ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.7, 4589–4596
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.7.27 Spec. Issue on RDESTM-2020

PRIVACY-PERSERVING KNN CLASSIFICATION PROTOCOL OVER ENCRPTED RELATIONAL DATA IN THE CLOUD

P. VINAYBHUSHAN[1] AND T. HIRWARKAR

ABSTRACT. With the recent popularity of cloud computing, clients presently have the chance to redistribute their data just as the data management tasks to the cloud. Notwithstanding, because of the ascent of different protection issues, touchy data (e.g., clinical records) should be encrypted before re- appropriating to the cloud. What's more, query preparing tasks ought to be taken care of by the cloud; in any case, there would be no good reason for redistributing the data at the primary spot. To process inquiries over encrypted data without the cloud ever decrypting the data is an extremely testing task. In this paper, we center on attempting the characterization issue over encrypted data. Specifically, we propose a safe k-NN classifier over encrypted data in the cloud. The proposed k- NN protocol ensures the privacy of the data, the client's information query, and data get to designs. As far as we could possibly know, our work is the first to build up a safe k-NN classifier over encrypted data under the standard semi-legit model. Likewise, we observationally investigate the proficiency of our answer through different examinations.

1. INTRODUCTION

The cloud computing worldview is reforming the associations' method of working their data especially in the manner they store, access, and procedure data. As a developing computing worldview, cloud computing draws in numerous associations to consider truly with respect to cloud potential as far as its

[1]*corresponding author*

2010 *Mathematics Subject Classification.* 68P25.

Key words and phrases. k-NN Classifier, privacy-preserving data mining, Encryption.

cost-productivity, adaptability, and offload of regulatory overhead. Regularly, associations delegate their computational activities notwithstanding their data to the cloud. In spite of huge focal points that the cloud offers, protection, and security issues in the cloud are forestalling organizations to use those focal points. At the point when data are profoundly touchy, the data should be encrypted before re-appropriating to the cloud. In any case, when data are encrypted, regardless of the fundamental encryption conspire, playing out any data mining tasks turns out to be trying while never decrypting the data.

Data Mining has wide applications in numerous territories, for example, banking, medication, logical research and among government offices. Characterization is one of the usually utilized tasks in data mining applications. For as far back as decade, because of the ascent of different protection issues, numerous hypothetical and useful answers for the characterization issue have been proposed under various security models. In any case, with the ongoing fame of cloud computing, clients presently have the chance to re-appropriate their data, in encrypted structure, just as the data mining tasks to the cloud. Since the data on the cloud is in encrypted structure, existing security safeguarding grouping methods are not pertinent. Data mining over encrypted data (indicated by DMED) on a cloud needs to ensure a client's record when the record is a piece of a data mining process. Besides, cloud can likewise determine valuable and delicate data about the real data things by watching the data get to designs regardless of whether the data are encrypted. Along these lines, the protection/security necessities of the DMED issue on a cloud are triple: (1) privacy of the encrypted data, (2) secrecy of a client's query record, and (3) concealing data get to designs.

The data owner re-appropriates his/her database and DBMS functionalities (e.g., kNN query) to an untrusted outside specialist co-op which deals with the data in the interest of the data owner where just believed clients are permitted to query the facilitated data at the specialist co-op. By re- appropriating data to an untrusted server, numerous security issues emerge, for example, data protection (shielding the classification of the data from the server just as from query guarantor).

1.1. **K-NN Algorithm.** The k-nearest neighbors' algorithm is a strategy for grouping objects dependent on the following preparing data in the element space. It

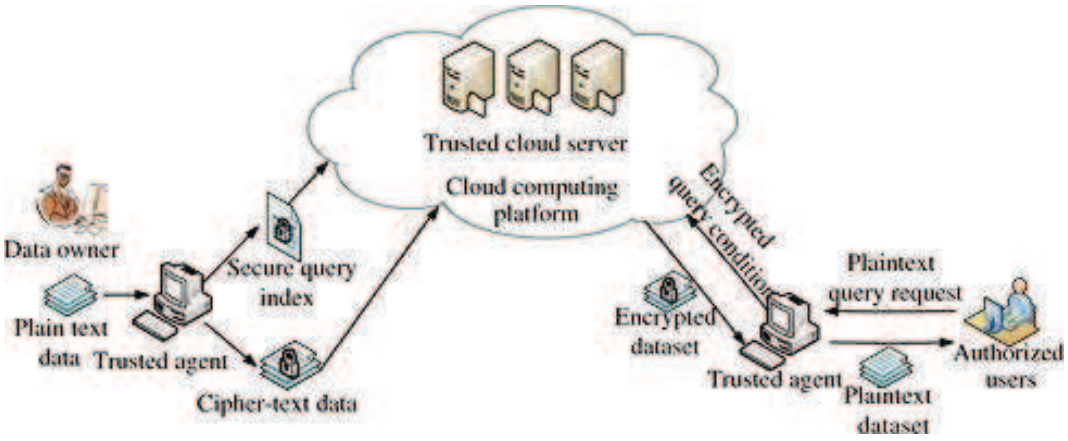

FIGURE 1. An efficient privacy preserved data model over cloud

is among least complex of all component learning calculations. The calculation works on a lot of d-dimensional vectors, $V = \{xi|I = 1...N\}$, where $xi \in kd$ indicates the i^{th} data point. The calculation is introduced by determination k focuses in kd as the underlying k bunch agents or "centroids". Methods for select these essential seeds incorporate inspecting aimlessly from the dataset, setting them as the arrangement of bunching a little subset of the data or annoying the worldwide mean of the data k times. At that point the calculation emphasizes between two stages till intersection:

Stage 1. Data Assignment every datum point is allot to its abutting centroid, with ties broken discretionarily. This outcomes in an apportioning of the data.

Stage 2. Relocation of "signifies". Each gathering delegate is migrating to the middle (mean) of all data focuses dole out to it. In the event that the data focuses accompany a chance measure (Weights), at that point the movement is to the desires (weighted mean) of the data allotments.

Privacy. Preserving Data Mining (either annoyance or secure multi-party calculation based methodology) can't take care of the DMED issue. Annoyed data don't have semantic security, so data irritation procedures can't be utilized to scramble exceptionally delicate data. Likewise the irritated data don't create exceptionally precise data mining results. Secure multi-party calculation (SMC) based methodology expect data are dispersed and not encrypted at each taking an interest party. What's more, many middle of the road calculations are performed dependent on non- encrypted data. Accordingly, in this paper, we

proposed novel strategies to adequately tackle the DMED issue expecting that the encrypted data are redistributed to a cloud. In particular, we center around the order issue since it is one of the most widely recognized data mining tasks. Since every grouping procedure has their own favorable position, to be solid, this paper focuses on executing the k-nearest neighbor order strategy over encrypted data in the cloud computing condition.

2. Proposed Model

Here, Ram owns database DA of n records $t1, ..., tn$ and $m + 1$ properties. Let t p,q means the q the property estimation of record tp . At first, Ricky scrambles his database trait astute, that is, he figures E pk (t p,q), for $1 \leq p \leq n, 1 \leq q \leq m+1$, where section (m + 1) contains the class marks. We expect that the hidden encryption plot is secure. Leave the encrypted database alone signified by DB́. We expect that John redistributes DB́ just as the further arrangement procedure to the re-appropriated data. Let Seetha be an approved client who needs to arrange his information record $r = hr$, r m p by applying the k-NN grouping strategy dependent on DB́ . We allude to such a procedure as security saving k-NN arrangement over encrypted data in the cloud. Officially, we characterize the security saving kNN protocol as: PP-kNN (DÁ , r) $\rightarrow$ cr Where c r means the class name for r subsequent to applying k-NN arrangement strategy on DÁ and r.

Let see, W is the Whole System Consist of W= Qu, PPKNN, E', PRKNN, PCMCK, PPP.

Where Qu is set of query entered by user.

Qu=q1, q2, q3,. . . .qm.

E' = Encrypted Data set.

PPKNN = privacy-preserving process k-NN, PRKNN = protected Retrieval of k-Nearest Neighbours, PCMCK = Protected Computation of Majority Class, PPP = Privacy-Preserving Primitives.

The proposed PP-kNN protocol fundamentally comprises of the accompanying two phases:

Stae 1: Protected Retrieval of k-Nearest Neighbors (PRkNN). In this stage, User at first sends his query q (in encrypted structure) to C1. After this, C1 and C2 include in a lot of sub-protocols to safely recover (in encrypted structure) the

class. Labels comparing to the k-nearest neighbors of the information query q. At the finish of this progression, encrypted class names of k-nearest neighbors are known distinctly to C1.

Stage 2: Protected Computation of Majority Class (PCMCk): C1 and C2 mutually process the class mark with a larger part casting a ballot among the k-nearest neighbors of q. At the finish of this progression, just User realizes the class name relating to his information query record q.

2.1. **PPP.** Here we present a lot of nonexclusive sub-protocols that will be utilized in building our proposed k-NN protocol. The entirety of the beneath protocols are considered under two-customer semi legitimate setting. Specifically, we consider the nearness of two semi legitimate customers P1 and P2 to such an extent that the Palliser's mystery key sk is known uniquely to P2 while ik is open. PMIN: In this protocol, P1holds private info (ú, ý) and P2 holds sk, where ú = ([u],Epk(su)) and ý = ([v],Eik(sv)). Here su (resp., sv) means the mystery related with u(resp., v). The objective of SMIN is for P1 and P2 to mutually Here we present a lot of nonexclusive sub protocols that will be utilized in developing our proposed k-NN protocol .All of the underneath protocols are considered under two-customers semi-legitimate setting. Specifically, we consider the nearness of two semi legit customers P1 and P2 with the end goal that the Palliser's mystery key sk is known uniquely to P2 while ik is open.

2.2. **PMINn.** In this protocol, we consider P1 with n encrypted vector's ([d1], [dn]) alongside their comparing encrypted privileged insights and P2 with sk. Here[dp] = hEik(dp,1), . . . ,Eik(dp,l)i where dp,1 anddi,l are the most and least noteworthy bits of whole number independently, for $1 \leq p \leq n$. The secretor dp is given by sdi . P1 and P2 mutually compute [min(d1, . . . , dn)]. Moreover, they register Epk(smin(d1,...,dn)). Toward the finish of this protocol ,the yield ([min(d1,..., dn)],Epk(smin(d1,...,dn)))is known distinctly to P1. During SMINn, no data in regards to any of dp's and their privileged insights is uncovered to P1 and P2. PF: Here P1 with private info (hEik(c1),...,Eik(cw)p, hEik(ć1),... ,Eik(ćk)p) and P2 safely process the encryption of the recurrence of cq , indicated by f(cq), in the listhc'1,..., c'kp, for $1 \leq q \leq w$. Here we expressly accept that cq́s are novel and cṕ $\in$ {c1,..., cw}, for $1 \leq p \leq k$. The yield Eik(f(c1)),..., Eik(f(cw))p will be known distinctly to P1. During the SF

protocol, no data in regards to cṕ, cq , and f(cq) is uncovered to P1 and P2, for $1 \leq p \leq k$ and $1 \leq q \leq w$.

3. PERFORMANCE EVALUATION

The exhibition of the PPkNN protocol is assessed under various parameter settings. The calculation expenses of Stage 1 in the PPkNN protocol were investigated first by changing estimations of number of k-nearest neighbors (Figure 2). The Paillier encryption key size K was either 512 or 1024 bits. The calculation cost of Stage 1 for K=512 bits fluctuated from 9.98 to 46.16 minutes when k differed from 5 to 25, separately. Conversely, the calculation cost of Stage 1 for K=1024 bits fluctuated from 66:97 to 309:98 minutes when k differed from 5 to 25, separately. In either case, the calculation time of Stage 1 developed straightly with k. Furthermore, for some random k, the expense of Stage 1 expanded by very nearly a factor of 7 at whatever point K multiplied. For instance, when k=10, Stage 1 required 19.06 and 127.72 minutes to create the encrypted class names of the 10 nearest neighbors under K=512 and K=1024 bits, individually.

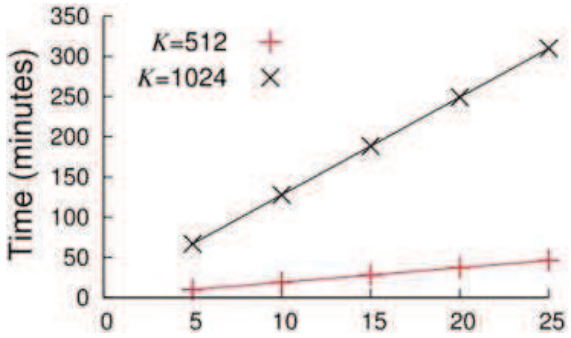

FIGURE 2. Computation costs of the PPkNN protocol

Besides, when k=5, one could see that around 66:29% of the expense in Stage 1 accounted due to SMINn which started k times in the PPkNN protocol (once in every cycle). Additionally, the expense brought about due to SMINn expanded from 66:29% to 71:66% when k differed from 5 to 25. Seetha's calculation cost in the PPkNN protocol was basically because of the encryption of his information query. In the dataset, Seetha's calculation cost was 4 and 17 milliseconds

when K was 512 and 1024 bits, separately. It was obvious that PPkNN protocol was productive from Bob's computational point of view which was particularly advantageous when he gave inquiries from an asset compelled gadget.

CONCLUSION

In this paper, to secure user protection, different protection safeguarding characterization procedures have been proposed. By the by, the current strategies are not appropriate in redistributed database conditions where the data lives in the encrypted structure on an outsider server. Along this bearing, this paper proposed novel protection saving k-NN arrangement protocol over encrypted data in the cloud. Our protocol secures the classification of the data, client's info query, and conceals the data get to designs. We likewise assessed the presentation of our protocol under various parameter settings. The exhibition of our PPkNN protocol, we intend to research option and progressively proficient answers for the SMINn issue, we utilized the notable k-NN classifier and built up a protection safeguarding protocol for it over encrypted data.

REFERENCES

[1] S. DE CAPITANI DI VIMERCATI, S. FORESTI, P. SAMARATI: *Managing and accessing data in the cloud: Privacy risks and approaches*, 7th International Conference on Risk and Security of Internet and Systems (CRiSIS), Cork, 2012, 1-9, doi: 10.1109/CRISIS.2012.6378956.

[2] Y. ELMEHDWI, B. K. SAMANTHULA, W. JIANG: *Secure k-nearest neighbor query over encrypted data in outsourced environments*, 30th IEEE International Conference on Data Engineering (ICDE), Chicago, IL, 2014, 664-675, doi: 10.1109/ICDE.2014.6816690.

[3] C. GENTRY, S. HALEVI: *Implementing gentry's fully-homomorphic encryption scheme*, EUROCRYPT, Springer-Verlag, (2011), 129–148.

[4] W. HENECKA, S. KÖGL, A.-R. SADEGHI, T. SCHNEIDER, I. WEHRENBERG: *Tasty: tool for automating secure two-party computations*, In ACM CCS, (2010), 451–462.

[5] B. HORE, S. MEHROTRA, M. CANIM, M. KANTARCIOGLU: *Secure multidimensional range queries over outsourced data*, The VLDB Journal, **21**(3) (2012), 333–358.

[6] H. HU, J. XU, C. REN, B. CHOI: *Processing private queries over untrusted data cloud through privacy homomorphism*, IEEE ICDE, (2011), 601–612.

[7] Y. HUANG, D. EVANS, J. KATZ: *Private set intersection: Are garbled circuits better than custom protocols?*, NDSS, 2012.

[8] Y. HUANG, J. KATZ, D. EVANS: *Quid-pro-quo-tocols: Strengthening semi-honest protocols with dual execution*, IEEE Symposium on Security and Privacy, (2012), 272–284.

[9] S. KESAVAN, S.E. KUMAR, A. KUMAR, K. VENGATESAN: *An investigation on adaptive HTTP media streaming Quality-of-Experience (QoE) and agility using cloud media services*, International Journal of Computers and Applications (2019), https://doi.org/10.1080/1206212X.2019.1575034.

[10] P. KUMAR, V. ANBARSU, R. VIJAYALAKSHMI, K. VENGATESAN: *Intellectual Resource Sharing Scheme in Cloud Environment*, Jour of Adv Research in Dynamical & Control Systems, **11**(10) (2019), 6 pages.

[11] K. VENGATESAN, E. SARAVANA KUMAR, S. YUVARAJ, P. SHIVKUMAR TANESH, A. KUMAR: *An Approach for Remove Missing Values in Numerical and Categorical Values Using Two Way Table Marginal Joint Probability*, International Journal of Advanced Science and Technology, bf 29(5) (2020), 2745-2756.

[12] K. VENGATESAN, S. YUVARAJ, S. SAMEE, P. SHIVKUMAR TANESH, A. KUMAR: *Prediction of the petrol consumptions by using data mining decision classifier classification algorithm*, Test Engineering and Management, (2020), 22206–22212.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,
SRI SATYA SAI UNIVERSITY OF TECHNOLOGY & MEDICAL SCIENCES,
MADHYA PRADESH, INDIA.
Email address: pillalamari.vinay@gmail.com

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,
SRI SATYA SAI UNIVERSITY OF TECHNOLOGY & MEDICAL SCIENCES,
MADHYA PRADESH, INDIA.