*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 6, Ver. IV (Nov.-Dec. 2016), PP 19-25*
*www.iosrjournals.org*

# Enhancing Cloud Security by DES through K-NN Computation

K. Sindhuja[1], B. Lakshmi Ramani[2], T. Praveen[3]
[1](CSE, PVPSIT/ JNTUK, India)
[2](CSE, PVPSIT/ JNTUK, India)
[3](ECM, PVPSIT/ JNTUK, India)

***Abstract:*** *Protecting databases or data contents from the web world environment is a tough task for a company. Because every Company/ Financial Institute/ Hospital was hiding their customers or end users list secretly and will not open for all. But now Tom's gang (Hackers) made this possible and tries stealing the data and major portion Tom's gang was successful in doing this one. In these conditions securing the data outsourcing area such as web hosting and cloud space storage option are becoming very prominent. To manage the situation many were out with secured sharing solutions. Now one more novel approach with high secured and efficient sharing option in data retrieving by end user is demonstrating in this paper. The technique is comprises with two famous algorithms one is DES an encryption scheme and the next is K-NN query passing and data retrieving code. Distance and verification for nearest values will generate the query passing value. The files were secured by using DES encryption mode.*

***Keywords:*** *Tom's gang, Cloud spacing, Secured, Sharing, DES, K-NN, Query.*

## I. Introduction

In this paper cloud computing as an emerging technology is expected to reshape information technology processes in the near future. Due to the overwhelming merits of cloud computing, e.g., cost-effectiveness, flexibility and scalability, more and more organizations choose to outsource their data for sharing in the cloud. As a typical cloud application, an organization subscribes the cloud services and authorizes its staff to share files in the cloud. Each file is described by a set of keywords, and the staff, as authorized users, can retrieve files of their interests by querying the cloud with certain keywords. In such an environment, how to protect user privacy from the cloud, which is a third party outside the security boundary of the organization, becomes a key problem. User privacy can be classified into search privacy and access privacy. Search privacy means that the cloud knows nothing about what the user is searching for, and access privacy means that the cloud knows nothing about which files are returned to the user. [5]

Range query is one of the most frequently used queries for online data analytics. Providing such a query service could be expensive for the data owner. With the development of services computing and cloud computing, it has become possible to outsource large databases to database service providers and let the providers maintain the range-query service. Without sourced services, the data owner can greatly reduce the cost in maintaining computing infrastructure and data-rich applications. However, the service provider, although honestly processing queries, may be curious about the hosted data and received queries. Most existing encryption based approaches require linear scan over the entire database, which is inappropriate for online data analytics on large databases. While a few encryption solutions are more focused on efficiency side, they are vulnerable to attackers equipped with certain prior knowledge. Random Space Encryption (RASP) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches [6].

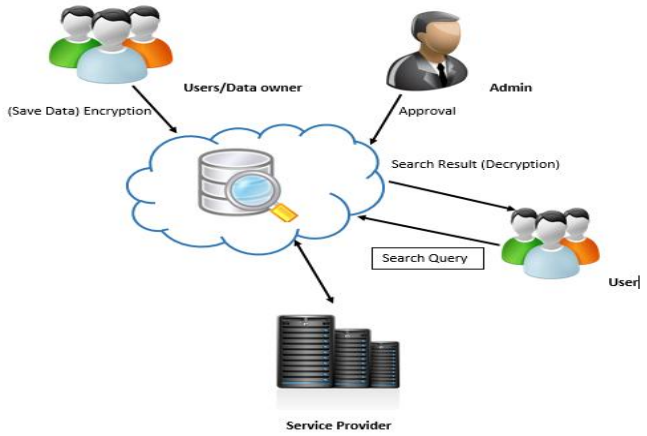


**Fig1:** Basic Data outsourcing through cloud space.

This is a very challenging task, as conventional encryption does not support processing on top of cipher texts, whereas more recent cryptographic tools such as homomorphism encryption are not flexible enough(they support only restricted operations),and they are also prohibitively expensive for practical uses. To address this problem, previous work such as has proposed privacy-preserving data transformations that hide the data while still allowing the ability to perform some geometric functions evaluation. However, such transformations lack the formal security guarantees of encryption. Other methods employ stronger-security transformations, which are used in conjunction with dataset partitioning techniques, but return a large number of false positives, which is not desirable due to the financial considerations outlined earlier.

## II. Literature Survey

Because of the specificity of such information, gathering and keeping up such data is a costly procedure, and moreover, a portion of the information might be delicate in nature. For example, certain dissident gatherings might not have any desire to discharge their occasions to the overall population, because of concerns that enormous enterprises or severe governments may mediate and trade off their exercises. Thus, a few gatherings may want to keep their geo-labeled datasets secret, and just available to trusted subscribed clients, for the dread of reaction from more moderate populace bunches. It is in this way essential to shield the information from the cloud benefit supplier. What's more, because of money related contemplations in the interest of the information proprietor, sub-scribing clients will be charged for the administration in light of a compensation for each result model. For example, a supporter who requests k-NN results will pay for k-items, and ought not to get more than k-results. Consequently, rough questioning strategies with low accuracy, for example, existing systems that arrival numerous false encouraging points notwithstanding the real results, are not attractive.

Inquiry handling that jellies both the information security of the proprietor and the question protection of the customer is another exploration issue. It demonstrates expanding significance as distributed computing drives more organizations to outsource their information and questioning administrations. Nonetheless, most existing studies, including those on information outsourcing, address the information protection and inquiry security independently and can't be connected to this issue.

Cell phones with geo-situating abilities (e.g., GPS) empower clients to get to data that is applicable to their present area. Clients are keen on questioning about purposes of intrigue (POI) in their physical vicinity, for example, eateries, bistros, progressing occasions, and so forth. Substances represented considerable authority in different zones of intrigue (e.g., certain specialty headings in expressions, diversion, travel) accumulate a lot of geo-labeled information that interest to subscribed clients. Such information might be delicate because of their substance. Moreover, staying up with the latest and applicable to the clients is not a simple errand, so the proprietors of such datasets will make the information available just to paying clients. Clients send their present area as the inquiry parameter, and wish to get as result the closest POIs, i.e., closest neighbors (NNs). Be that as it may, normal information proprietors don't have the specialized intends to bolster handling questions on a substantial scale, so they outsource information stockpiling and questioning to a cloud benefit supplier. Numerous such cloud suppliers exist, who offer capable stockpiling and computational frameworks easily. In any case, cloud suppliers are not completely trusted, and commonly carry on in a fair yet inquisitive manner. In particular, they take after the convention to answer questions accurately, yet they additionally gather the areas of the POIs and the endorsers for different purposes. Spillage of POI areas can prompt to security ruptures and in addition money related misfortunes to the information proprietors, for whom the POI dataset is an essential wellspring of income. Divulgence of client areas prompts to protection infringement and may deflect endorsers from utilizing the administration inside and out. In this paper, we propose a group of strategies that permit preparing of NN inquiries in an un-trusted outsourced environment, while in the meantime securing both the POI and questioning clients' positions. Our procedures depend on variable request saving encoding (mOPE), the main secure request safeguarding encryption technique known to-date. We additionally give execution improvements to diminish the computational cost natural to handling on encoded information, and we consider the instance of incrementally upgrading datasets. We show a broad execution assessment of our procedures to outline their feasibility practically speaking.

In distributed computing, information proprietor utilizes information and questioning administrations for outsourcing on the cloud information. Amid this procedure, information is the different and private resource of the information proprietor thus that must be secured against cloud and questioning customer. Question which is let go by the customer may reveal the delicate points of interest/data of the customer. Henceforth ought to be shielded in cloud and from information proprietors. Along these lines, one of the significant issues in distributed computing is to ensure both, information protection and inquiry security among the information proprietor, the customer, and the cloud allude Fig-1. The informal communication is one of the rising segments confronting such kind of protection issue [2]. Distributed computing is new stage to sending, overseeing, and giving answer for the different sorts of capacity, stage issues utilizing web based handling. Be that as it may, it is exceptionally delicate issue to transfer our own information on the cloud since information protection is the enormous issue

and real issue of security. Delicate data must be encoded before outsourcing, which makes the powerful foundation. The administrations, for example, Goggle Docs, Amazon EC2, Microsoft Azure, and Online record stockpiling and so on are the cases of distributed computing and they are broadly utilized by many individuals around the world. Information use administrations and that is huge testing assignment. One of the strategies of recovery called Symmetric Searchable Encryption (SSE) of scrambled information on the cloud yet at the same time there is spillage of information security. Secure server –side positioning, which depends on the request saving Encryption (OPE), additionally incorporates the closeness significance and heartiness [3]. For the security of the information, different general arrangement in as of late research papers are saved to show consider on the information protection, the broadest arrangement in as of late done research papers are encryption. It implies information kept administration supplier must be scrambled to dodge data spillage on the cloud. Agrawal et al [4] proposed one of the arrangements in order to arrange protecting encryption plot (OPES) by which, lists can be assembled straightforwardly on figure content. The different SQL articulations, for example, MAX, MIN, COUNT, GROUP BY and ORDER BY can then be revamped and prepared over the scrambled information. In any case, OPES does not bolster SUM or AVG proclamations, if there should arise an occurrence of SUM and AVG unique information must be unscrambled first. In private Information recovery (PIR) for concealing a client's inquiry totally and giving solid security and secrecy, question anonymisation more often than not utilizes k-Anonymity [5] and its variations to blend the client's question with other loud question information. In [6], [7], client security and information protection is viewed as together. Yonghong Yu and WenyangBai talked about how to uphold information security and client protection over outsourced database benefit in [8]. Hu et al. [9] proposed one of the arrangements in light of secure traversal structure and protection homomorphism based encryption plot.

In the original k-nearest neighbor (KNN) classification method, no classifier model is built in advance. KNN refers back to the raw training data in the classification of each new sample. Therefore, one can say that the entire training set is the classifier. The basic idea is that the similar tuples most likely belongs to the same class (a continuity assumption). Based on some pre-selected distance metric (some commonly used distance metrics are discussed in introduction), it finds the k most similar or nearest training samples of the sample to be classified and assign the plurality class of those k samples to the new sample. The value for k is pre-selected. Using relatively larger k may include some pixels not so similar pixels and on the other hand, using very smaller k may exclude some potential candidate pixels. In both cases the classification accuracy will decrease. The optimal value of k depends on the size and nature of the data. The typical value for k is 3, 5 or 7.

The steps of the classification process are:

1) Determine a suitable distance metric.
2) Find the k nearest neighbors using the selected distance metric.
3) Find the plurality class of the k-nearest neighbors (voting on the class labels of the NNs).

Assign that class to the sample to be classified.

## III. Methodology

In this paper, we propose a family of techniques that allow processing of NN queries in an un-trusted out-sourced environment, while at the same time protecting both the POI and querying users' positions. Our techniques rely on data encryption scheme (DES), which guarantees in-distinguish-ability under ordered chosen-plaintext attack (IND-OCPA).We also provide performance optimizations to decrease the computational cost inherent to processing on encrypted data, and we consider the case of incrementally updating datasets. Inspired by previous work in that brought together encryption and geometric data structures that enable efficient NN query processing, we investigate the use of Voronoi diagrams and Delaunay triangulations to solve the problem of secure outsourced kNN queries. We emphasize that previous work assumed that the contents of the Voronoi diagrams is available to the cloud provider in plaintext, whereas in our case the processing is performed entirely on cipher-texts, which is a far more challenging problem.

Our proposed methods for secure nearest-neighbor evaluation perform query processing on top of encrypted data, and for this reason they are inherently expensive. It is a well-known fact that achieving security by processing on encrypted data comes at the expense of significant computational overhead. Next, we propose two optimizations that aim at reducing this cost and secure protocols for processing k-nearest-neighbor queries (kNN) on R-tree index is given. In the authors following work [7], they integrated indexing techniques with secure multiparty computation (SMC) based protocols to construct a secure index traversal framework. In this framework, the service provider cannot trace the index traversal path of a query during evaluation, and hence keep privacy of users. Their protocols for query are complex, and hard to implement. To solve private processing of more specific queries, different techniques have been implemented, e.g. public data column and private data column are implemented by hashing in. But join by hashing is unable to retrieve other specific as well as relevant data columns. Some time before a paper published by researchers proposes k-NN queries by processing private & remotely using homomorphism encryption [2].

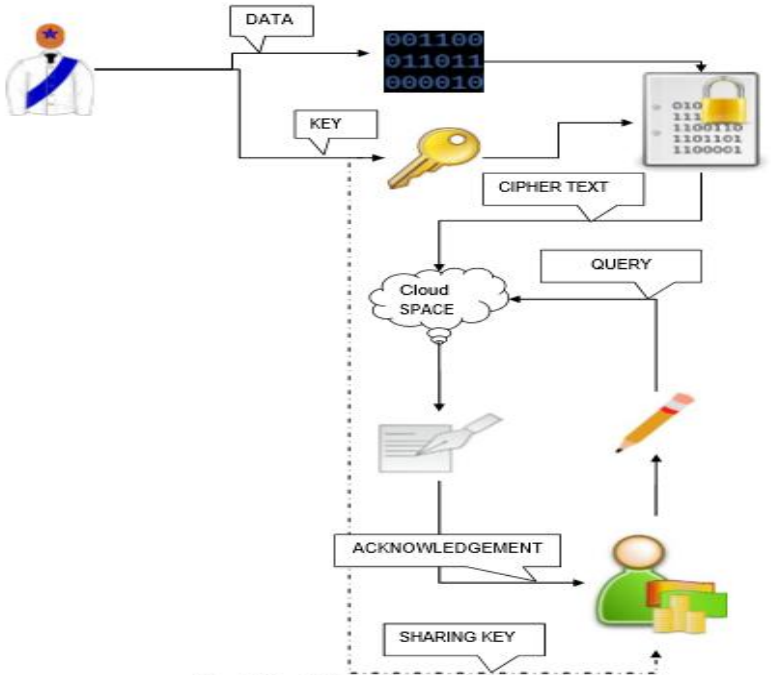


**Fig2:** Block Diagram of Proposing Scheme

Theoretical protocols using homomorphic encryption have been proposed to process private document search by specific keywords in a line of documents .These protocols are still too costly to use practically, and they perform only approximated search. Finally, we are not concerned to private query processing on outsourced encrypted data although our data bucketization is inspired by the data bucketization idea in a work from that area [12].Our approach may also apply to protect query privacy in outsource scenarios.Spatial database is a database that is optimized to store and query data that represents objects defined in a geometric space. Most spatial databases allow representing simple geometric objects such as points, lines and polygons. Some spatial databases handle more complex structures such as 3D objects, topological coverage's, linear networks, and TINs. While typical databases are designed to manage various numerics and character types of data, additional functionality needs to be added for databases to process spatial data types efficiently.

As mentioned previously, the dataset of points of interest represents an important asset for the data owner, and an important source of revenue. Therefore, the coordinates of the points should not be known to the server. We assume an honest-but-curious cloud service provider. In this model, the server executes correctly the given protocol for processing kNN queries, but will also try to infer the location of the data points. It is thus necessary to encrypt all information stored and processed at the server. To allow query evaluation, a special type of encryption that allows processing on cipher texts is necessary. In our case, we use the DES technique from [6].DES is a provably secure order-preserving encryption method, and our techniques inherit the IND-OCPA security guarantee against the honest-but-curious server provided by DES. Furthermore, we assume that there is no collusion between the clients and server, and the clients will not disclose to the server the encryption keys.

The server receives the dataset of points of interest from the data owner in encrypted format, together with some additional encrypted data structures (e.g., Voronoi diagrams, Delaunay triangulations) needed for query processing. The server receives kNN requests from the clients, processes them and returns the results. Although the cloud provider typically possesses powerful computational resources, processing on encrypted data incurs a significant processing overhead, so performance considerations at the cloud server represent an important. The client has a query point Qand wishes to find the point's nearest neighbors. The client sends its encrypted location query to the server, and receives k-nearest neighbors as a result. Note that, due to the fact that the data points are encrypted, the client also needs to perform a small part in the query processing itself, by assisting with certain steps.

Data Owner sends to Server the encoded Voronoi cell vertices coordinates, MBR boundaries for each cell, encoded right-hand side, and encrypted, for each cell edge. Client sends its encoded query point to the Server. Server performs the filter step, determines for each kept cell the edges that intersect the vertical line passing through the query point and sends the encrypted slope, of the two edges to the Client.

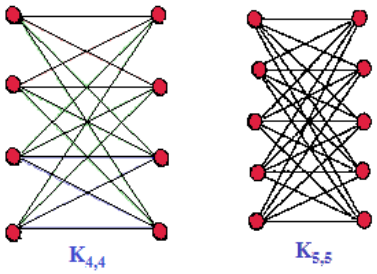


**Fig3:** K-NN Clustering diagram

Client computes the left-hand side, encodes it and sends it to the Server. Server finds the Voronoi cell enclosing the query point and returns result to Client.

To support secure kNN queries, where k is fixed for all querying users, we could extend the VD-1NN method from by generating order-k Voronoi diagrams .However, this method, which we call VD-k NN, has several serious drawbacks:

(1) The complexity of generating order- k Voronoi diagrams is either depending on the approach used or significantly higher than for order-1 Voronoi diagrams.

(2) The number of Voronoi cells in an order- k-Voronoi diagram, or roughly when k<<n. That leads to high data encryption over head at the data owner, as well as prohibitively high query processing time at the server (a k-fold increase compared to VD-1NN). Motivated by these limitations of VD-kNN, we first intro-duce a secure distance comparison method (SDCM).

Next, we devise Basic kNN (BkNN), a protocol that uses SDCM as building block, and answers kNN queries using repetitive comparisons among pairs of data points. BkNN is just an auxiliary scheme, very expensive in itself, but it represents the starting point for Triangulation kNN (TkNN), presented TkNN builds on the BkNN concept and returns exact results for k=1. For k>1, it is an approximate method that provides high-precision kNN results with significantly lower costs.

**Algorithm for Data Encryption with KNN**

**Input:** Data to be encrypted (Di)
**Step 1:** Every 8th bit of the unknown key is an odd parity
**Step 2:** Remove parity based on permutations
i.e, i= first bit of last byte of 8 bytes
(Ki),K(i-8),….K(i+1), K((i+1)-8)…., K(i+2), K((i+2)-8),…
key permutated bits after removing parity bits
**Step 3:** Split keys into right block and left block for the remaining 36 bits
L(0)=P(1:28)
R(0)=P(29:56)
**Step 4:** For 1<=i<=16 (i.e ., 16 sub keys need to generate)
**Step 5:** Applying left circular shift to generate 16 sub keys
**Step 6:** L(i)=LS[i] L[i-1]
**Step 7:** R(i)= LS[i] R[i-1] //here LS is left shift
**Step 8:** K[i]=P2[C(i) D(i)]
**Step 9:** End for
**Step 10:** Process 64 data bits (db)
**Step 11:** Permutation of db result in
J= second bit of last byte of 8 bytes
Db= db(j) d(j-8) …. d(j+2) , db((j+2)-8) , ….
**Step 12:** Split db into L and R blocks
dl(0)=dp(1:32)
dr(0)= dp(33:64)[i-1]
**Step 13:** For i<=1 <=16
**Step 14:** dl[i]=R[i-1]
**Step 15:** dr[i]=L[i-1] XOR F(R[i-1],k[i])
**Step 16:** Chipper text is cb=PP[(drc16) d(c16)]

## IV. Results

Experimental Results For Accuracy Versus Various Clusters

| CLUSTERS | K-NN in % | K-NN DES in % |
|---|---|---|
| 3 | 80 | 90 |
| 5 | 80 | 85 |
| 7 | 75 | 83 |
| 9 | 73 | 80 |
| 10 | 72 | 78 |
| 20 | 70 | 75 |
| 30 | 65 | 75 |

**Table:** different clusters versus accuracy.

As the number clusters increasing the data size get increasing so the retrieving possibility is getting reduced time by time. Even the time of execution is getting increased, so to push these entire flaws aside and trying to attain data retrieving condition formidably. But this needs to be more relevant while passing the query an encrypted will find less in size than a normal. So, the encrypted data is easily retrievable and due cluster framing accurate results are obtained all these were tested under different modes of uploading, sharing, encrypting, decrypting and downloading the data with in different cluster. This helps in obtaining an average accuracy of above 80% for overall performance under different cluster schemes.

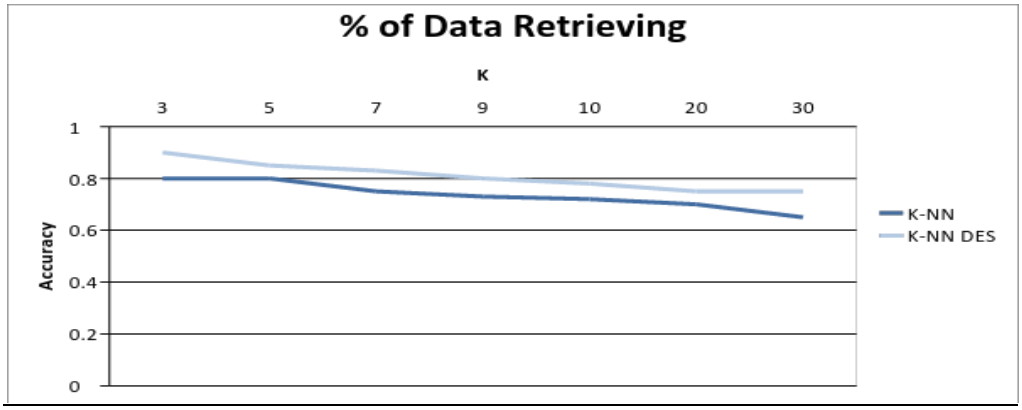


**Fig4:** Chart for Accuracy.

## V. Conclusion

Acquiring an efficient encrypting scheme with a secured data retrieving scheme provides data for analyses and this helps in not only providing security to data but also helps in acquiring the data speedily in terms of time and execution in a cloud storage. So, the possibility of data retrieving and passing query was very much easy in this arena. Previously encryption and decryption of cloud data with other clustering algorithms were mentioned but a result of mixed versions. Now, a dual security approach on cloud data by using k-NN query passing with DES encryption provides perfect data sharing and results in accurate search, which helps many users identify the data.

## References

[1]. Yousef Elmehdwi, Bharath K. Samanthula and Wei Jiang, *"Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments",*

[2]. Hu, Haibo, et al. *"Processing private queries over untrusted data cloud through privacy homomorphism." Data Engineering (ICDE), 2011 IEEE 27th International Conference on. IEEE, 2011.*

[3]. Nandhini, N., and P. G. Kathiravan. *"An Efficient Retrieval of Encrypted Data In Cloud Computing."*

[4]. Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and YirongXu. *Order preserving encryption for numeric data. In Proceedings of the 2004 ACM SIGMOD international conference on Management of data, SIGMOD '04,pages 563–574, New York, NY, USA, 2004. ACM.*

[5]. P. Samarati and L. Sweeney. *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, 1998.*

[6]. TingjianGe, Stanley B. Zdonik, and Stanley B. Zdonik. *Answering aggregation queries in a secure system model. In VLDB, pages 519–530, 2007.*

[7]. Haibo Hu and JianliangXu. *Non-exposure location anonymity.* In Yannis E. Ioannidis, DikLun Lee, and Raymond T. Ng, editors, *ICDE, pages 1120–1131. IEEE,2009.*

[8]. Yonghong Yu and WenyangBai. *Enforcing data privacy and user privacy over outsourced database service. JSW, 6(3):404–412, 2011.*

[9]. Hakan Hacgm, BalaIyer, and Sharad Mehrotra. *Efficient execution of aggregation queries over encrypted relational databases.* In Yoon Joon Lee, Jianzhong Li, Kyu-Young Whang, and Doheon Lee, editors, *Database Systems forAdvanced Applications, volume 2973 of Lecture Notesin Computer Science, pages 125–136. Springer Berlin Heidelberg, 2004.*

[10]. Varghese, Jiss, and Lisha Varghese. *"Homomorphic Encryption for Multi-keyword based Search and Retrieval over Encrypted Data."*

[11]. C. Gentry. *Fully homomorphic encryption using ideal lattices. In STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing, pages 169– 178, 2009.*

[12]. Josep Domingo-Ferrer. *A provably secure additive and multiplicative privacy homomorphism. In Proc. 5th International Conference on Information Security, 2002*