International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Enhanced Security Framework for Storage using PSO in Cloud Computing

Dinesh Parkash*[1], Dr. Sumit Mittal[2]

Submitted: 24/09/2023 **Revised**: 17/10/2023 **Accepted**: 11/11/2023

Abstract: The expansion in web usage demands security issues with it. Poor programming can influence the activity of the frameworks and information secrecy because of the security holes in the frameworks. Intrusion Detection System (IDS) has been developed to identify and report assaults. To encourage IDS frameworks, Artificial Intelligence based approaches have been utilized. The development in the field of cloud computing has provided a multifunctional view for the clients such as normalized applications to clients on the web that can be maintained on a routine basis. Cloud Computing is utilized for information stockpiling, so that information security and protection issues such as Confidentiality, Availability, and Integrity should be met. The universality of cloud computing allows individuals to store their information on the web. In our research idea, first of all we have designed an IDS (Intrusion Detection System) for Cloud Security using Artificial Bee Colony Algorithm (ABC) and then we have designed an enhanced framework for cloud storage with a secure environment using Particle Swarm Optimization Algorithm (PSO). Firstly, we considered three performance measures like recall, accuracy, and precision, and then to improve the efficiency of the proposed model, we have added two new metrics like F-measure and confusion matrix. After that we have compared this proposed model with the existing models. In this research proposal, the developed model trained with the NSL-KDD dataset is being presented.

Keywords: *Naïve Bayes (NB), Particle Swarm Optimization (PSO), Artificial Bee Colony Algorithm (ABC), KDD-99, NSL-KDD dataset, Cuckoo Search Algorithm (CSA).*

1. Introduction

Cloud Computing is web based registering where virtual shared servers give programming, framework, stage, gadgets and different assets. Cloud Computing being a Web based framework conveys information development and registering assets, for example, working frameworks, stockpiling administrations, network foundation, equipment gear, and, surprisingly, whole programming applications to clients in a minimal expense way. Rather than keeping information on your own hard drive or refreshing applications for your requirements, you utilize a help over the Web, at another area, to store your data or utilize its applications. Cloud computing offers significant types of assistance like Software as a service (SaaS), Expert as a service (EaaS), Platform as a service (PaaS) and Infrastructure as a service (IaaS) as shown in figure 1. [6]

Cloud security is a grouping of methodology and innovation intended to address outer and inward dangers to business security. Associations need cloud security as they advance toward their computerized change methodology and consolidate cloud-based apparatuses and administrations as a component of their framework. Cloud storage is a model of information stockpiling in which the computerized information is put away in sensible pools, the actual stockpiling traverses various servers, and the actual climate is ordinarily possessed and overseen by a facilitating organization.

Expanding web use has likewise brought numerous security holes. Numerous innovations, for example, firewall, information encryption, client validation are utilized to prevent these security holes. These security instruments prevent many sorts of assaults. The rising worldview of cloud computing emerges security gambles that unfavorably influence its unique partners. CSA based IDS trained with a dataset (KDD 99) led to a secure framework which ensures protection against attacks. [1]

1,2 M.M. Institute of Computer Technology and Business Management, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana 133203, India

ORCID ID : 0000-0001-5422-13841

** Corresponding Author Email: dineshgcccs@gmail.com1*

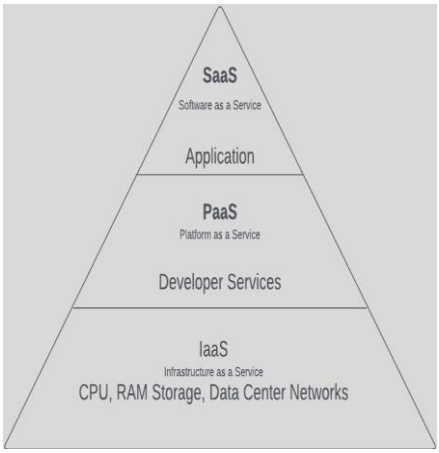

Fig.1. Cloud Computing Services

2. Literature Review

De los Reyes et al [2] developed a novel security design for the portable venture which utilizes cloud computing network-based security and computing to both improve and upgrade the security of accomplishments, and restore the right now vanishing security edge. The proposed new security engineering offers added adaptability for the present requirements. Parkash D. & Mittal S. [6] presented a survey of the accessible Cloud Security models. This approach also introduced the performance of Naïve Baye's and SVM (Support Vector Machine) using the KDD99 dataset. This approach examined the tabular presentation of the security boundaries like confidentiality, privacy and integrity.

Markandey et al [3] discussed how to accomplish cloud storage information security. The methodologies are contrasted against the outcome possibilities of existing information. Hendre et al [4] discussed a comprehensive review to audit the possible risks faced by cloud clients and chosen the consistency models. In light of this review, they have fostered a metaphysics portraying the cloud security controls, risks and compliances.

Abuhussein et al [5] developed a stakeholder-oriented taxonomical methodology that decides the security and confidentiality issues for different Cloud Computing models according to a partner's point of view. It suggests a thorough rundown of safety and security credits that are connected with these issues. The objective is to give partners the protection issues related with their cooperation with the cloud. Parkash D. & Mittal S.,[7] proposed a protected system for a Cloud computing climate. We have additionally achieved improvements utilizing the CSA. The boundaries like precision, recall and accuracy have been considered as CSA improvements. In contrast to the preceding NB and SVM IDS systems, the proposed IDS have an upper edge in the terms of precision, accuracy and recall parameters. This IDS based on CSA accomplished a maximum of the accuracy of 87.5% precision of 0.8 and recall of 1.0 values. Hajimirzaei B. et al [9] proposed new IDS with three steps (a) Training (b) Validation (c) Testing. Identification of attacks are done with MLP (Multilayer Perceptron) Network. The Simulation process has been carried out using cloud-sim simulator and NSL-KDD Datasets. The evaluation methods such as MAE (Mean outright mistake), RMSE (Root Mean Square Blunder) and kappa statistics have been raised with existing models and performed better. Salem R. et al [10] have made another calculation performed as a combination of ABC and Multi Objective Enhancement. Thus, access and enhanced copy arrangement are introduced at the suitable best spot for the base distance and smallest expense way and, for direct honey bees, for briefest courses in distance and lower cost. Parkash D. & Mittal S.,[8] proposed a methodology towards IDS and made a new fitness function for the Artificial Bee Colony Algorithm. The implemented IDS has been developed with the KDD Cup 99 Dataset. In this experimental work, we have organized a safeguarded system for the Cloud Comp. climate and comparatively accomplished an update utilizing ABC Calculation. The cutoff points like recall, accuracy and precision have been engaged with the thought for the ABC improvement methodology. From the foundation drive for the planned IDS close by NB and SVM. It has been observed that the implemented IDS found to be superior in comparison to SVM and NB.

Nil C. et al [11] identified not only the best AI model, but also surveyed the fundamental datasets, freely accessible, used to prepare and test security arrangements that utilize current grouping calculations for inconsistency discovery. Consequently, DARPA 1998 and KDD were considered as primary drives, while NSL-KDD, ISCX IDS 2012 and CICIDS 2017 will be examined in future. Thomas C. et al [12] introduced attempts to give supporting realities to the utilization of the DARPA IDS assessment dataset. The signature-based IDSs namely Grunt and Cisco IDS, along with two inconsistency indicators, have been utilized for this IDS assessment using DARPA dataset. Xiong Fu et al [13] proposed a model which stored the data and index in the same block known as Innode which was stored in the different files in the traditional file system. To handle the recovery and failure a round chain structure was introduced. Two Innode segments such as IOF (Index of files) and COF (Cache of files) have been considered for storing the small files. In this approach if block size is less than or equal to 64 MB then they have been considered small files of size 2 MB else 10 MB. In this approach a block replica placement algorithm has been proposed and performed better than Sequence file and Hadoop archives (HAR).

Kumar A. et al [14] presented a methodology that permits clients to access and store information from cloud storage safely with assurance that nobody can access the information. Clients additionally need not stress over cloud

suppliers accessing their information illicitly. The authors have exploited the method of elliptic bend cryptography encryption to safeguard information documents in the cloud. Two pieces of the cloud server has been moved along the exhibition during capacity and getting information. The ECC Encryption calculation has been utilized for encryption as another benefit to work on the exhibition during the coding and decoding process. Ingre B. and Yadav A.[15] proposed an IDS based on ANN and used NSL-KDD dataset for binary classification with other attack detection classification false positive rate, accuracy and detection rate parameters has been used for result evaluation. The results have been compared with other models and better performance for the developed IDS has been observed. The developed IDS has performed better for binary classification as well as for finding the class of attacks. Roempluk, T., & Surinta, O.[17] proposed an IDS used to identify the DDoS attacks from the KDD99 and NSL-KDD Datasets using the three machine learning algorithms (a) KKN (K-nearest neighbor) (b) SVM (c) MLP (Multilayer perceptron). For the identification process the datasets are divided into several classes like 2,6 and 7 classes from the experimental work. KNN algorithm has been performed better from the other MLP and SVM algorithms.

Proposed Method

An Efficient IDS that uses an ABC algorithm for Security and Grouped PSO algorithm for Storage purpose has been designed. First of all, we have labeled all the features of the NSL KDD dataset with numbers. The identification process of attack and normal data in the network traffic is done. A graph is generated to show the number of normal and attack values. In this proposed model, we took five parameters for calculating the performance and evaluated the parameters of the proposed IDS using NB (Naïve Bayes) and SVM (Support Vector Machine). The implemented IDS using SVM has achieved superior performance as compared to the implemented IDS using NB. For the less memory consumption in the proposed IDS, we have applied the G-PSO algorithm.

In the proposed IDS, we have considered 10 features, the characteristic number 5,6,23,24,30,33,34,35,36 and 37 using the NSL KDD Dataset. Using these feature data, we have suggested our new function of fitness for the desirable solution using ABC algorithm.

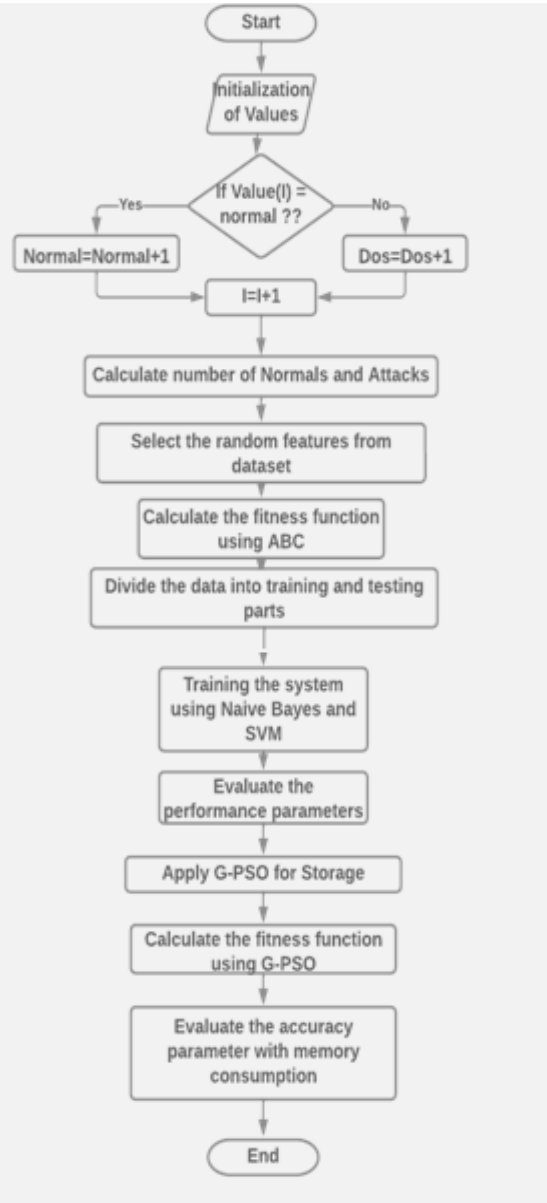

Fig.2. Flow chart of the proposed IDS

3. Experimental Results

3.1. Dataset

Defense Advanced Research Projects Agency (DARPA) was utilized to prepare and test the presentation of IDS. In DARPA, a total of 5 million association records of 100 bytes with 7 weeks of organization traffic are utilized. Knowledge Discovery and Data mining (KDD-99) Dataset is formed by extracting the DARPA Dataset. It is a dataset that contains two types of connections, bad connections are called attacks or intrusions and with good connections are called normal. It consists of 41 features and one more feature for class.

The NSL-KDD is a changed variation of KDD Cup 99 dataset. The dataset has 41 properties, divided into 5 classes, which are ordinary and 4 assault gatherings, and they are depicted later. The 42nd characteristic is the class property, which contains data about these gatherings; this

quality has positive or negative occurrences. The Attacks are further classified into four classes (1) (DoS) Denial of Service (2) (PRB) Prob (3) (R2L) Remote to Local (4) (U2R) User to Root.[16][17]

It excludes duplicate records in the training dataset.

In NSL KDD datasets the duplicacy of records is not present.

The values can be easily obtained from the dataset as the quantity of records is less

The quantity of chosen records from every trouble level gathering is contrarily relative to the level of records in the first KDD dataset. Thus, the characterization paces of different AI strategies shift in a more extensive territory, which makes it more effective to have a precise assessment of various learning methods.[17]

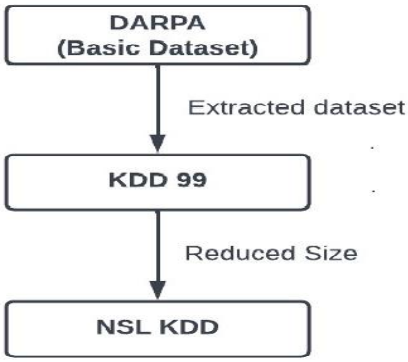

Fig.3. Relationship between Datasets

DARPA 1999 Dataset		
S.No.	**Attack Type**	
1.	DoS	63
2.	Probes	37
3.	R2L	53
4.	U2R	37
	190 instances of 57 Attacks	

Table1. Description of DARPA Dataset

S. No.	Properties	KDD 99	NSL-KDD
1.	File Size	92 MB	15 MB
2.	Usability	5.63	6.25
3.	Data Explorer	35.69 MB	55.87 MB
4.	File Summary	8	10

Table2. Properties of KDD 99 and NSL-KDD Dataset

3.2. Evaluation Metrics

To build the exhibition of the proposed IDS model; Accuracy, Precision, Recall ought to be determined. We have picked up Accuracy, Precision, Recall, F1 Measure and Confusion matrices for assessment.

- **Accuracy:** Exactness addresses the quantity of accurately grouped information occurrences over the absolute number of information cases.

Accuracy= (Total Positive + Total Negative)/ Total Instances

- **Precision:** Out of all the positives anticipated which rate is really certain?

Precision= True Positive / (False Positive + True Positive)

- **Recall:** Out of complete positives, which rates are anticipated?

Recall= True Positive / (False Positive + True Negative)

- **F- measure:** Recall and precision's Harmonic mean

F- measure= 2*(recall*precision)/(recall + precision)

- **Confusion Matrix:** It is a framework of size 2×2 for paired characterization with genuine qualities on one hub and anticipated on another.

3.3. Evaluation Criteria and Results

In the evaluation phase, first of all our proposed IDS identifies the normal and attack data of the KDD dataset and calculate them in numbers. The values of normal and attacks after input analyzed by the proposed IDS model using NSL KDD dataset as shown in Table3 and figure 4. All the five parameters are determined by the proposed IDS among NB and SVM represented in the figure 5, figure 6 respectively as well as in Table4 and Table5. After evaluating the parameters we compared both the methods

and represented in the Table6 and Fig.7 also. To reduce the memory consumption of the proposed IDS, we have applied Grouped PSO. After applying G-PSO we have accuracy and memory consumption as shown in Table7 and Fig.8 using proposed IDS model with NB and SVM. Memory consumption before and after with accuracy is represented in Table8 and Fig.9 using the proposed IDS. Reduction in memory consumption is also shown in Fig.9.

Analysis of Normal & Attack of NSL KDD Dataset	Values
Normal	264
Attack	236

Table3. Analysis of Normal & Attack

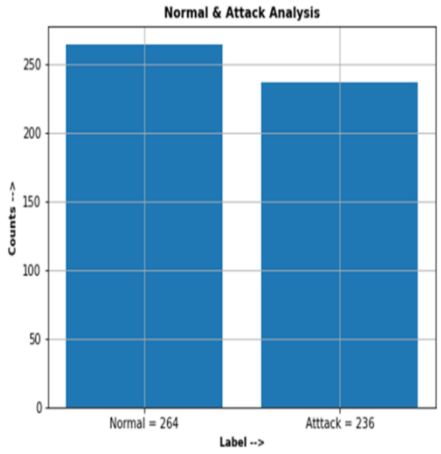

Fig.4. Analysis of Normal & Attack of NSL KDD Dataset

Evaluation of Proposed IDS (NB)	Values
Accuracy of NB	86.0 %
Precision of NB	90.0 %
Recall of SVM	79.0 %
F-measure of NB	84.0 %
Confusion Matrix of NB	[[73 6] [15 56]]

Table4. Evaluation of Proposed IDS (NB)

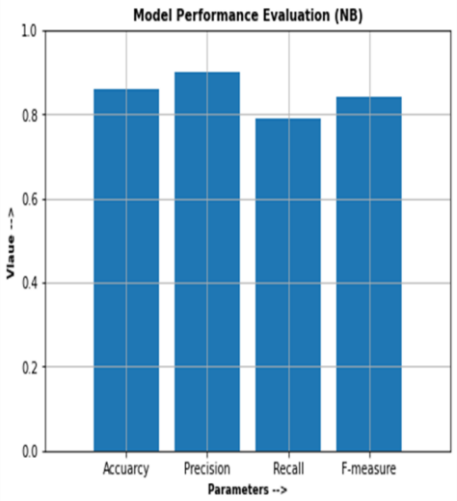

Fig.5. Evaluation of Proposed IDS (NB)

Evaluation of Proposed IDS (SVM)	Values
Accuracy of SVM	96.0 %
Precision of SVM	95.0 %
Recall of SVM	97.0 %
F-measure of SVM	96.0 %
Confusion Matrix of SVM	[[75 4] [2 69]]

Table5. Evaluation of Proposed IDS (SVM)

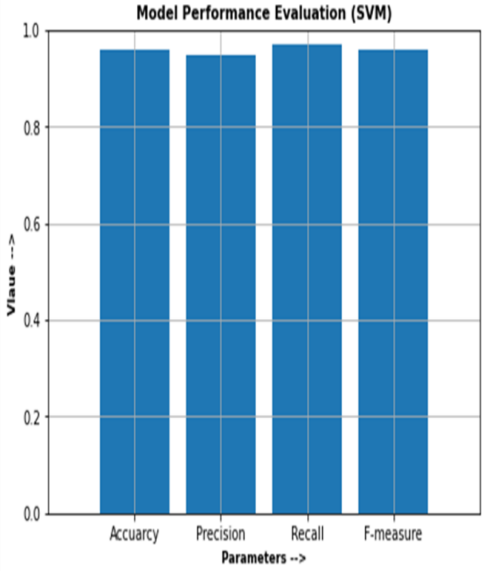

Fig.6. Evaluation of Proposed IDS (SVM)

Comparison Of Proposed IDS's among NB & SVM	Values
Accuracy %age ABC & NB	86.0 %
Accuracy %age ABC & SVM	96.0 %

Table6. Comparison of Proposed IDS's among NB & SVM

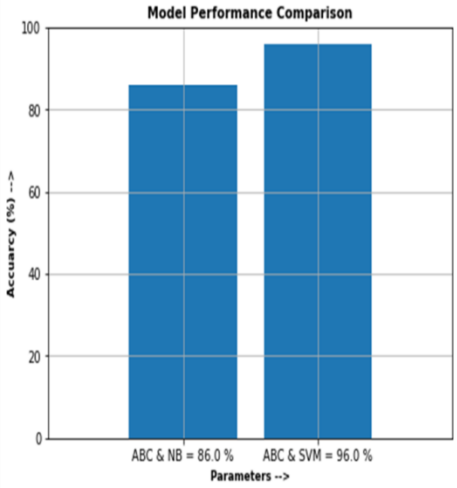

Fig.7. Comparison of Proposed IDS's among NB & SVM

Comparison Of Proposed ABC(NB) and ABC(SVM) among Storage	Values
Accuracy %age ABC & NB	88.0 %
Accuracy %age ABC & SVM	96.0 %
Memory Consumption (Before)	545.94 kb
Memory Consumption (After)	116.44 kb

Table7. Comparison of Proposed ABC (NB) and ABC (SVM) among Storage

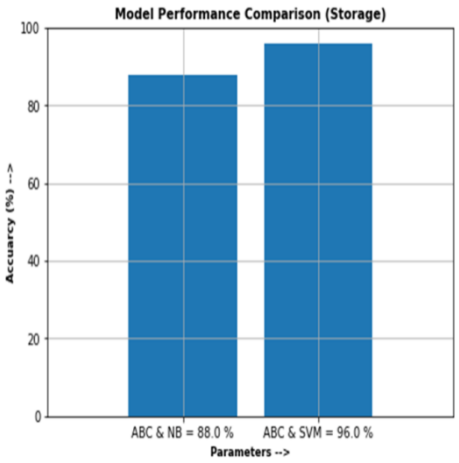

Fig.8. Comparison of Storage based Proposed ABC (NB) and ABC(SVM)

4. Conclusion

In this research proposal, a IDS based on a combination of an artificial bee colony algorithm (ABC) and Grouped Particle Swarm Optimization Algorithm (G-PSO) has been developed. Both abnormal and normal network traffic packets are identified by the proposed IDS as shown in Table3 and Fig.4. By using ABC algorithm, it is found that the planned IDS using SVM achieved finer than proposed IDS using NB as shown in Fig.7. By using G-PSO, it is found that memory consumption used in the proposed IDS is less. In this research work, we have conducted 10 trials with different amounts of data as shown in Table8. It is observed that the planned IDS achieved superior ABC using SVM as shown in Fig.8. To reduce the memory space G-PSO has been applied. It is observed that memory consumption is reduced as compared to previous as shown in Fig9. In this research work, we will also apply a Cuckoo Search Algorithm like optimization algorithm for the future.

				Using ABC Only Accuracy		Using ABC & PSO Accuracy		Memory Consumption (KB)	Memory Consumption (KB)
Sr. No.	**Data**	**Normal**	**Attack**	**NB(ABC)**	**SVM(ABC)**	**NB(Storage)**	**SVM(Storage)**	**Before M.C. (kb)**	**After M.C. (kb)**
1	500	264	236	86	96	88	96	545.94	116.44
2	1000	515	485	74	85	92	88	571.97	115.32
3	1500	750	750	72	82	96	91	424.28	83.59
4	2000	1038	962	70	83	86	91	457.97	90.31
5	2500	1307	1193	86	88	92	89	536.15	106.11
6	3000	1571	1429	80	89	90	94	609.55	120.20
7	3500	1839	1661	78	80	91	78	599.55	115.20
8	4000	2110	1890	74	79	90	77	623.45	223.16
9	4500	2358	2142	76	77	87	76	749.37	227.98
10	5000	2647	2353	70	73	83	77	783.85	339.89

Table8. Accuracy and Memory Consumption

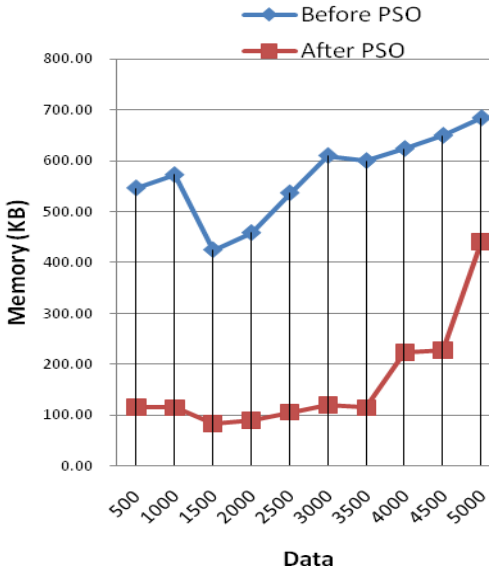

Fig.9. Memory Consumption before and after applying G-PSO on the Proposed IDS.

References

[1] Abuhussein, A., Bedi, H., & Shiva, S. (2013). Towards a Stakeholder-Oriented Taxonomical Approach for Secure Cloud Computing. 2013 IEEE Sixth International Conference on Cloud Computing. doi:10.1109/cloud.2013.132.

[2] De los Reyes, G., Macwan, S., Chawla, D., & Serban, C. (2012). Securing the mobile enterprise with network-based security and cloud computing. 2012 35th IEEE Sarnoff Symposium. doi:10.1109/sarnof.2012.6222759.

[3] Markandey, A., Dhamdhere, P., &Gajmal, Y. (2018). Data Access Security in Cloud Computing: A Review. 2018 International Conference on Computing, Power and Communication Technologies (GUCON). doi:10.1109/gucon.2018.8675033

[4] Hendre, A., & Joshi, K. P. (2015). A Semantic Approach to Cloud Security and Compliance. 2015 IEEE 8th International Conference on Cloud Computing. doi:10.1109/cloud.2015.157

[5] Abuhussein, A., Bedi, H., & Shiva, S. (2013). Towards a Stakeholder-Oriented Taxonomical Approach for Secure Cloud Computing. 2013 IEEE Sixth International Conference on Cloud Computing. doi:10.1109/cloud.2013.132.

[6] Parkash D. ,Mittal S.(2020) Comparative study and performance analysis of various data security and cloud storage models. J Solid State Technology, 63(2s), 6318–6331.

[7] Parkash D.,Mittal S.(2022) An Enhanced Secure Framework Using CSA for Cloud Computing Environments.International Conference on Innovative Computing and Communications, Proceedings of ICICC 2022,Volume 2(349-356). Lecture Notes in Networks and Systems (471) by Springer(https://link.springer.com/chapter/10.1007/978-981-19-2535-1_27).

[8] Parkash D.,Mittal S.(2022) An Efficient Security Framework Using ABC in Cloud Computing.2nd International Conference on Research Trends in Engineering and Management, Proceedings of ICRTEM- 2022,(60-64).

[9] Hajimirzaei, B., & Navimipour, N. J. (2018). Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm. ICT Express. doi:10.1016/j.icte.2018.01.014.

[10] Salem, R., Salam, M. A., Abdelkader, H., awad, A., & Arafa, A. (2019). An Artificial Bee Colony Algorithm for Data Replication Optimization in Cloud Environments. IEEE Access, 1–1. doi:10.1109/access.2019.2957436.

[11] Nil C., Patriciu V. V. and Bica I., "MACHINE LEARNING_DATASETS FOR CYBER SECURIT_APPLICATIONS" Vol. 3 (2019), Issue 3, pg(s) 109-112

[12] Thomas C., V. Sharma, N. Balakrishnan, "Usefulness of DARPA dataset for intrusion detection system evaluation" in Proceedings of SPIE - The International Society for Optical Engineering, March 2008.

[13] Fu, X., Liu, W., Cang, Y., Gong, X., & Deng, S. (2016). Optimized Data Replication for Small Files in Cloud Storage Systems. Mathematical Problems in Engineering, 2016, 1–8. doi:10.1155/2016/4837894.

[14] Kumar, A., Lee, B. G., Lee, H., & Kumari, A. (2012). Secure storage and access of data in cloud computing. 2012 International Conference on ICT Convergence (ICTC). doi:10.1109/ictc.2012.6386854.

[15] Ingre, B., & Yadav, A. (2015). Performance analysis of NSL-KDD dataset using ANN. 2015 International Conference on Signal Processing and Communication Engineering Systems. doi:10.1109/spaces.2015.7058223.

[16] Tavallaee, M., Stakhanova, N., & Ghorbani, A. A. (2010). Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(5), 516–524. doi:10.1109/tsmcc.2010.2048428.

[17] Roempluk, T., & Surinta, O. (2019). A Machine Learning Approach for Detecting Distributed Denial of Service Attacks. 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON). doi:10.1109/ecti-ncon.2019.8692243.