Chapter 4
Black Hole Algorithm for BigData Anonymization

U. Selvi and S. Pushpa

Introduction

Over the decades, Hardware and software have had a rapid growth toward the advancement of storage capability. All personal information regarding an individual is available online and usage of Databases has been booming. Small set of data can be easily handled by the Data Mining tools available but handling large set of data is a challenging task. Preserving privacy is another challenging issue when dealing with sensitive data. When dealing with data using traditional tools like relational Databases, data is found in the term of tuples where we have attributes that describe individuals. Four types of attributes that explicitly describes the individuals are as follows: explicit identifiers1, quasi-identifiers2, sensitive attributes, and non-sensitive attributes [1]. The widespread methodology for conserving privacy is anonymization which is to hide the explicitly identifying individual attributes and standard Algorithm for this is k-anonymity. But though such approach seems to be simple in implementation, which is not sufficient and individuals can be re-identified.

The flaw of straightforward k-anonymity was revealed [2, 3] and was further confirmed by de Montjoye et al. in [4]. By combining two datasets, L. Sweeney was successful in identifying the individuals by attack named "linking attack". K-anonymity was proposed to address those attacks and it becomes the base algorithm for the next subsequent algorithm related to anonymization. In k-anonymity, each record cannot be renowned among a minimum of k-1 records. Throughout this process, dataset is divided into several groups based on similarity classes and the proceedings of each group are generalized. Hence it becomes difficult to spot individuals in group, since all the individuals of equivalent groups are similar. Thus

U. Selvi (✉) · S. Pushpa
Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Avadi, Chennai 600054, India
e-mail: slvunnikrishnan@gmail.com

Ch. Satyanarayana et al. (eds.), *Proceedings of the International Conference on Computer Vision, High Performance Computing, Smart Devices and Networks*,
Advanced Technologies and Societal Change,
https://doi.org/10.1007/978-981-19-4044-6_4

the objective of anonymization is achieved. This approach seems to be similar to clustering-based approach in which each equivalence class is grouped as cluster. Clustering-based k-anonymity has a data utility as they group related records organized. Although the k-anonymity constructed clustering is theoretically simple, the computational complication of discovering an optimal k-anonymous solution is NP-hard [5]. In this context, great effort is required to provide a complete search for optimal solutions. In this case, it should be feature with minimum information loss. However, this process fails when the dataset increases and data suffer bad data quality. Many meta-heuristics approaches were found to be effective in those areas but more exploration is needed in terms of privacy and anonymity.

Meta-heuristic algorithms are optimization methods but require expensive computation time. Meta-heuristic algorithms are simple to implement and have a simple structure and reduced number of parameters; it is called as region Algorithm (BHA) and is free from parameter setting issues. BHA algorithm has never been applied to the problem of privacy-preserving and anonymization. BHA algorithm is summarized as follows: It is a population-centered algorithm that has the region of space. It has a gravitational force in which any object in the universe gets disappeared if it gets close to it. The BHA Algorithm [6] applied to k-anonymity problem signifies the k-anonymous solution and the top explanation given by black hole. The algorithm starts with an initial population of candidate clarifications produced accidentally. Its objective is to select the optimal k-anonymity result which has the minimum information loss. To enhance info quality, clustering algorithm group the similar quasi-identifiers within the group having a minimum of k records. The similarity is calculated supported information loss as a distance and cost metric. This makes sure that fewer misrepresentations are required to anonymize the record in a cluster which enhances data quality.

The rest of the paper is systematized as follows. Subsequent segment surveys correlated work around k-anonymity-centered approaches. Our algorithm is presented in Sect. 4.3 and is experimentally assessed in Sect. 4.4. We determine this paper in Sect. 4.5.

Related Work

k-anonymization becomes the standard Benchmark and base algorithm for much privacy-preserving algorithms. Since our proposed work is around k-anonymity, some literature survey is done around k-anonymity and cluster algorithm.

Anonymity based on clustering approaches

Clustering-based anonymization on attributes hierarchies by local recording was proposed [7]. Equivalence class was created and this approach tries to select the correspondence class of size lesser than k. It then calculates the space among C and catches the similarity C$'$ with the small distance to C. Lastly the two similar class is group and generalized. The process is repeated until the equivalence class has a minimum distance of k records. Weighted Feature C-means clustering for k-anonymity was projected [8]. The process begins as follows: C random records were selected as seeds and by calculating the number of equivalence classes. Then the algorithm starts to assign weights to each quasi-identifier. The process continues to identify records close to the seeds and feature weight is updated to reduce information loss. This process is iterated until no change is applied to the clusters of record. The algorithm merges small equivalence class which has k records with larger correspondence classes to satisfy the k-anonymity constraint.

K-member clustering was projected by Byun et al. [9]. The algorithm tries to build a cluster selecting record around the seed and form k-1 nearest record. Then, the algorithm selects the replacement record for the record farther from the seed and iterates the process to create a cluster. And also assigns the unassigned record to any closer clusters. The process ends when all the records are assigned. Greedy k-anonymity algorithm was proposed by Loukides and Shao [10]. This is similar to k-member clustering but differs from assigning cluster to a group awaiting a user demarcated threshold is extended. The cluster which has record lesser than ‘k’ records will be deleted.

One pass k-means Algorithm (OKA) was projected by Lin and Wei [11] to achieve anonymity-based clustering. During the first phase, k-mean algorithm was applied and during the second phase, clusters having records more than k records are adjusted by moving records to cluster having less than k records. Clustering-primarily based K-anonymity with a set of rules known as GCCG was suggested by Ni et al. [12]. This methodology is composed of four steps namely Grading, Centering, Clustering, and Generalization. In Grading and Centering steps, the facts are looked after primarily constructed totally at the rating computed of every file then the primary X facts are selected as centroids. The next stage is foundation of clusters through including to every centroid the k-1 closest facts. In the very last phase, the facts are generalized. To decorate the overall piece of GCCG set of rules, the authors additionally suggest a parallelized model of GCCG.

A clustering-primarily based totally k-anonymity set of rules which deliberates the general delivery of quasi-identifier businesses in a multidimensional space was projected by Zheng et al. [13]. The proposed set of rules first alternatives erratically a file r as a centroid of the primary cluster and provides the k-1 nearby facts to it, which will shape the primary cluster. Then the set of rules picks the file which has the biggest distance among itself and the primary centroid and sets it to the second one centroid. The ith centroid is created through with inside the identical way, primarily based totally on the space among the ith file and all of the happened centroids. Subsequently every step of centroid formation, the algorithm provides the k-1 closest facts to the centroid to shape the clusters. At the stop of this procedure, all of the clusters formed incorporate k facts, if there are ungrouped facts. The set of rules repeats the closing facts and enclosure every file into the nearest cluster, i.e., having the slightest space with its centroid.

An adaptive k-anonymity set of rules, known as AKA was proposed by Arava and Lingamgunta [14]. It is primarily centered totally on KOC’s regular approach for locating the fine seed values. AKA begins off evolved with computing the variety of clusters p = no tuples/k value. For complete file in every institution, it computes k-closeness with each different file and types them in descending order. Then, it units in each institution the facts with minimal and most closeness as preliminary centroids (i.e., 2 * p seeds) and builds the clusters. The closing facts are allotted to their adjacent clusters, such that each cluster ought to have k cluster individuals. The extra facts (i.e., that have sizes special to k) are restructured and attached to their nearest clusters. For the clusters with scopes advanced to k, the procedure generates new clusters with insignificant of k facts. A present clustering set of rules can be implemented to the stay clusters, with sizes not as good as k, to distribute their facts.

Weighted k-member clustering set of rules known as (WKMCA) was proposed by Byun et al. [9]. The proposed set of rules is a changed k-individual to lessen the have an effect on outliers at the clustering effect. For this, WKMCA provides a biased level wherein a chain of weighting signs was assigned to assess the outlyingness of facts which will expedite filtering out the outliers. Thereby, k-individuals are primarily centered totally on the ones signs to acquire k-anonymity.

k-anonymity primarily based totally on nature-stimulated optimization procedures:

Lunacek et al. [15] projected a brand new crossover operator and carried out a Genetic Algorithm- primarily centered totally k-anonymity method with the suggested crossover operator to reveal the gain of the usage of the brand new operator over traditional crossover operators. Lin and Wei [16] projected a Genetic Algorithm (GA)-primarily centered totally clustering method for accomplishing k-anonymity. In this method, the preliminary populace of GA is fashioned primarily based totally on Hybrid Method anticipated [17]. A candidate answer of populace encoded through a chromosome and includes no rarer than k genes, wherein every gene suggests the index of a report with inside the authentic dataset. The set of rules makes use of most effective choice and crossover operations of GA. Mutation isn't always completed because of the set of rules makes use of the authentic report indexes which cannot be reformed.

Run et al. [18] proposed a hybrid seek technique primarily based totally on Tabu Search (TS) and Genetic Algorithm (GA) to acquire k-anonymity. In the projected technique, TS is embedded right into a traditional GA to carry out the position of mutation. Bhaladhare and Jinwala [19] anticipated a Fractional Calculus-primarily centered totally Bacterial Foraging Optimization Algorithm referred to as FC-BFO to generate a most reliable clustering. The goal of FC-BFO is to enhance the optimization cap potential and convergence velocity of BFO set of rules through making use of to it the idea of FC in its chemotaxis step. Effectively, the FC-BFO gives a higher records loss and execution time than BFO.

Wai et al. [20] proposed a huge statistics private maintenance method primarily based totally on Hierarchical Particle Swarm Optimization (HPSO). The suggested method is constructed upon MapReduce Hadoop groundwork to deal with the scalability problems of huge statistics. It includes stages; The first degree is HPSO clustering. The set of rules generates a MapReduce activity to provide the predefined amounts of intermediate clusters, characterized through particles, then a MapReduce activity of HPSO clustering is accomplished on every cluster through iteratively appearing Map and Reduce stages till the quantity of statistics individuals in every particle beat k. In the second one degree, the occasioned clusters are generalized to be converted into their anonymized paperwork. The Map step truly permits all statistics individuals of every intermediary cluster to its corresponding Reduce step which per-paperwork HPSO clustering activity to provide k-anonymized clusters.

Madan and Goswami proposed hybrid optimization algorithms to acquire k-anonymity referred to as Dragon-PSO [21] and GWO-CSO method [22]. The Dragon-PSO set of rules associations the Dragonfly Algorithm (DA) and Particle Swarm Optimization (PSO) through adapting the replace system of DA the usage of PSO. Madan and Goswami [23] projected an anonymity version for statistics issuing primarily centered totally on K-DDD degree, Dragonfly operators-primarily centered totally Genetic Algorithm referred to as Duplicate-Divergence-Different homes enabled Dragon Genetic (DDDG) set of rules. The head step, referred to as k-DDD anonymization, is the transformation of authentic database to k-DDD database primarily centered totally at the projected k-DDD degree. k-DDD degree transforms the authentic database through producing "k" quantity of identical facts, "k" quantity of Divergence in touchy attributes, and "k" quantity of Unlike provider companies in every cluster of the database. The subsequent phase is implemented D-Genetic set of rules on k-DDD database. D-Genetic Algorithm is shaped via the change of Genetic Algorithm (GA) with the Dragonfly Algorithm (DA).

k-anonymity in BigData MapReduce

Several privacy preservation algorithms fit into the MapReduce framework to perform parallel execution of large datasets. This is to address the scalability problem of BigData. MapReduce framework computes using map and reduces functions. Data from distributed file system is divided into number of chunks and assigned to map function and secretes a list of key/value pairs. In the subsequent section, Reducer syndicates the values fitting to every distinct key permitting to several functions and engraves the result to an output file. Thus the MapReduce function solves the scalability problem of BigData. LeFevre et al. [24] spoke the scalability problem of anonymization algorithms using scalable decision trees and sampling techniques. Fung et al. [25] projected the Top-Down Specialization methodology to yield anonymous datasets without data exploration problem. Ke et al. [26] proposed the Bottom-up generalization to anonymize large datasets. Yavuz et al. [27] proposed the data anonymization in large-scale dataset generated in real-time application using the Apache spark.

Proposed Approach

K-anonymization based on black hole algorithm in Big Data (KAB-BD) was proposed in the following section. The KAB-BD starts with initial populations of stars in each chuck of Map phase in MapReduce framework, clustering-centered k-anonymous solutions, and then progresses the population to discover the superlative k-anonymous solution, i.e., having the lowest information loss which is done in Reduce phase of the MapReduce framework. As the projected algorithm originates from black hole algorithm, the development of population to an optimal resolution is completed by moving all the stars to the best solution, characterized by the black hole.

Algorithm: k-anonymity-based black hole algorithm (KAB) in MapReduce

i. Initialize the map function with original population of stars
ii. In each chunk of Map phase, estimate the fitness of every star and customary the top star as the b-hole
iii. The fitness is updated and all stars are moved toward the b-hole
iv. If a star touches the finest location than the b-hole, it suits the b-hole and vice versa
v. If a star becomes too close to the b-hole, a new star is created to replace the old one and its fitness is evaluated
vi. Output from each chunk of Map is fed as input to Reduce phase and step (iii), (iv), and (v) is repeated
vii. If determined amount of iteration is reached, the algorithm stops

Step: 1 Generation of Initial Population

In the first step, on each map function, candidate solutions were selected based on clustering algorithm where solution signifies a k-anonymous clustering. The records in a cluster must be as alike as likely to acquire worthy data quality. This guarantees that fewer distortions are desirable to generalize the records from the identical cluster as a consequence subsequent of getting worthy data quality. *Normalized Certainty Penalty* (NCP) [28] is one such metric used to attain this objective and cost dimension of clustering algorithm. NCP is a proficient and easy to usage metric that deals the degree of information loss produced by the anonymization method.

Step: 2 Evaluation and Selection of the Black Hole

In the evaluation step followed by initialization, fitness function for every star is calculated to find the best star. The star with the finest fitness value is the black hole selected. After modifying the population, the fitness function of every star is weighed and the finest star, which has the finest fitness value, is selected as the black hole.

Step: 3 Update the Positions and the Fitness of the Stars

In the next subsequent step, totally the candidates move near the top candidate which is the black hole. This move can be done by shifting the location of each star.

Three possible scenarios are found in these situations:

i. After shifting the stars to new locations, star can reach the finest location than black holes with lesser fitness value. In this case, star befits the black hole and vice versa (interchange their locations and fitness).

ii. Star marks the event horizon of the black hole; in such a case the star will be absorbed by the black hole and switched by a new star.
iii. Neither of the two previous scenarios nor in this occasion is the locations and the fitness just rationalized. Once all the stars are relocated, subsequent reiteration precedes place with the novel positions of stars and black holes and their equivalent objective functions.
iv. Step. 2 and step. 3 is iterated in Reduce phase until the supreme quantity of iteration is seen and the algorithm terminates.

Experimental Result

In this section, we calculate the quality of our proposed algorithm. For preserving privacy, data utility and information loss are the main objectives of every anonymization algorithm. KAB is measured by evaluating these two objectives. Anonymized data utility is measured based on the different privacy levels characterized by k.

Evaluation Metrics and Experimental Setup:

Metric preferred for this is Classification Accuracy (CA), to calculate the rate of k-anonymous clusters in the anonymized data. A cluster is measured *correctly classified* if it fulfills k-anonymity criterion. To assess the data utility of our algorithm, we have used metrics namely *Classification Metric* (CM) and *Average Equivalence Class Size Metric* (*C*AVG) [24]. Information loss experienced by the unique data after anonymization process can be measured by two commonly used metrics: *Total Information Loss* (Total-IL) and *Normalized Certainty Penalty* (NCP). Total Information Loss is defined as the loss of accuracy when take a broad view of specific attributes. NCP *measures* the classification errors by penalizing equivalence classes that contain rows with different class labels.

The carrying out tests were accomplished on a Desktop PC with Intel Core 2.10 GHz CPU and 4 GB of RAM under Ubuntu operating system. The executions were built and run with Hadoop. The adult dataset was taken from the UCI machine learning repository for anonymization. The dataset comprises census data has a total of 32,561 instances underneath 15 attributes. Each record has the personal information and personal income-related information. Preprocessing is done at initial stage for the removal of duplicate and missing values. In adult dataset, age and education are the numerical attributes and remaining attributes are the categorical attributes. Totally, dataset has 30,162 records.

Results and Discussion

Data Utility with different levels of privacy is done by considering three different algorithms (k-anonymization, KAB-based k-anonymization, and BHA-based k-anonymization) and the experimental results are shown in Figs. 4.1, 4.2, 4.3, and 4.4 Data Utility with respect to information loss is increasing with the increasing value of k represented by *GenTotal-IL* and *GCP*, correspondingly for all the three algorithms is depicted in Figs. 4.1 and 4.2.

K represents the number of records in a cluster. If the value of k is large, extra records are found in single cluster. It is found that, the KAB-based k-anonymization presents the minimum information loss with respect to privacy level. BHA-based

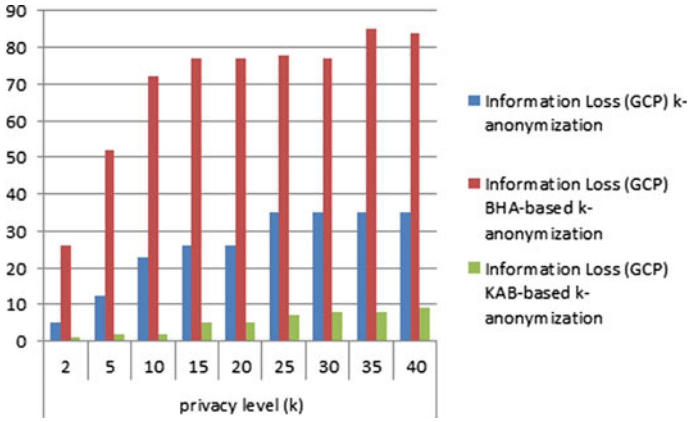

Fig. 4.1 Information loss (GCP) versus privacy level

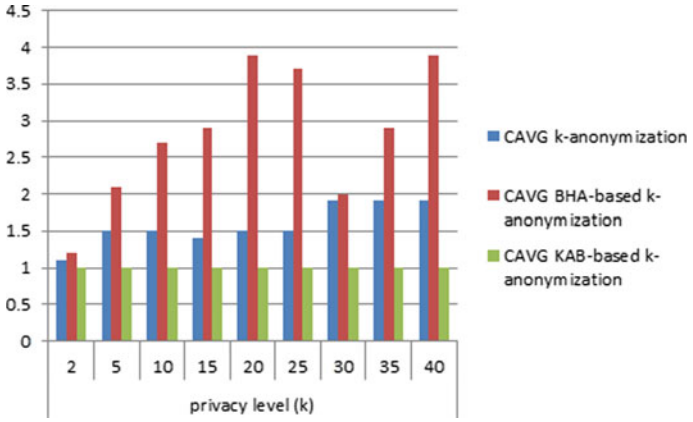

Fig. 4.2 *C*AVG versus privacy level

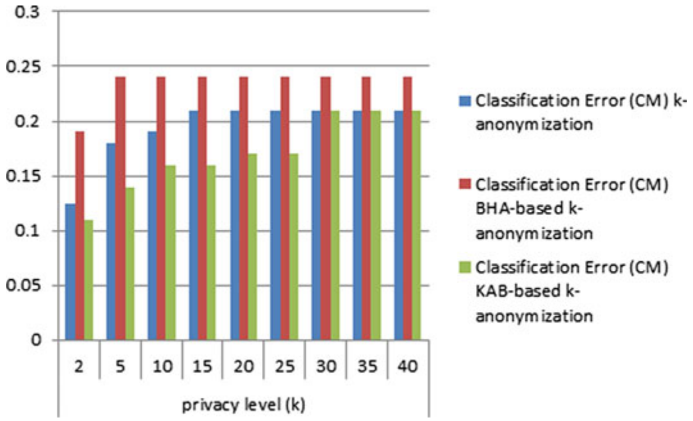

Fig. 4.3 Classification error (CM) versus privacy level

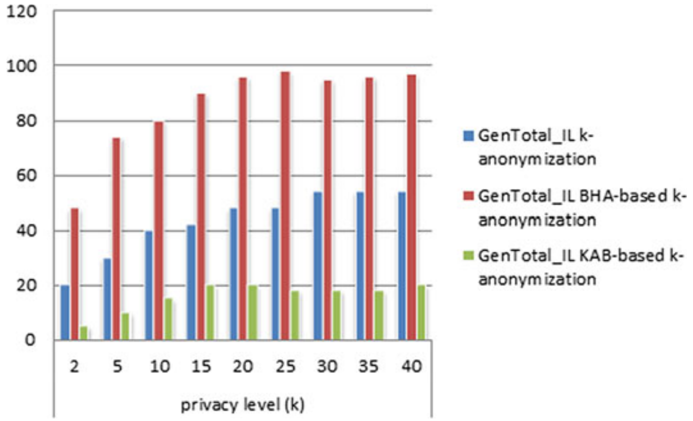

Fig. 4.4 Generalization Total Information loss (GCP) versus privacy level

k-anonymization has the poorest information loss since it doesn't follow any reliable metrics to organize the records in the cluster and the records are placed randomly. Also, BHA-based k-anonymization has the small convergence rate and requires a more quantity of iteration to cover.

It is found that KAB algorithm is an enhancement of BHA. KAB declines the distance among preliminary results and optimal results to accelerate the convergence rate BHA to create an optimal solution. Figure 4.3 considers the data utility of the entire three algorithms with respect to $CAVG$ metric as the value of k rises. It is the reflection of the information lost from previous figures. Based on the observation, it is found that Mondrian Multidimensional forms equivalence classes, of sizes adjacent to ideal case which clarifies the reasonable information loss familiarized by k-anonymization. KAB algorithm generates correspondence classes of ideal sizes, i.e., equal to 1, which has subsidized to reducing information loss of KAB-based k-anonymization. BHA-based k-anonymization generates equivalence classes of variable sizes, because the amount of clusters, which governs the size of the equivalence classes, is calculated randomly; the slighter the number of clusters is, the greater the sizes of equivalence classes are, and higher the information loss it.

Figure 4.4 reports data utility, of the algorithms, with respect to *CM* as the value of k increases. It shows that classification errors of the KAB-based k-anonymization and k-anonymization increase with the increase of k-value. Instinctively, the higher the class dimensions are, and better the probability of result classification errors is. Generally, CM presented by BHA-based k-anonymization is the same. From the figure, it is detected that KAB-centered k-anonymization declares less classification errors than the other algorithms.

Conclusion

K-anonymization centered on BHA (KAB) is proposed in this work. The main objective is to find the optimal clustering-based k-anonymity in Map Reduce framework. This starts with a population of clustering-centered k-anonymous candidate solutions, on which BHA is applied. To measure the efficiency, our approach is compared with k-anonymity, BHA-centered k-anonymity, and clustering-centered k-anonymity techniques, in terms of data utility and scalability. To reduce the execution time, the above said algorithm is implemented in MapReduce framework. The simulation results report that KAB algorithm in MapReduce outperforms all the compared techniques in terms of data utility and scalability. Data Utility can be further enhanced by increasing the number of iterations and/or stars. In our future work, this can be improved by implementing in machine learning based approach.

References

1. Ciriani, V., di Vimercati, S.D.C., Foresti, S., Samarati, P., Yu, T.: k-anonymity. In: Jajodia, S., Yu T. (eds.) Security in Decentralized Data Management. Springer (2006)
2. Sweeney, L.: Datafly: A system for providing anonymity in medical data. In: Database Security XI, pp. 356–381. Springer, Boston, MA (1998). https://doi.org/10.1007/978-0-387-35285-5_22
3. Sweeney, L.: k-anonymity: a model for protecting privacy. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **10**(05), 557–570 (2002). https://doi.org/10.1142/S0218488502001648
4. De Montjoye, Y.A., Radaelli, L., Singh, V.K.: Unique in the shopping mall: On the reidentifiability of credit card metadata. Science **347**(6221), 536–539 (2015). https://doi.org/10.1126/science.1256297
5. Meyerson, A., Williams, R.: On the complexity of optimal k-anonymity. In: Proceedings of the Twenty-Third ACM SIGMOD-SIGACT-SIGART Symposium on Principles of database systems, pp. 223–228 (2004). https://doi.org/10.1145/1055558.1055591. Moon, B., Jagadish, H.V., Faloutsos, C., Saltz, J.H.: Analysis of the clustering properties of the hilbert space-filling curve. IEEE Trans. Knowl. Data Eng. **13**(1), 124–141 (2001). https://doi.org/10.1109/69.908985

6. Hatamlou, A.: Black hole: A new heuristic optimization approach for data clustering. Inf. Sci. **222**, 175–184 (2013). https://doi.org/10.1016/j.ins.2012.08.023
7. Li, J., Wong, R.C.W., Fu, A.W.C., Pei, J.: Achieving k-anonymity by clustering in attribute hierarchical structures. In: International Conference on Data Ware-housing and Knowledge Discovery, pp. 405–416. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/11823728_39
8. Chiu, C.C., Tsai, C.Y.: A k-anonymity clustering method for effective data privacy preservation. In: International Conference on Advanced Data Mining and Applications, pp. 89–99. Springer, Berlin, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73871-8_10
9. Byun, J.W., Kamra, A., Bertino, E., Li, N.: Efficient k-anonymization using clustering techniques. In: International Conference on Database Systems for Advanced Applications, pp. 188–200. Springer, Berlin, Heidelberg (2007). https://doi.org/10.1007/978-3-540-71703-4_18
10. Loukides, G., Shao, J.: Capturing data usefulness and privacy protection in k-anonymization. In: Proceedings of the 2007 ACM symposium on Applied computing, pp. 370–374 (2007). https://doi.org/10.1145/1244002.1244091
11. Lin, J.L., Wei, M.C.: An efficient clustering method for k-anonymization. In: Proceedings of the 2008 International Workshop on Privacy and Anonymity in Information Society, pp. 46–50 (2008). https://doi.org/10.1145/1379287.1379297
12. Ni, S., Xie, M., Qian, Q.: Clustering based K-anonymity algorithm for privacy preservation. IJ Netw. Secur. **19**(6), 1062–1071 (2017)
13. Zheng, W., Wang, Z., Lv, T., Ma, Y., Jia, C.: K-anonymity algorithm based on improved clustering. In: International Conference on Algorithms and Architectures for Parallel Processing, pp. 462–476. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-05054-2_36
14. Arava, K., Lingamgunta, S.:Adaptive k-anonymity approach for privacy preserving in cloud.Arab. J. Sci. Eng.,1–8(2019).https://doi.org/10.1007/s13369-019-03999-0
15. Lunacek, M., Whitley, D., Ray, I.: A crossover operator for the k-anonymity problem. In: Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, pp. 1713–1720 (2006). https://doi.org/10.1145/1143997.1144277
16. Lin, J.L., Wei, M.C.: Genetic algorithm-based clustering approach for k-anonymization. Expert Syst. Appl. **36**(6), 9784–9792 (2009). https://doi.org/10.1016/j.eswa.2009.02.009
17. Lin, J.L., Wei, M.C., Li, C.W., Hsieh, K.C.: A hybrid method for k- anonymization. In: 2008 IEEE Asia-Pacific Services Computing Conference, pp. 385–390. IEEE (2008). https://doi.org/10.1109/APSCC.2008.65
18. Run, C., Kim, H.J., Lee, D.H., Kim, C.G., Kim, K.J.: Protecting privacy using k-anonymity with a hybrid search scheme. Int. J. Comput. Commun. Eng. **1**(2), 155 (2012)
19. Bhaladhare, P.R., Jinwala, D.C.: Novel approaches for privacy preserving data mining in k-anonymity model. J. Inf. Sci. Eng. **32**(1), 63–78 (2016)
20. Wai, E.N.C., Win, A.T., Tsai, P.W., Pan, J.S.: Privacy preservation in big data by particle swarm optimization. University of Computer Studies (Taunggyi) (2017)
21. Madan, S., Goswami, P.: A privacy preserving scheme for big data publishing in the cloud using k-anonymization and hybridized optimization algorithm. In: 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), pp. 1–7. IEEE (2018). https://doi.org/10.1109/ICCSDET.2018.8821140
22. Madan, S., Goswami, P.: A novel technique for privacy preservation using k-anonymization and nature inspired optimization algorithms. In: Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur, India (2019). https://doi.org/10.2139/ssrn.3357276
23. Madan, S., Goswami, P.: k-DDD measure and mapreduce based anonymity model for secured privacy-preserving big data publishing. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **27**(02), 177–199 (2019). https://doi.org/10.1142/S0218488519500089
24. LeFevre, K., DeWitt, D.J., Ramakrishnan, R.: Incognito: Efficient full-domain k-anonymity. In: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, pp. 49–60 (2005). https://doi.org/10.1145/1066157.1066164

25. Fung, B.C.M, Wang, K., Yu, P.S.: Anonymizing classification data for privacy preservation. IEEE Trans. Knowl. Data Eng. **19**(5), 711–725 (2007)
26. Ke, W., Yu, P.S., Chakraborty, S.: Bottom-up generalization: A data mining solution to privacy protection. In: Proceedings of 4th IEEE International Conference on Data Mining, ICDM'04, pp. 249–256 (2004)
27. Yavuz, C., Seref, S.: BigData anonymization with spark. In: IEEE International Conference on Computer Science and Engineering (UBMK) (2017)
28. Xu, J., Wang, W., Pei, J., Wang, X., Shi, B., Fu, A.W.C.: Utility-based anonymization for privacy preservation with less information loss. ACM SIGKDD Explor. Newsl. **8**(2), 21–30 (2006). https://doi.org/10.1145/1233321.1233324
29. Pramanik, M.I., Lau, R.Y., Zhang, W.: K-anonymity through the enhanced clustering method. In: 2016 IEEE 13th International Conference on e-Business Engineering (ICEBE), pp. 85–91. IEEE (2016). https://doi.org/10.1109/ICEBE.2016.024