International Journal of Scientific Research & Engineering Trends
Volume 9, Issue 4, Jul-Aug-2023, ISSN (Online): 2395-566X

Category Attack-Based Searchable Symmetric Encryption Using Des Algorithm

Dr.G.Ramesh , Dr.J.S. Kanchana, V.Lekhaa
ramesh11182013@gmail.com kanchjs@gmail.com, lekhaakavi@gmail.com, Sivagangai Dist,
Department of Information
K.L.N. College of Engineering
Tamilnadu,India

Abstract- Symmetric searchable encryption (SSE), which allows a facts consumer to soundly seek and dynamically replace the encrypted documents stored in a semi-trusted cloud server, hasreceived considerable attention in recent years. We design the new data structure Category Attack-based SSE to support dynamic updating and boost verification and Leverage the timestamp mechanism within side the scheme to save you the malicious cloud from launching a replay attack. We can achieve more efficient query and verification with Data Encryption Standard Algorithm. By sampling the data, we can solve the problem of unbalanced distribution of network data. To look for functions that high-quality replicate the distinction among anomalous behaviors and normal behaviors Feature selection is enabled for various subsets of each category attacks. To determine the best sampling ratio of each category, DES is used to optimize the sampling ratio of each category and the performance of SSE is used to evaluate candidate sampled data. We verify the effectiveness of the data optimization proposed in this system, the precision, recall, and F1 score obtained by testing. Then we offer an in-depth overall performance analysis. Finally, we compare our scheme through complete experiments. The results are consistent with our analysis and show that our scheme is secure, and more efficient compared with the previous methods with the same functionalities.

Keywords-Des, Preprocessing, Sampling, False Alarm Rate, Feature Extraction

I. INTRODUCTION

Cloud computing entails using shared technology which include virtualization and cloud orchestration. Thus, through exploiting vulnerabilities in any a part of those technologies, attackers can motive good sized harm to many cloud users. Cloud Computing is a brand new surroundings in computer-orientated services. This system has some similarities of distributed system, according to this similarities cloud computing also uses the features of networking. Therefore, the security is the biggest problem of this system, because the services of cloud computing are based on the sharing on demand services via Internet is provided by cloud computing using large amount of virtual storage. Though cloud carriers use cryptographic algorithms to guard records in storage, they commonly use limited reasserts of entropy (along with the time) to mechanically generate random numbers for statistics encryption.

For instance, Linux-primarily based totally digital machines generate random keys handiest from the precise millisecond. This won't be sufficient for robust records encryption, however, as attackers additionally use state-of-the-art deciphering mechanisms to hack information. Thus, cloud builders have to reflect on consideration on a way to stable information earlier than it actions to the cloud. Its maximum vital function is that person has no want to set up expensive computing infrastructure and pay very much less for its services. Security is massive venture of cloud computing. Cloud computing lets in the person to get admission to assets everywhere whenever through net that is certainly the primary purpose in the back of the a couple of types of attacks. This paper consists of a exam primarily based totally on a theoretical survey on cloud computing that communicated diverse possible threats and also taxonomy model where at each layer a number of various types of category attacks enter from the utilization of various cloud services, furthermore for those assaults proposed mechanisms and the answers to be had earlier.Many SSE schemes use two-party model Data owners and servers, here we use three-party model administrators, Database servers, and users.

This scheme proposes an efficient SSE scheme based on the attack category using the DES algorithm Although there have been many schemes we use, but still attack category are many in implementing this scheme. Data sampling can resolve the trouble of unbalanced distribution of community data. First, we differentiate this process of attack category using the dataset. Preprocessing of the dataset is the input to the dataset in data samplings. The Dataset parameters and the feature description will be shown in this module implementation. Sampling the dataset into various features with the values referenced accordingly Second, we categorize different

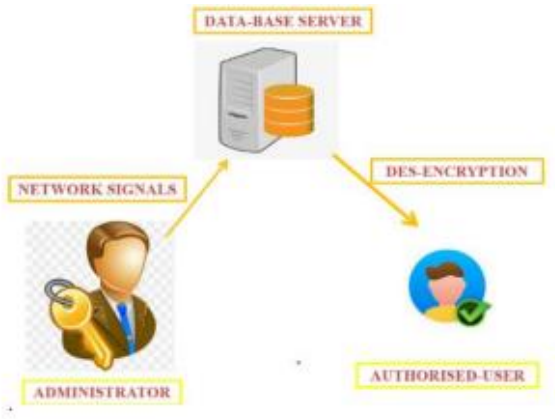
types of attacks with different types of protocols and label category encryption as 0 or 1. Then the number of occurrences of each category of attack is labelled to show the optimal feature subset of each category of anomalous behaviours Finally, we obtain the best sampling ratio of each category by using DES to optimize the sampling ratio of each attack and the performance of SSE used to evaluate candidate sampled data. Here Our proposal is an efficient SSE scheme that uses the attack category and the DES algorithm to analyse the attack category. To evaluate this scheme through comprehensive experiments, and show that scheme is secure, and more efficient compared with the previous schemes with the same functionalities.

II. RELATED WORK

1. System Model Administrator- Administrator is an individual, who's chargeable for configuring, Commissioning and preservation of community infrastructure and services. In an organization, Network Administrator commonly don't normally get involved without delay with users, rather attention upon configuring, tracking and protection of community additives inside organization's LAN/WAN infrastructure. DATABASE SERVER: The systematic approach to collecting data with predetermined properties with clients of the database servers in a structured manner. network signals with source IP, Destination IP, source Port, Destination port and various feature with attribute values holding state of the protocol and services be to offered. Containment set service provided to users, specifically for the request user query.

Authorized User- The user updates the query to the database server and analyses the search query with the database server, where authorized users can search the data of the cloud server. We consider cloud servers to be untrusted and have no trust in them. malicious server may perform active attacks. The search results may contain sensitive information that the server may extract from the search results. Moreover, when a client queries, the servers may return partial search results to the user. The server can perform re-play attacks. We are attempting to mitigate these attacks in this scheme.

To protect the server or user from various attacks, we use our methodology and analyse the attack types based on the frequent visits of other attacks to prevent data from malicious servers. We categorize the attacks according to the various protocols and features. Compared with the past methodology to prevent malicious server from returning partial results or tampered results, integrity check should be supported. However, most of the schemes ignore this or other issues. The integrity of the search results returned by the server to the user [4], [2], verifiable SSE schemes have also been extensively studied [1], [2], [3], [4], , unfortunately, these scenarios only support validation in static databases [8], [9].



Our goal is to design an efficient and privacy protection system for encrypted data, these categories of attacks are beyond this document will be discussed. Safeguarding Privacy: safeguarding privacy is the main requirement; encrypted sets can privilege attackers with anomalous servers so that it should be kept secret from cloud servers. Efficiency: In order to achieve an improvement in the effectiveness of attackers from the sample datasets are taken as input provided with the features. To do so. in the proposed system, we also aim to attain efficiency.

III. THE PROPOSED METHOD

In this approach, we present our scheme in sampling data set of network distribution over LAN/WAN into a defined set of features whereas, we introduce feature extraction in input data sets, that may be easy in the analysis of DES, Finally, the random distribution of the network is summarized into the defined set with features so that we can discover the attacker's occurrences in the network.

1. Preprocessing Of Data - The issue of an imbalanced distribution of network data can be resolved via data sampling. The Australian Centre for Cyber Security (ACCS)'s cyber security research team has produced the UNSW-NB15 dataset. The dataset is broken up into a training set and a testing set and has 2,540,044 records with 42 attributes. The test set has 82,332 records, whereas the training set has 175 341 records. In this module's implementation, the dataset parameters and the feature description are displayed. The precision, recall, F1 score, Accuracy, and FAR are achieved by evaluating the provided model in order to confirm the efficacy of the data optimization proposed in this system.

2. Attack Category Classification -The best feature subset for each category of anomalous behaviours is displayed in this classification. by integrating the machine learning method with the sampling technique to extract representative training data. Finding traits that most accurately capture the distinction between abnormal and typical behaviour is the process of feature selection. This

module displays the ideal sampling ratio for each Attack-Category discovered during data sampling. It displays each category of anomalous behaviours' ideal feature subset. It should be noticed that among the subset of ideal characteristics, the Normal category has the most features. Each category attack's frequency is normalized into a table to indicate how frequently it occurs through different transaction protocols and is referenced by a numerical value.

3. Des Analysis- The performance of SSE is used to assess candidate sampled data in order to find the best sampling ratio for each category, and DES is used to optimize the sampling ratio for each category. The F1 score is taken as the fitness function in this suggested system. The F1 score is a harmonic function that considers recall and precision. The F1 score of the best feature subset for each category of anomalous behaviors is displayed. Additionally, it displays the precision and false alarm rate (FAR) across all categories.

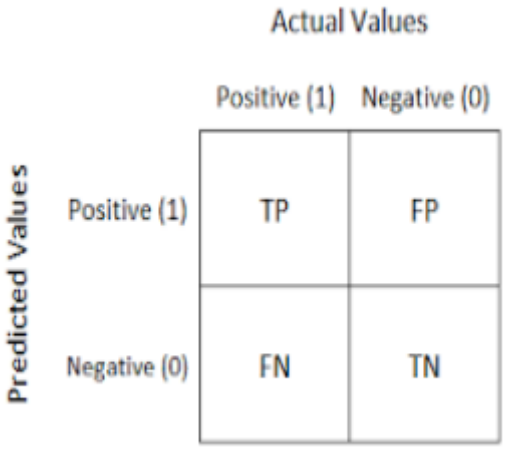
4 . Experiments - Experiments are carried out to validate the proposed methods, including the attack occurrence of each type of attacks via various protocols from the source IP and the destination IP in numbers, so that we may configure our results in our experimentally proposed approach. All trials are run on a 64-bit version of Windows 7, and our PC has an i5-6500 CPU, a 500 GB hard drive, and 5 GB of RAM. By empirically assessing the functional points of each attack category with the predominate values of precision and recall and by computing with confusion matrix values of each attack occurrence with the formulation of the score reached, we calculate the fitness score.

Table 1 Sse Des

ATTACK-CATEGORY	PRECISION	RECALL	F1-SCORE
Normal	0.91	0.971	0.939
Exploits	0.835	0.761	0.796
Fuzzers	0.957	0.415	0.578
Backdoor	0.25	0.513	0.336
DoS	0.045	0.472	0.459

Fitness Score The following formulas are used for the Fitness score measurements. If there are just two potential results, such as yes or no, positive or negative, etc., these notions are applied in binary classification. A false positive occurs when a positive outcome is predicted or determined incorrectly, a false negative occurs when a negative outcome is incorrectly predicted or measured, and a true positive occurs when you correctly predict or measure a positive outcome. Another indicator of how closely a measurement or prediction corresponds to the actual value is precision. It is determined by dividing the total number of accurate positive forecasts or measurements by the total number of positive forecasts or measurements. True positive and false positive are denoted by TP and FP, respectively, in the formula you previously used. When there are only two possible outcomes, such as yes or no, positive or negative, etc., these concepts are employed in binary classification. ATTACKCATEGORY PRECISION RECALL F1-SCORE Normal 0.91 0.971 0.939 Exploits 0.835 0.761 0.796 Fuzzers 0.957 0.415 0.578 Backdoor 0.25 0.513 0.336 DoS 0.045 0.472 0.459 When you accurately forecast or measure a positive outcome, it is called a true positive; when you do so, it is called a false positive.

Table 2 Confusion Matrix



In binary classification, precision is calculated as follows: Precision = $TP / TP + FP$ TP-True Positive value TN-True Negative value FP-False Positive value FN-False Negative Value For recall and f1 score summarized by the equation Another measure of how well a measurement or prediction fits the actual value is recall. It is determined by dividing the total number of actual positive outcomes by the proportion of correctly positive predictions or measurements. True positive and false negative are denoted by TP and FN, respectively, in the formula you previously used.

When there are only two possible outcomes, such as yes or no, positive or negative, etc., these concepts are employed in binary classification. A false negative is when you mistakenly forecast or measure a negative outcome when the actual outcome is positive. A true positive is when you accurately predict or measure a positive outcome. Recall in binary classification is calculated as follows: The percentage of accurate positive predictions or measurements is provided by this formula Recall = $TP / TP + FN$ An evaluation of a model's performance in binary classification—where there are only two possible outcomes—is done by calculating its F1 score. It is calculated as the harmonic mean of precision and recall, two more metrics for model effectiveness that

you are already familiar with. In binary classification, the F1 score is calculated as follows:

$F1$ score = 2 × Precision × Recall / Precision + Recall This equation provides you with a number between 0 and 1, with 1 being the best and 0 being the worst. A model with a high F1 score will have high recall and accuracy, which implies that it will accurately predict or measure the majority of positive outcomes while not making a lot of erroneous predictions or measurements. For normal attacks.

$F1$ score = 2 × Precision × Recall / Precision + Recall =2*0.910*0.971/0.910+0.971 =0.939 A low F1 score indicates that the model has low precision, low recall, or both, implying that it misses many good events or makes many incorrect predictions or measurements. The following equation determines binary classification accuracy: Accuracy = $TP + TN$ / $TP + FP + FN$+ TN The frequency with which a system or model issues a false warning or alarm when no actual event or result has occurred is known as the false alarm rate. It is determined by dividing the total number of negative results by the proportion of falsely positive predictions or measurements.

The letters FP stand for false positive and TN for true negative in your previous equation. When there are only two possible outcomes, such as yes or no, positive or negative, etc., these concepts are employed in binary classification. A genuine negative is when you accurately forecast or measure a negative consequence, whereas a false positive is when you predict or measure a positive outcome when the actual outcome is negative. The binary formula for false alarm rate is formulated by the notion of the formula as FAR = FP / $FP + TN$ A confusion matrix, also known as an error matrix is a specific table layout in the field of machine learning that allows Each row of the matrix represents an actual class, whereas each column represents a predicted class, or vice versa - both variations are available in the literature.

The name comes from the fact that it makes it clear whether the system is confusing two classes (i.e. frequently mislabelling one as another. True Positive (TP) is the number of actual anomalous records classified as anomalous ones, True Negative (TN) is the number of actual normal records classified as normal ones, False Positive (FP) is the number of actual normal records classified as anomalous ones, False Negative (FN) is the number of actual anomalous records classified as normal ones Evaluating Category attacks Classifier How can we use those metrics and what we can read from the confusion matrix? For instance, let's consider a classical problem of normal and category attacks, by using binary classification model. Our dataset consists of175 341 82,332 attack category types that are normal, and category attacks that are to evaluate the performance of our developed model, which labels category attacks as normal and other vulnerable attacks as 0 or 1, we can use confusion matrix, where the outcome is formulated in a 2×2 contingency table or a confusion matrix :Altogether, the classifier made 82,332 predictions (82,332 category attacks were classified in normal, and category attacks) Out of 2,540,044 records with 42 attributes of other category attacks, the test set has 82,332 records, whereas the training set has 175 341 records. This result to 95% accuracy.Further,82,332 out of 175 341 category attacks were classified falsely: category attacks, which were actual Category attacks, were not predicted as Category attacks (False Negative). And more important, no category attacks were falsely predicted as Category attacks (False Positive), which is very desired in this case. We can observe that our model is very conservative when it comes to predicting Category attacks. Therefore, the precision of this of this model is very high: 1.0.By computing additional measures (also called rates) from the classification matrix, we can get additional insight about our model.

Macro score The arithmetic mean, also known as the unweighted mean, of all the per-class F1 scores is used to calculate the macro-averaged F1 score, also known as the macro F1 score. Regardless of the support values, all classes are treated equally by this method. The macro-averaged F1 score in our categorization report is the same as the figure of 0.641 that we calculated above.

Table 3 Overall Macro score

TABLE 3
Overall Macro score

Sl.No	Feature Measurements	Macro-Values
1	Accuracy	0.975
2	False-Alarm Rate	0.343
3	Macro-Precision	0.619
4	Macro -Recall	0.665
5	Macro-F1-Score	0.641

Sl.No Feature Measurements MacroValues 1 Accuracy 0.975 2 False-Alarm Rate 0.343 3 Macro-Precision 0.619 4 Macro -Recall 0.665 5 Macro-F1-Score 0.641 5

IV.PERFORMANCE ANALYSIS AND COMPARISON

The precision, recall, and F1 score achieved by evaluating the proposed model are displayed in order to demonstrate the efficacy of the data optimization suggested in this system. As can be shown, SSE-DES has performed well in detecting network anomalies that exhibit uneven data distribution. This module compares the false alarm rate

(FAR) and accuracy of simple RF and DO IDS across all categories. Additionally, it compares the proposed method's FAR and accuracy to those of competing machine learning techniques. It is clear that the accuracy and FAR of the suggested method have significantly improved. Attacks are the strategies that attackers use to take advantage of the vulnerabilities in applications. Attacks are regularly harassed with vulnerabilities, so please try and make sure that the assault described is something that an attacker might do, as opposed to a weak spot in an application.

SSE-DES F1 VALUE COMPARISON Here, we compare several SSE scheme features with various categories of attacks. To obtain meaningful results for each experiment, we evaluate multiple attack strategies. We contrast our plan with the RF approach and the intrusion detection method for data optimization. Here, a high negligible value comparison is done to determine the value of privacy preservation.

To get threshold values that are additive, different attack category values are normalized ACCURACY COMPARISON Here we compare the accuracy of every schemes used in the method. Accuracy of every attack category is described and compared with the already existing models like RF method, Intrusion detection method While IDS solutions are important tools in monitoring and detecting potential threats, they are not without their challenges. the existing scheme measures the accuracy because of two solutions False alarms: Also referred to as fake positives, those depart IDS answers liable to figuring out ability threats that aren't a real threat to the organization.

Table 4 Attackcategory Rf Method Do-

ATTACK-CATEGORY	RF-METHOD	DO-IDS METHOD	SSE-DES METHOD
Normal	0.865	0.935	0.951
Exploits	0.902	0.926	0.936
Fuzzers	0.876	0.953	0.964
Backdoor	0.978	0.98	0.985
DoS	0.915	0.931	0.945
Generic	0.989	0.99	0.998
Reconnaissance	0.984	0.988	0.989
Shellcode	0.984	0.992	0.995
Analysis	0.972	0.982	0.989
Worms	0.989	0.998	1

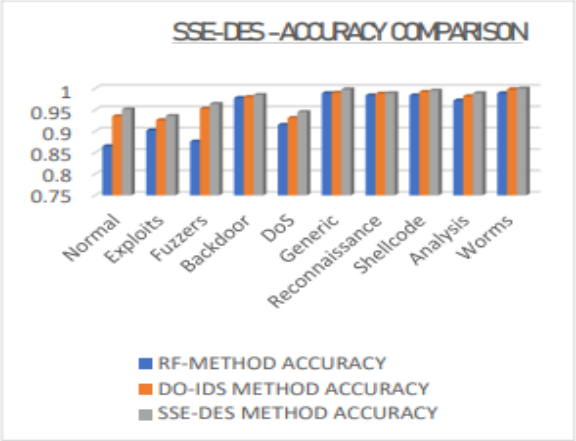


Figure 2

To keep away from this, companies have to configure their IDS to recognize what ordinary seems like, and as a result, what must be taken into consideration as malicious activity. False negatives: This is a larger concern, because the IDS answer errors an real protection hazard for valid traffic. An attacker is authorized to by skip into the organization's network, with IT and safety groups oblivious to the truth that their structures were infiltrated FALSE ALARM RATIO (FAR) fashionable magnificence of strategies of detecting anomalies in a laptop gadget which might be primarily based totally on heuristics or synthetic intelligence techniques.

These strategies are to differentiate among ordinary and anomalous device behaviour's foremost weak point of those strategies is a fake alarm rate that is generally measured through counting fake-positives value on a pattern set representing ordinary behaviour this dimension a base price of anomalous behaviour in a.stay surroundings isn't always taken into consideration and that results in a base-price fallacy. This trouble can substantially have an effect on a actual range of fake alarms which may be appreciably more than anticipated.

Table 5

Attack-Category	Rf-Method	Do-Ids Method	Sse-Des Method
Normal	0.124	0.033	0.03
Exploits	0.303	0.337	0.327
Fuzzers	0.971	0.619	0.607
Backdoor	0.937	0.597	0.54
DoS	0.583	0.539	0.519
Generic	0.033	0.031	0.029
Reconnaissance	0.849	0.18	0.169
Shellcode	0.183	0.22	0.21
Analysis	0.997	0.939	0.815
Worms	0.218	0.205	0.187

Attack Category RfMethod Do-Ids Method Sse-Des Method Normal 0.937 0.597 0.54 DoS 0.583 0.539 0.519 Generic 0.033 0.031 0.029 Reconnaissance Every

measurement's macro score value is calculated by averaging the results from each method to get a finite value matching to other schemes. Implementing The SSE-DES method yields accurate results for other macro values referenced when compared to other existing method.

Table 6

METHOD	ACC*	FAR	PRE*	RECALL	F1-S*
RF-METHOD	0.865	0.519	0.489	0.487	0.488
DO_IDS-METHOD	0.928	0.37	0.616	0.63	0.623
SSE-DES-METHOD	0.975	0.343	0.619	0.665	0.641

Performance Evaluation -In this part, we assess the scheme's effectiveness using statistics on the accuracy and false alarm rate of various types of attacks. We can observe that the false alarm rate is very low when the precision is good. This allows us to distinguish between common assaults and server-vulnerable attacks. We have put our suggested plan into practice utilizing a PHP front end and a MYSQL database and XAMPP framework on the back end to demonstrate its viability. The prototype contains more than 4,000 lines of code. We used a 64-bit version of Windows 7, and our computer's specs include

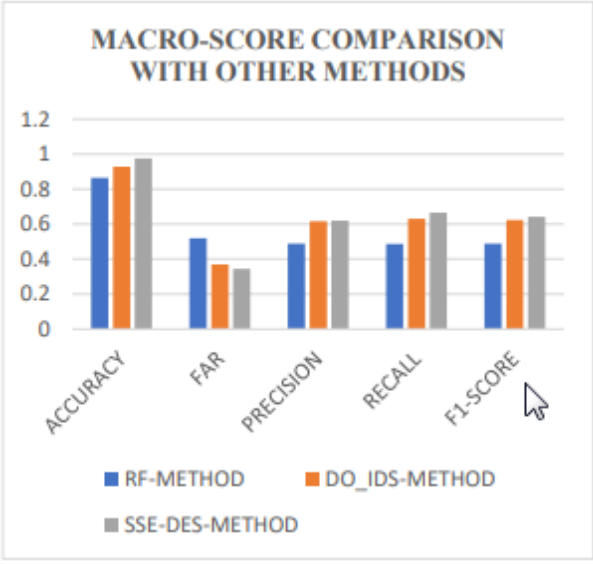


Figure 4

Computer's specs include an i5-6500 CPU, a 500 GB hard drive, and 5 GB of RAM. The Australian Centre for Cyber Security (ACCS)'s research team on cyber security generated the UNSW-N B15 dataset, which we use.

Comparison With Existing Schemes- A great deal of SSE techniques uses data outsourcing mechanisms instead of security assaults. This SSE method assesses the attacks using a dataset that has a vast number of entries and offers incredibly promising results. By contrasting this approach with the wellknown dynamic SSE solution and the globally verified solution, we confirm its viability. The research team in cyber security at the Australian Centre for Cyber Security (ACCS) recently generated the UNSW-NB15 dataset. The dataset, which is split into a training set and a testing set, has 2, 540,044 records with 42 properties.

Figure 5

When compared to other existing approaches such as KNN, MLP, SVM, DO-IDS, the intrusion detection systems appear to be extremely accurate and overhead for keeping a large number of records with 42 characteristics. We validate the solution's viability by comparing it to the well-known dynamic SSE solution and the universal verifiable solution

VI. FUTURE DIRECTION

Future study can consider more complex searches, such as conjunction query search and multiple keyword extraction. It could also be used in other areas of anomaly detection, such as fraud detection. Because training classifiers takes a long time, the search strategy should be enhanced further. as the amount of graph data, including biological and social networks, grows. The existing SSE cannot fully satisfy graph data queries. Future study could concentrate on more complex search directions such as matrix queries and graph adjacency inquiries, among others. Because of the blockchain's unique properties, research into combining it with searchable encryption is also a possibility.

VII. CONCLUSION

In this research, we proposed an efficient and privacypreserving technique. Finally, the suggested system implements an efficient SSE method based on attackcategory with DES, allowing for safe verification, dynamic updating, and multi-user requests. The proposed method also enables for rapid updating. Create the authenticator by encrypting and signing the DES's root and timestamp after the token has been produced. The authenticator enables users to validate the accuracy of the server's results. Finally, the strategy will be put into action, and thorough trials will be carried out to evaluate it. The results are appropriate and consistent with our performance analysis. 8

REFERENCES

[1]. W. Lou, Y. T. Hou, W. Sun, X. Liu, and H. Li, "Catch you if you lie to me: Efficient verifiable conjunctive keyword search over large dynamic encrypted cloud data," in Proc. IEEE Conf. Comput. Commun., 2015, pp. 2110–2118.

[2]. C. Wang, X. Yuan, Q. Wang, J. Zhu, Q. Li, and K. Ren, "Enabling generic, verifiable, and secure data search in cloud services," IEEE Trans. Parallel Distrib. Syst., vol. 29, no. 8, pp. 1721–1735, Aug. 2018.

[3]. J. Yang, L. Xiong, J. Liu, and J. Pei, "Secure and efficient skyline queries on encrypted data," IEEE Trans. Knowl. Data Eng., vol. 31, no. 7, pp. 1397–1411, Jul. 2019.

[4].Jiadong Ren, Jiawei Guo, Wang Qian , Huang Yuan ,Xiaobing Hao , and Hu Jingjing," Building an Effective Intrusion Detection System by Using Hybrid Data Optimization", Hindawi Security and Communication Networks Volume 2019

[5]. Zhenkui Shi , Xuemei Fu , Xianxian Li , and Kai Zhu," Enabling Efficient, Secure, Verifiable Searchable Symmetric Encryption," in Proc. IEEE Trans.Knowl. Data Eng, Vol. 34, No. 7, July 2022.

[6] Chai and G. Gong, "Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers," 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 2012, pp. 917-922, doi: 10.1109/ICC.2012.6364125.

[7] X. Liu, R. H. Deng, K. -K. R. Choo and J. Weng, "An Efficient Privacy-Preserving Outsourced Calculation Toolkit With Multiple Keys," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 11, pp. 2401-2414, Nov. 2016, doi: 10.1109/TIFS.2016.2573770.

[8] Kui Ren Bo Zhang Ruitao Xie Kan Yang, Xiaohua Jia. Effective data access control for multi-authority cloud storage systems. IEEE TIFS, 8:1790–1799, 2013.

[9] Lokesh M. Gupta Karl A. Nielsen Matthew G. Borlick, Lokesh M. Gupta. Method, system, and computed program product for distributed storage of data in a heterogeneous cloud. INTERNATIONAL BUSINESS MACHINES CORPORATION, 10:171, 2019.

[10] Amrit Jassal Daniel H. Jung Gregory B. Neustaetter Sean H. Puttergill etc Hakan Ancin, Xi Chen. Systems and methods for facilitating access to private files using a cloud storage system. In Inc. Mountain View, CA, page 585, 2019.

[11] Kristin Lauter Seny Kamara. Cryptographic cloud storage. In LNCS, pages 136–149, 2010.