. While performing Encryption and Decryption, the application ought to fulfill the benchmarks of approval and confirmation of the client.

2. RELATED WORK

Security saving is a noteworthy issue as of late. There are number of existing framework for security of private information. Already there are such a variety of methodologies are characterized for protection conservation of information. According to the examination done by Seema Kedar and few others, they have composed idea of existing security conservation of information mining methods and how to achive their efficeiency [1, 17]. This overview on Pk-anonymity can be useful for finding the restrictions of existing information mining approaches. It guarantees effective security protection of information. The single-level trust Pk-anonymity problem via information annoyance has been generally examined in the writing. In this, a proprietor of the information have certain trusts on all beneficiaries of its information consistently and it circulates just a single bothered duplicate of the information. A broadly utilized and acknowledged approach to irritate information is by added substance bother.

In the investigation by Vaishali Borade, R.N.Phursule, they had inspected in the event that we apply nonlinear conspiracy assault on Pk-Anonymization for isolation maintainance approach, it is conceivable to recreate unique data [6, 14, 21]. When same nonlinear intrigue assault is connected on proposed framework it can't reproduce unique duplicate information implies it protect the security. In that work, they may apply nonlinear procedures to infer unique information and recoup more data. Under the multilevel trust situation, at more elevated amounts information diggers can get to less irritated duplicates. In any case, these less annoyed duplicates are not available by information mineworkers at bring down confide in levels. At various put stock in levels, information excavators may interest to share the bothered duplicates among themselves. Thus, it is regular that information diggers approach more than one bothered duplicate [14]. In this, security is safeguarded, straight change calculation which produces common data into unique information. Also, when same nonlinear assault is connected then unique information can't be reproduce.

Security safeguarding intended to avoid data presentation and guarantee legitimate access to the information. Consequently, protection safeguarding is not quite the same as existing information security, get to control and encryption innovation which tries to anticipate data exposure against ill-conceived implies. The significance of Personal information covering up has been brought up in different applications. One of the mainstream approaches for data stowing away is steganography. This approach utilized by R. Valarmathi and G. M. Kadhar Nawaz builds framework security by utilizing

steganography with encryption alongwith key management [16]. Here One more approach of giving multilayer security to client's close to home information presented by mrs. Rabbit Ram Sah and G. Gunasekaran. In burrow ital security the undeniable advantage of steganography over encoded information is that, this messages don't draw in consideration of information excavators to themselves, to delivery people or to recipient [3].

Adaptability [17, 28] is important for current security safeguarding approaches, in light of the fact that the size of informational indexes is too extensive to be in any way prepared by existing brought together calculations. This decrease is an exchange off that outcomes in some loss of viability of information administration or mining calculations keeping in mind the end goal to increase some security.

3. MTHODOLOGY OF PK-ANONYMIZATION

The data sets are accumulated in private by data vendors of cloud and cannot be utilized by other users. Data vendors indicate confidential necessities and present them towards the confidential maintainance structure [28]. The below Fig.1 stipulate the on the whole presentation occurred in PK-Anonymization.

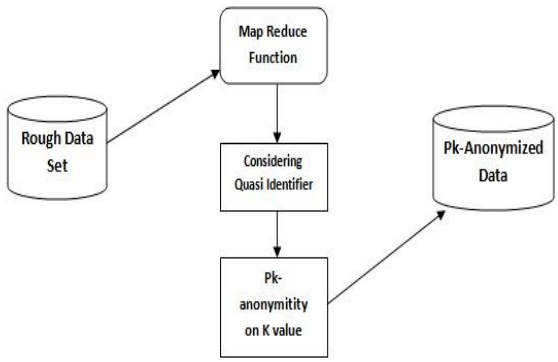

Figure 2. Structural design

Figure 2 illustrates architecture of the entire process of anonymization in which K value is considered for performing Pk-Anonymization method on the data set considered. In the first place the Bigdata experience a procedure called outline, it gives the information won't permit the copy information it contains just the predictable information without duplication. Furthermore, just the predictable information will go for the procedure of PKanonymization. Guide decreasing [8] comprises of three procedure to play out the reliable information. It incorporates mapping, rearranging and decreasing the information. The resultant arrangement of the guide diminishing information can experience the procedure of the PK - anonymization. To start with to distinguish the semi identifier. Semi identifiers [12] are snippets of data that are not of themselves extraordinary identifiers, but rather adequately all around related with an element that they can be joined with other semi identifiers to make a special identifier.

3.1 Algorithm Pk –Anonymization

Randomization: k-anonymization is accomplished just with confidential uprooting of information. Pk-Anonymization is a randomization technique with wellbeing proportionate to k-anonymization.[1]. Machine learning gives adjust information by evaluating information appropriate for investigation utilizing the attributes of randomization. Anonymity in enormous information [16, 27] is asked for under the present translation of the legitimate framework, behind which are an assortment of requirements with the expectation of complimentary utilization of individual information by making them unknown. It is conceivable to process individual information into an express that has altogether wiped out singularity incorporated into information. k-Anonymity ensures that, with any mix of estimations of semi recognized characteristics in the distributed smaller scale data set $T_0(A_1,...,A_n)$, in which k records allocating that blend of qualities. Hence, given a person in an outer non-unknown informational index, the likelihood of playing out the correct linkage back to the comparing record in the distributed small scale informational index, and in this way the likelihood of taking in its private traits, is 1/k. In this way probabilistic k-secrecy is characterized.

Algorithm Pk-anonymization

Require: The k-anonymised dataset $\mathcal{D}'_T$, the original dataset $\mathcal{D}_T$, the set of trajectories for the attacks BK_T and anonymity threshold k.

for $h = 1, \ldots, M$ where M is the length of the longest trajectory in $\mathcal{D}_T$ **do**

  for t' of length h in BK_T **do**

    $N(t')_{\mathcal{D}_T} \leftarrow |\{t \in \mathcal{D}_T | t' \preceq t\}|$.

    $N(t')_{\mathcal{D}'_T} \leftarrow |\{t \in \mathcal{D}'_T | t' \preceq t\}|$.

    if $N(t')_{\mathcal{D}_T} \geq k$ *and* $N(t')_{\mathcal{D}'_T} \leq N(t')_{\mathcal{D}_T}$ **then**

      $\Pr(re\_id|t') \leftarrow 1/N(t')_{\mathcal{D}_T}$.

    else

      $\Pr(re\_id|t') \leftarrow 1/N(t')_{\mathcal{D}'_T}$.

    end if

  end for

end for

return Cumulative Distribution of $\Pr(re\_id|t')$ for all h.

4. PK-ANONYMIZATION FLOW CHART

Here, first the enormous data from the cloud is considered by a user and then the information will be identified by semi identifier to give important information of a desired user. After that the guide diminishing will be passed to maintain a strategic distance from the copy information System and then the sensitive information is convereted into unreadable format and then cryptographic strategies are applied on them for providing high security for data stored in the cloud.

Figure 3 delineate the segments of framework.

- User Login: Here, client right off the bat logon to the framework.
- User database: It contains client's close to home information or some other data.
- Cryptography: It is use to secret unique message from client to unintelligible shape that is encoded arrange.
- Steganography: Here, the utilization of steganography is to install a message which is in encoded design into picture document to conceal the first information. What's more, this resultant picture document is exchange to server.
- Server database: It stores all the client related information.

• Unauthorized client: He is noxious client or information excavator who tries to get to client's close to home data or, on the other hand different points of interest.Figure 4 describes the anonymization process by applying cryptographic techniques along with anonymization methods and finnaly stored the data in cloud.

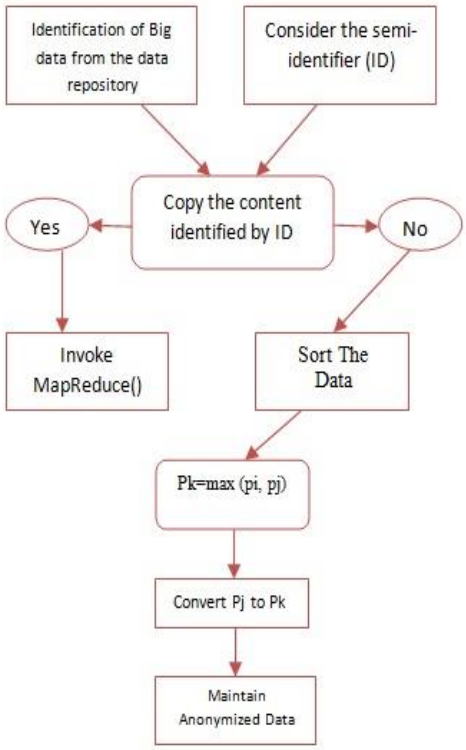

Figure 3. Framework architecture

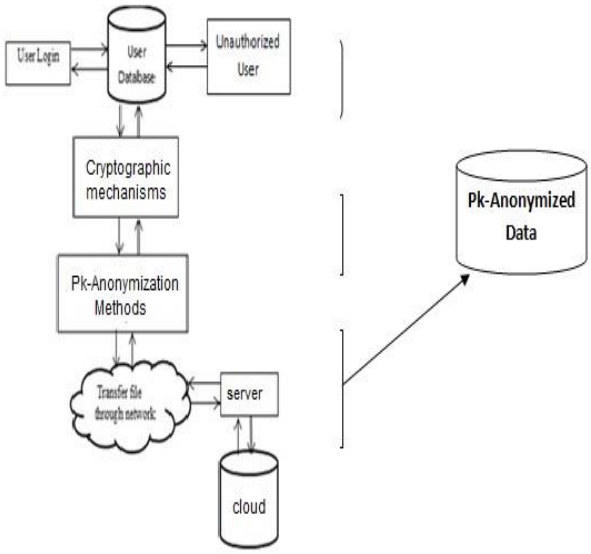

Figure 4. Pk-anonymization process

5. CRYPTOGRAPHIC MECHANISMS FOR STORAGE OF DATA IN CLOUD

Cloud security concerns have risen to be of an expanding interest and significance, in the connected cryptography and PC inquire about group, while requesting sufficient measures for cloud challenges. For this reason, we think about a capacity situation, where the customer outsources information in remote servers. The cloud servers go about as conveyed secret elements, from the users view. The systems exhibited in this section are in part to be utilized at the specialist co-op's side or the customer's side, yet regardless ought to secure the interests of both, to set up an effective also, dependable administration.

Cloud Service Provider (CSP) is untrusted outsider which gives information storerooms, computational offices. Subsequent to performing Pk-Anonymization procedure the information is changed over into a secured arrangement and now to store the information safely in cloud we present new substance call as „Cryptserver", which is trusted gathering and assume liability of encode/unscramble the document, mystery key administration and send scrambled/decoded record to entity(users,CSP) and evacuates weight of encode/decode records by users,key administration and trade key with clients by proprietor, proprietor not be constantly online when the client needs get to the information and not uncover any unique certifications to CSP.

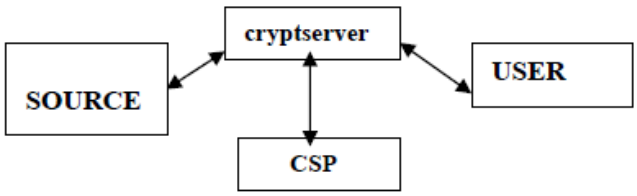

Figure 5. Cloud secure data storage

Predominantly 3 substances in our proposed demonstrate:

i. Users: clients of the framework clients are separated into two sorts.

• Source: need to share possess information to different parsons and furthermore need to allocate get to rights to parsons, get to control list (ACL) is doled out by proprietor to CSP in view of CSP control access on shared information.

• User: get to the mutual document by proprietor in light of access rights allocated by proprietor.

ii. Grave server: Trusted Party take all obligations scramble, decode of shared records, age and administration of the encryption key K.

iii. CSP: Untrusted party give store offices and to sharing information, keep up ACL allocated by client and in view of that control access of encoded store record.

Algorithm

```
Selection of data (type)
If (type==1)
{
Structured data
Perform Pk-Anonymization
Encrypt()
Send the data to cloud
}
Else
{
Unstructured data
Perform k-anonymization
Perform Pk-Anonymization
Encrypt()
Send the data to cloud
}
```

The data is considered as structured or unstructured data.

The above algorithm explains the process of performing anonymization and storing it in the cloud.

- **Key Generation**

  Select two large primes p and q, such that $p \neq q$.
  $n = p * q$.
  $\varphi(n) = (p-1)*(q-1)$, where φ is Euler's totient function.
  Select an integer e such that $1 < e < \varphi(n)$ and $\gcd(e, \varphi(n))=1$ (coprime).
  $d = e^{-1} \bmod \varphi(n)$
  Public key $\leftarrow (e,n)$
  Private key $\leftarrow d$

- **Encryption**

  $c = m^{e} \bmod n$

- **Decryption**

  $m = c^{d} \bmod n$

The proposed algorithm above is used to encrypt the anonymized data which uses Pk-anonymization. After the data get encrypted the data will be successfully stored in the cloud.

6. SIMULATION RESULTS

Anonymization methods result in contortions of the information. Over the top anonymization might decrease the nature of information considering it inadmissible for various examinations, and potentially result in off base or one-sided comes about. Along these lines, it is vital to adjust the measure of Pk-anonymization being performed against the measure of data misfortune. It is in this way imperative to see accurately the sorts of assaults that can be propelled on an informational collection and the distinctive approaches to legitimately anonymize the information before it is revealed. Here, concentrated on Pk-namelessness, which is a mainstream approach for securing protection utilizing k-implies incremental grouping to deal with incremental cloud information.

6.1 Data anonymization

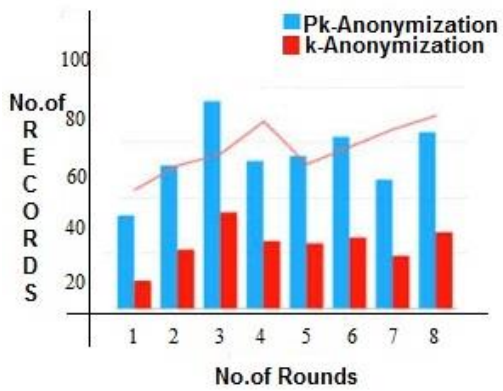

Figure 6. Experimental results on performance of K and Pk-anonymization

By contrasting the methodologies of Pk-Anonymization and cryptography. Pk-Anonymization gives less data misfortune than the k anonymization.Consequently, the security safeguarding structure can essentially enhance the capacity and effectiveness contrasted and existing cutting edge anonymization approach. The outline obviously exhibits the correlation among k and the Pk-anonymization. Lastly the thrashing of data have to be underneath half to give improved anonymization. To assess the fundamental segments of the protection safeguarding system through directing the tests on certifiable informational indexes. Utilizing the general residents healing facility data set to the procedure of anonymization. Semi attribute resolve to be distinguished accurately for the improved anonymization. Contrasting k and Pk-anonymization, the PK-anonymiation has enhanced effectiveness of the information with small loss of data. It is explained on the beneath Fig. 6. Here comparision is based on no.of rounds anonymization is performed Vs No.of Records.Here in the process of Pk-anonymization the data which is to be stored in the cloud is to be anonymized and converted to unreadable format as the hacker cannot misuse the data.The experimental results show that Pk-anonymization performs well.

6.2 Execution instance

The time amid where a set of codes are running, rather than different periods of a program's lifecycle, for example, accumulates time, interface instance and load instance is called execution instance. On the off chance that the record in the big data is little then the run time of K-anonymization is superior to the Pk-Anonymization. It is ascertained by Consumed Time = AllRecords() – ExecutedRecords().

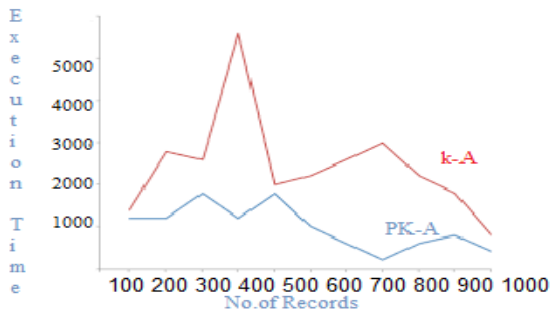

Figure 6. Execution time analysis

Here when compared with different types of anonymization strategies PK-Anonymization is performing well in all aspects of converting the data and securely storing them in cloud.

7. CONCLUSION

Anonymization systems outcome in twisting the sensitive information. Unnecessary Anonymization procedures result in contortions the information. Over the top anonymization may lessen the nature of the information making it inadmissible for some investigation, and conceivably result in inaccurate or one-sided comes about. In this manner, it is essential to adjust the measure of anonymization being performed against the measure of data problem. It is along these lines essential to see unequivocally the sorts of re-distinguishing proof assaults that can be propelled on an informational index and the diverse approaches to appropriately anonymize the information before it is uncovered. A adaptable, versatile, dynamical and practical protection safeguarding structure in light of performing Map Reduce on cloud with Pk-Anonymity is proposed in this manuscript . The security protecting system can anonymize substantial scale informational indexes and deal with the mysterious informational collections in a very adaptable, versatile, proficient and savvy mold.

This venture gives adaptable protection system on customary Bigdata and not for spilling of information. With the assistance of the tempest the spilling of information can be refreshed successfully. A few information handling system will be coordinated to play out the anonymization all the more successfully. Pk-Anonymization can be utilized to convert distinctive big data set in a compelling way. Information Security is a vital issue now-a-days. The current innovations for confidential maintainance of information are great upto some broaden. In any case, to keep from assorted variety assault and other nonlinear assaults and furthermore for greater improvement, we proposed the above framework that has been configuration to give a multilevel security to client's information while exchanging of data in clouds. The Pk-Anonymization techniques and cryptography,steganography are utilized for multilevel security. The motivation behind this plan is that the steganography can connected on encoded information, on account of this the framework offers double layer information security. The joined approach of both these procedures give hearty and secure information transmission and maintaianance in clouds.

REFERENCES

[1] Altman E, Jiménez T. (2003). NS Simulator for beginners [Online]. Available: citeseer.ist.psu.edu/altman03ns.html.

[2] Azimi R, Bhatia G, Rajkumar R, Mudalige P. (2011). Vehicular networks for collision avoidance at intersections. Society for Automotive Engineers (SAE) World Congress, Detroit, MI, USA

[3] Bishop R. (2000). A survey of intelligent vehicle applications worldwide. Proceedings of the IEEE Intelligent Vehicles Symposium 2000, Dearborn, MI, USA, pp. 25-30.

[4] Boneh D, Boyen X, Shacham H. (2004). Short group signatures. In: Franklin, M.K. (ed.) CRYPTO 2004. Springer, Heidelberg, pp. 227-242.

[5] Karp B, Kung HT. (2000). GPSR: greedy perimeter stateless routing for wireless networks. Proceedings of the 6th annual international conference on Mobile computing and networking, MobiCom.

[6] Bresson E, Stern J, Szydlo M. (2002). Threshold ring signatures and applications to ad-hoc groups. In Proc. CRYPTO 2002, USA, Lecture Notes in Computer Science, 2442, Springer-Verlag, pp. 465 480.

[7] Cao Z, Hu J, Chen Z, Xu M, Zhou X. (2008). FBSR: Feedback based secure routing protocol for wireless sensor networks. J. Pervasive Comput. & Comm.

[8] Chang S, Chen L, Chung Y, Chen S. (2004). IEEE Transactions on Intelligent Transportation Systems.

[9] Chaum D., Hevst EV. (1991). Group Signature. In Eurocrypt 1991, volume 547 of LNCS, pp. 257-265.

[10] Baltimore TSG, Chen ZY, Gangopadhyay A. (2008). A privacy protection model for patient data with multiple sensitive attributes. International Journal of Information Security and Privacy 2(3): 28-44.

[11] Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. (2007). L-diversity: Privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data 1(1): 3.

[12] Fung BCM, Wang K, Chen R, Yu PS. (2010). Privacy-preserving data publishing: A survey of recent developments, ACM Comput. Surv. 42(4): 1-53.

[13] LeFevre K, DeWitt DJ. (2005). Raghu Ramakrishnan, Incognito: Efficient fulldomain k-anonymity. In: Proceedings of 2005 ACMSIGMOD. International Conference on Management of Data, SIGMOD'05, pp. 49-60.

[14] Fung BCM, Wang K, Yu PS. (2007). Anonymizing classification data for privacy preservation. IEEE Trans. Knowl. Data Eng. 19(5): 711-725.

[15] LeFevre K, DeWitt DJ. (2006). R.Ramakrishnan, mondrian multidimensional k-anonymity. In: Proceedings of 22nd International Conference on Data Engineering, ICDE'06, p. 25.

[16] Xu J, Wang W, Pei J, Wang X, Shi B, Fu AWC. (2006). Utility-based anonymization for privacy preservation with less information loss. ACMSIGKDD Explor. News l8(2): 21-30.

[17] Dean J, Ghemawat S. (2010). MapReduce: A flexible data processing tool. Commun. ACM 53(1): 72-77.

[18] Samet S, Miri A. (2013). New incremental privacy – preserving clustering protocols. Lecture Notes On Software Engineering 1(3).

[19] Xiao X, Tao Y. (2007). M-invariance: Towards privacy preserving republication of dynamic datasets. In: Proceedings of the 2007 ACMSIGMOD International Conference on Management of Data, SIGMOD' 07, pp. 689-700.

[20] Lavrac N, Bohanec M, Pur A, Cestnik B, Debeljak M, Kobler A. (2007). Data mining and visualization for decision support and modeling of public health-care resources. Journal of Biomedical Informatics 40: 438-447.

[21] Rahman N, Harding JA. (2012). Textual data mining industrial knowledge management and text classification: A business oriented approach. Expert Systems with Applications 39: 4729-4739.

[22] Koluguri A, Gouse S, Bhaskara Reddy P. (2014). Text Steganography Methods and its Tools. International Journal of Advanced Scientific and Technical Research 2(4): 888-902.

[23] Pinkas B. (2002). Cryptographic techniques for privacy preserving in data mining. SIGKDD Explorations 4(2): 12-19.

[24] Agrawal A, Singh V. (2014). Securing video data: A critical review. International Journal of Advanced Research in Computer and Communication Engineering 3(5).

[25] Kantarcioglu M, Jiang W. (2013). Incentive compatible privacy-preserving data analysis. IEEE Transactions on Knowledge and Data Engineering 25(6).

[26] Kantarcioglu M, Kardes O. (2009). Privacy preservation of data mining in the malicious model. International Journal of Information and Computer Security 2: 353-375.

[27] Sweeney L. (2002). K-anonymity: A Model for protecting privacy. International Journal on Uncertainty, Fuzziness based Systems, 557-570.

[28] Lohiya S, Ragha L. (2012). Privacy preserving in data mining using hybrid approach. Fourth International Conference on Computational Intelligence and Communication Networks.