Hybridization of Adaptive Cuckoo's Search Algorithm with Core Vector Machine for Feature Selection

Srinivasa Rao Pokuri[1], Dr. Nagaraju Devarakonda[2]
[1]Research scholar, Department of CSE,
Acharya Nagarjuna University, Nagarjuna Nagar, AP, India ,
[2]Associate Professor, School of Computer science and engineering,
VIT-AP University, Amaravati, AP, India,
[1]srinivas.pokuri2@gmail.com,[2]dnagaraj_dnr@yahoo.co.in

Article Info
Page Number: 4740 - 4748
Publication Issue:
Vol 71 No. 4 (2022)

Article History
***Article Received**: 25 March 2022*
***Revised**: 30 April 2022*
***Accepted**: 15 June 2022*
***Publication**: 19 August 2022*

Abstract

As an important catalyst in a fast digitizing world, Cloud computing offers great opportunities in creating scalability in resource sharing to perform transparent computation, while allowing seamless transfer of information as well. Researchers face an uphill task to ensure that data, information sources, software and cloud materials remain safe and secure. Cloud security measures integrated into the cloud computing ecosystem help in securing the cloud resources.In such a background, Anomaly Detection Systems (ADS) offer us the best possibilities in building detection control mechanisms. However, when network traffic data is efficiently managed with the application of a machine learning algorithm, it results in building the accuracy capability of the ADS. In this paper, AdaptiveCuckoo's Search Algorithm is hybridized within a CoreVector Machine to ensure better feature selection results.

Keywords:*Anomaly Detection Systems (ADS), Network Security, Feature Selection, Cuckoo Search, Main Vector Machine*

1. Introduction

A software application known as anomaly detection system (ADS) monitors a computer network to detect suspicious activity or protocol breaches and generates notifications for a control station. Some new developments revealed that networking and computer devices are regularly affected by many bugs, sometimes measured in the range of 20-40 every month on average. Such vulnerable and unstable computing / network ecosystem is aggravated or augmented by various device flaws. As a result of this vulnerable climate, the scope for intrusion detection and prevention is constantly expanding. Anomaly detection devices, the cybersecurity version of a burglary warning, help the distressed firewall. Researchers around the world have acknowledged detection systems abnormalities as providing a secure safety net while monitoring the network traffic, thereby demonstrating its dexterity in spotting suspicious behaviour easily. Theree are 3 broad areas that ADS is quite effective in, such as;

- To monitor and analyze Traffic
- To identify suspicious activities.
- To assess magnitude of the intrusion and raise warning signal

In Figure 1, the anomaly detection system and its basic architecture is displayed.

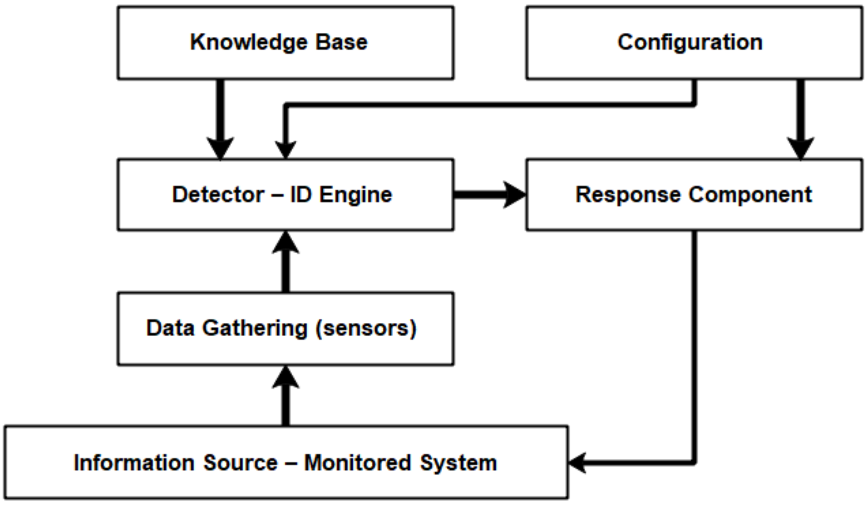

Fig. 1. Architecture of anomaly detection systems

The architectural framework is composed of the following parts:

Data collection device: Here data collection from the monitored system takes place.

Detector - ID Engine: In the event of an attack, it processes the data gathered from the sensors to detect disruptive activity and sends a warning signal to the response component.

Database: It contains expertinformation networks wherein pre-processing and collection is done by sensors

Configuration device: The current status of ADS is provided here.

Response component: A response (active or inactive) is triggered here at the sight of an intrusion.

1.1. Intrusion detection methods

An intrusion detection system can either be host-based or network-based based on the nature of targeted data.

- **Host Based Anomaly Detection Systems (HADS)**: Such a system comprises host delegate who examines device calls to detect intrusions, besides monitoring activities taking place in programme logs, file system updates, and other host behavior.
- **Network Based Anomaly Detection Systems (NADS):**This stand alone platform detects intrusions by carrying out inspection of the network and its activities, thereby tracking movement of different hosts.The network traffic is accessed by the NADS when a network hub and a network switch are connected to it. A network switch is given optimization touches for port mirroring, or highlighting network stations.

This article primarily focuses on troubleshooting intrusion detection issues wherein administrator is supported to achieve superior data preprocessing and classification performance. This article focuses on examining how hybrid swarm intelligence algorithms i.e. Adaptive Cuckoo Search Algorithm (ACSA) combines with Core Vector Machine (CVM) in classifying detection data set anomalies by using feature selection methods.

2. Related work

Al-Jarrah et al. planned ADS with a data preprocessing and machine learning approach to increase the performance of ADS, [3]. A feature assortment supported a filter was counseled by Ambusaidi et al. to extend the performance of intrusion detection [4]. El-Khatib applied the knowledge gain magnitude relation (IG) metric and so the k-means classifier to pick out the foremost necessary characteristics of the dataset and boost ADS accuracy [5]. Also, Mishra et al. did some good work on intrusion detection in an exceedingly} very cloud setting [6]. Viejas et al. projected a feature selection technique to increase ADS performance, whereas operative the on-board system on reduced power [7]. Mistry planned a feature selection approach that mixes genetic formula (GA) with particle swarm optimization (PSO) rule to accurately observe facial expression or cues [8]., Lagrange projected a feature selection technique to create distinct classification of remote sensing footage [9]. Gülsen and Taskn proposed a system for classifying hyperspectral pictures victimization feature selection [10]. Wang et al. counseled a feature selection approach with PSO to eliminate obsolete and redundant features from high-dimensional space, thereby enhancing preciseness at intervals the functioning of the classification algorithm [11]. it' getting to be surmised from the offered literature that intrusion detection techniques have many proved applications in laptop computer and network settings to spice up its security and stability, on top of all. Furthermore, anomaly-based ADS is important to maintenance of the protection of cloud-based resources. Here, a machine learning formula has been applied to develop ADS supported anomaly

By victimization info preprocessing techniques further as feature collection, ADS preciseness levels are going to be increased. Hence, this paper introduces a hybrid swarm intelligence totally based intrusion detection model that mixes the adaptative Cuckoo Search formula (ACSA) and Core Vector Machine (CVM). The counseled methodology permits the exactitude of this intrusion detection model to extend further.

3. Proposed method

3.1. Anomaly Detection Systems (ADS) using ACSA

Figure 3 shows a schematic diagram depicting the trajectory in which the intrusion detection model is developed. A data set is created by collecting data related to network flow. The dataset's obsolete and redundant functionality is then disabled with ACSA. The anomaly detection model is developed after the Core Vector Machine is assigned the related functionalities. This anomaly detection model detects anomaly or attacker packets, and a warning signal like an attack or a regular one is sent depending on the packet appearing on the network. In response to the ADS's warning message, the intrusion protection mechanism then takes protective action for safeguarding the system.

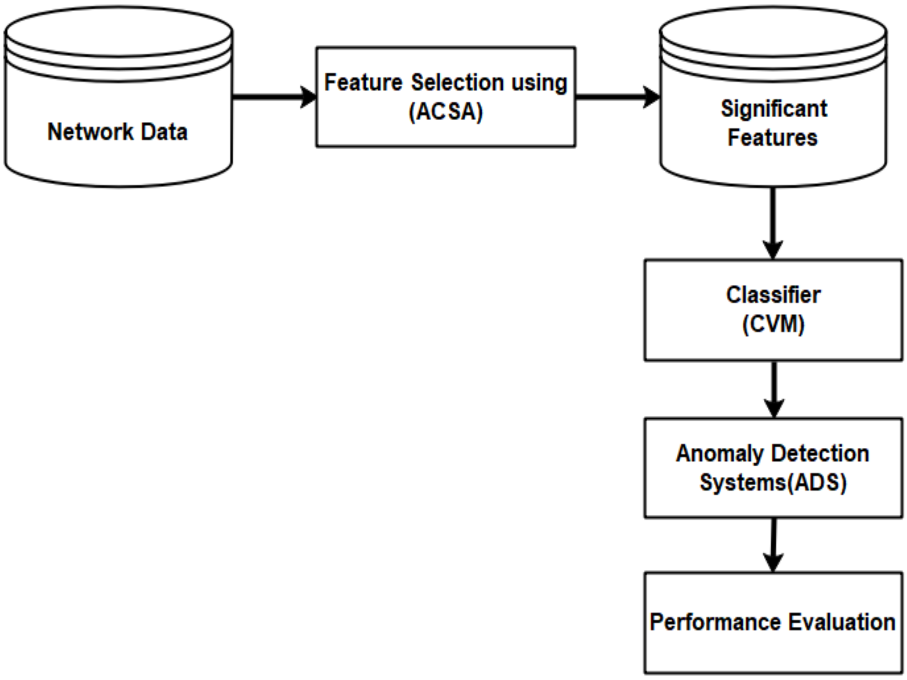

Fig. 2. Schematic Diagram of Anomaly Detection System (ADS) Development

3.2. Adaptive Cuckoo Search Algorithm (ACSA)

Cuckoo's regular search algorithm does not regulate the size of a step during iteration process to achieve total minima or maxima. Therefore, we do not attempt any step like integrating a phase size equaling the physical dimensions of each human nest either in the search room or the current crop.

Algorithm: *Adaptive Cuckoo Search Algorithm (ACSA)*

```
1.  Randomly initialize thenumber of guest nests $A_i = (a_i^1, ..., a_i^d, ..., a_i^n)$ for $i = 1, 2, ..N$ for an
    $n-$dimensional problem and define the fitness function $fit(A)$. Initially, take $t = 1$ and
    evaluate the fitness function of the guest nests $fit(A_i)$ for $i = 1, 2, .., N$; for the first time.
2. Iterative algorithm
        A. Find the best and the worst of the current generation among guest nests.
        B. Calculate the step size using:
        $Step_i(t+1) = \left(\frac{1}{t}\right)^{\left|\frac{bestfit(t)-fit_i(t)}{bestfit(t)-worstfit(t)}\right|}$
        C. Then calculate the new position of the cuckoo nests using:
        $A_i(t+1) = A_i(t) + rand_n \times Step_i(t+1)$
        D. Evaluate the objective function of host nests $fit(A_i)$ for $i = 1, 2, .., N$
        E. Then randomly choose a nest, $j$, among $N$
        If $(fit_i > fit_j)$
         Update $j^{th}$ Nest by the new solution.
         Finish
        F. The worst nests are abandoned with a probability $(P_a)$ and a new one is built.
        G. $t = t + 1$.
        H. Check $(t \le t\max)$ or, if yes, go to A.
         otherwise end.
3. end
```

As a fairly uncomplicated protocol, CVM can effectively function without relying upon complex heuristics unlike decomposition methods, while faster convergence is possible in lesser iterations in a CSA.Varying training data are used to allow the CVM to perform training, thereby helping it locate the coreset found among the host of training data points. The data points for all classes are randomly selected in this model. The CVM classifier is designed to find the support vectors from the selected points. In CVM, the clustering system is required for selecting the data as part of the process. As a result, the cuckoo search algorithm is used to find the best cluster centroids. The cluster points are applied to the training set. Finally, the CVM is retrained after the revised training set is applied.

Algorithm: *CVM-ACSA hybrid approach*

```
Given an input training dataset.
Let acc be initialized to 0
To start
        Let acc be the precision rate during execution initially 0;
        While $acc < ACC$ do // ACC is the precision rate threshold
for k = 1.. . , n do       // n is the number of iterations
Perform training using the CVM classifier .
Perform clustering using the cuckoo search algorithm .
Finish
Build classifiers; Update acc
End during
Finish
```

In comparison, CVM and CSA produce two normal data profiles in this hybrid solution. As a result, in order to minimize the incidence of false negatives, the data object is only confirmed as usual if all classifiers interpret it as such. In case of consensus among all classifiers about irrelevance or inappropriateness of the data, the CSA classifier assigns a particular type to the anomaly/intrusion. In case the classifiers' ranking of the results are inconsistent with its findings, it is labeled as a new category of network traffic anomaly / interference, which eventually helps the ADS enhance its own performance.

4. Results

4.1. Reduced functionality

The list of diminished characteristics is shown in Tables.1 along with the number of characteristics generated by the proposed method.

Table 1. Reduced functionality using the proposed method

Datasets	Reduced functionality	# of features
DoS + 10% normal	1,3,4,7,11,13,17,23,25,27,32,32,34,35	14
Probe + 10% normal	1,3,4,5,7,8,11,13,23,24,25,27,29,30,32,34,35	17
R2l + 10% normal	1,2,3,4,6,7,9,12,13,14,18,22,23,25,26,32,34	17
U2R + 10% normal	1,2,3,4,6,7,13,14,18,22,23,25,26, 27,31,34	16

The dataset i.e. NIMS 2 obtained from a freely accessible dataset repository is used in this experiment. The first 100 instances in this dataset are derived from the 'GTALK' class, while another batch of 100 instances belong to the 'PRIMUS' class.The remaining 100 instances are derived from the 'ZFONE' class, while the data set is strengthened by the addition of all 21 instances belonging to the 'ONLINE BANKING' class. As a result, the final data set reaches a tally of 321 instances, 22 features, and 4 classes.

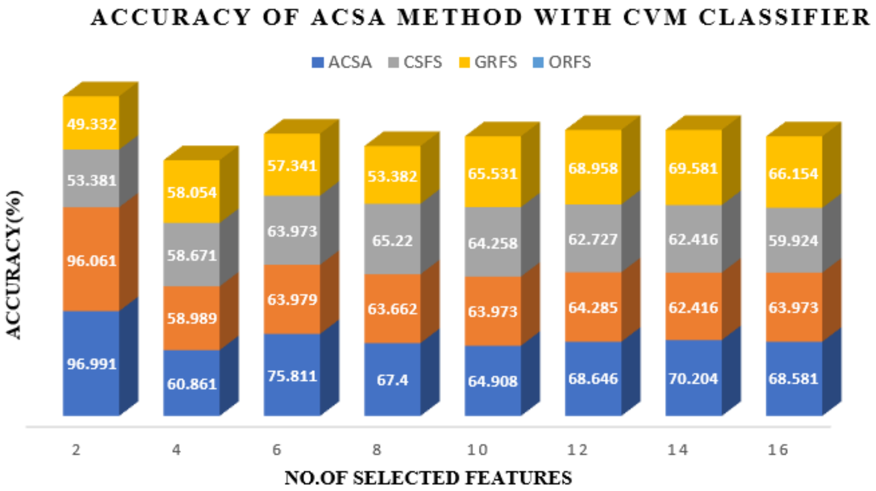

Fig. 3. Accuracy of CVM classifier with feature selection methods versus number of selected features

The currently available feature selection algorithms are used here, such as chi-square feature selection (CSFS), information gain enabled feature selection (GRFS), and single feature selection (ORFS). Measure performance in comparison mode. Suggested model. In addition, an anomaly detection model is created using the CFS classification algorithm and provides an appropriate assessment of the effectiveness of the feature selection method. Set the sum value k (the number of features to be selected), where k is the number of features to be selected. Then the number of k entities is selected through the ACSA selection process.

5. Conclusion

In this article, we have presented a CVM-ACSA model with the primary aim to improve the accuracy of ADS in network security. The suggested method is built upon a search technique based on adaptive cuckoo optimization and is able to select the most significant characteristics of the Core Vector Machine classification algorithm.When compared with other available methods for accuracy, our proposed method produces better results and is seen outperforming them as well.

References

[1] Krutz RL, Vines RD. Cloud computing security architecture. Cloud Security: A Complete Guide to Secure Cloud Computing. Indianapolis, IN: Wiley; 2010. pp. 179–80. To print.

[2] Tan Z, Nagar UT, He X, Nanda P, Liu RP, Wang S, Hu J. Improving the security of big data through collaborative intrusion detection. 2014 IEEE Cloud Computing Transactions; 1 (3): 27–33.

[3] Al-Jarrah OY, Alhussein O, Yoo PD, Muhaidat S, Taha K, Kim K. Data randomization and cluster-based partitioning for Botnet intrusion detection. IEEE Transactions on Cybernetics 2016; 46 (8): 1796–806.

[4]J. N. Rao and M. Ramesh, "A Review on Data Mining & Big Data Machine Learning Techniques", *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 7, no. 6S2, pp. 914-916, April 2019

[5] El-Khatib K. Impact of reduced functionality on the efficiency of wireless intrusion detection systems. IEEE Transactions on Parallel and Distributed Systems 2010; 21 (8): 1143–9.

[6] Mishra P, Pilli ES, Varadharajan V, Tupakula U. Intrusion detection techniques in a cloud environment: an investigation. Journal of Network and Computer Applications 2017; 77 (2): 18–47.

[7].J.N Rao, Dr.Rambabu Busi, Dr. G Rajendra Kumar, U. Surya Kameswari, " Content image Retrieval Based on using open Computer Vision and Deep Learning Techniques "International Journal of Advanced Science and Technology,Volume29Issue03Pages 5926 – 5939, 2020)

[8] Mistry K, Zhang L, Neoh SC, Lim CP, Fielding B. A micro-GA integrated PSO function selection approach for intelligent recognition of facial emotions. IEEE Cybernetics Transactions 2016; 47 (6): 1496–509.

[9] Lagrange A, Fauvel M, Grizonnet M. Selection of large-scale features with Gaussian mixing models for the classification of high-dimensional remote sensing images. IEEE Transactions on Computational Imagery 2017; 3 (2): 230–42.

[10] Kaya GT, Kaya H, Bruzzone L. Feature selection based on high dimensional model representation for hyperspectral images. 2017 IEEE Image Processing Transactions; 26 (6): 2918-28.

[11] Wang Y, Wang J, Liao H, Chen H. Selection of unsupervised function based on optimization of Markov coverage and particle swarm. Journal of Systems Engineering and Electronics 2017; 28 (1): 151–61.

[12] Ma L, Li M, Gao Y, Chen T, Ma X, Qu L. A new wrapper approach for feature selection in object-based image classification using polygon-based cross-validation. IEEE Geoscience and Remote Sensing Letters 2017; 14 (3): 409–13.

[13] Huda S, Yearwood J, Jelinek HF, Hassan MM, Fortino G, Buckland M. A selection of hybrid characteristics with ensemble classification for unbalanced health data: a case study for brain tumor diagnosis. IEEE 2016 access; 4: 9145–54.

[14] Nguyen TM, Wu QJ. Selection of online features based on fuzzy clustering and its applications. IEEE Transactions on Fuzzy Systems 2016; 24 (6): 1294-306.

[15] Abedinia O, Amjady N, Zareipour H. A new feature selection technique for load and price prediction of power supply systems. IEEE Power Systems Transactions 2017; 32 (1): 62–74.

[16] Yang Y, Xu HQ, Gao L, Yuan YB, McLaughlin K, Sezer S. Multidimensional intrusion detection system for SCADA networks based on the IEC 61850 standard. IEEE transactions on Power Delivery 2017; 32 (2): 1068–78.

[17] Marchang N, Datta R, Das SK. A new approach for an efficient use of the intrusion detection system in mobile ad hoc networks. 2017 IEEE Vehicle Technology Transactions; 66 (2): 1684–95.

[18] Ha T, Yoon S, Risdianto AC, Kim J, Lim H. Suspicious stream transfer for multiple intrusion detection systems on software-defined networks. IEEE 2016 Network; 30 (6): 22–7.

[19] Zhou C, Huang S, Xiong N, Yang SH, Li H, Qin Y, Li X. Design and analysis of anomaly intrusion detection systems based on several models in the automation of industrial processes. IEEE Transactions on Systems, Man and Cybernetics: Systems 2015; 45 (10): 1345–60.

[20] Lo CH, Ansari N. Consumer: a new hybrid intrusion detection system for distribution networks in a smart grid. IEEE Transactions on Emerging IT Topics 2013; 1 (1): 33–44.