Calibration Factors based intrusion detection (CFID) in Cloud Computing

Ponnuru. Sowjanya, K. V. N. Sunitha

ABSTRACT--- ***The distributed computing is the buzz in recent past, cloud computing stands first in this category. This is since, the users can adapt anything related to data storage, magnificent computing facilities on a system with less infrastructure from anywhere at any time. On other dimension such public and private cloud computing strategies would also attracts the foul players to perform intrusion practices. This is since, the comfortability that the cloud platform providing to end users intends them to adapt these services in regard to save or compute the sensitive data. The scope of vulnerability to breach the data or services over cloud computing is more frequent and easier, which is since, these services relies on internet protocol. In this regard, the research in intrusion detection defense mechanisms is having prominent scope. This manuscript, projecting a novel intrusion detection mechanism called "calibration factors-based intrusion detection (CFID)" for cloud computing networks. The experimental study portrayed the significant scope of the proposal CFID to detect the intrusion activities listed as remote-to-Local, Port Scanning, and Virtual-Machine-Trapping.***

Keywords: calibration factors-based intrusion detection (CFID), cloud computing (CC), cloud services (CS), IDC (International Data Corporation), ANN (artificial neural networks).

I.INTRODUCTION

The objective of cloud computing (CC) is to offer on demand, convenient access of network for sharing combined configurable resources of computing such as servers, applications, networks that can provisioned rapidly and released through minimum management service interactions of provider [1]. Services provided by the cloud are in several ways: platform as service (PaaS) [2], infrastructure as service (IaaS) [3], and software as service (SaaS) [4], open Nebula [5], Eucalyptus [6] and Microsoft's Azure [7].

Since cloud services (CS) are provisioned by internet: privacy & security of the CSs are the important problems to be faced. The work [8] presents that IDC (International Data Corporation) survey envisioned that security is the CC biggest challenge.

The work [9] presents that contemporary security of CC white paper through "Lockheed martin cyber security division" presents that concerning the substantial security after security of data would be the intrusion detection & safeguarding in the infrastructures of the cloud. The infrastructure of the cloud utilizes the virtualization methods, combined methods, which performs by benchmark protocols of internet. These might fascinate intruders because of more vulnerability included in it.

CC suffers from several outdated attacks like DoS, flooding, Distributed DoS, DNS poisoning. The work [10] presents that the attack of DoS underlying on cloud infrastructure of Amazon caused a Bitbucker.org a hosted site on the AWS for remaining inaccessible for small number of hours. The work [11] presents that cost of computing utilizing the contemporary cryptographic methods will not be overlooked aimed at cloud. For preventing the external attacks, the use of firewall is the best option, but is not suitable for internal attacks. The effective IDS (Intrusion detection system), "IPS (intrusion prevention system)" need to be included in the infrastructure of cloud for lessening these kinds of attacks.

1.1 Overview of Intrusion Detection Technologies for Cloud Computing

The contemporary Research in the domain of cloud computing has mainly aiming towards the security. This is since, the cloud computing is public accessible distributed network-based service-oriented platform.

Majority of research contributions in security of cloud computing are aiming to overcome the constraints of the intrusion detection and defense mechanisms for cloud computing.

The contributions models in recent past opted the method of defining set of rules [12] to observe the resource that is an active participant of the corresponding cloud computing network is prone to intrude or not. However, the method of rule-based IDS intricates to defend intruder in virtual and encrypted networks, as it might identify only outside attacks.

The other format of the intrusion detection in cloud computing are grid and cloud-based intrusion detection using deep learning methods [13].

These methods built on ANN (artificial neural networks) of both recurrent (relations between neurons are cyclic) and feed forward (no cyclic relations between neurons).

Revised Manuscript Received on August 30, 2019.

* Correspondence Author
Ponnuru. Sowjanya*, Research Scholar, JNTUH, Hyderabad, Telangana, India.
(Email: sowjanya.ponnuru@gmail.com)
K. V. N. Sunitha, BVRIT Hyderabad College of Engineering for Women, Hyderabad, Telangana, India.
(Email: k.v.n.sunitha@gmail.com)

However, the collective constraint of these methods is, demand of high volume of training data and process time, and false alarming is inversely proportionate to the both quantity and quality of the training corpus. The hidden Markov Method (HMM) based intrusion detection is the other contemporary contributions of the recent literature [14]. Since, the HMM performs based on the correlations discovered between the given parameters during the training phase, often these methods cluster the given training corpus in to multiple sets of training corpus such that each set contains the training records having highly correlation among them. These contributions also carry the constraints evinced in ANN based intrusion detection strategies.

The other dimension of the intrusion detection in cloud computing are built on traditional strategy that derives signatures based on the given training corpus [15], which further used in the classification process. However, these methods are not reliable to handle the zero-day attempts, and the intrusion attempts those slightly deviate from the signatures derived during learning phase.

The other format of the intrusion detection strategies built on using evolutionary computational techniques such as genetic algorithm, which derives the fitness of the parameter value produced through a request to cloud computing [16]. The crucial constraints of these methods are the probabilistic detection accuracy and computational complexity that often evince due to the inadequate fitness function.

In the year 2013, the work [17] suggested CC-oriented structure for constructing the disseminated intrusion identification called keeping the proves in several cloud platform levels, data transferring, where the probes are gathered towards engine security by proxy, & introducing incidents of security with the assistance of several engines of security. This model deliberated individual layers safety on the platform of cloud, yet it could not tell the way automatically the security engine introduced the events.

The work [18] presents that rapid compression algorithms & neural networks which are impulsive for analyzing the anomaly network traffic in the environment of CC, has kept the thought of intrusion detection based on network on the platform of cloud, yet it could not provide clear anomaly definition. The work [19] presents that several researchers are summed up in contemporary years on the CC platform of intrusion detection, and kept these contributions into 3 classes as per detection method called detection based on the tag, anomaly detection & hybrid method. They made brief defects examination in these models in coverage of data, coverage of attack and effectiveness of detection and noticed that these models are not complete. In the year 2015, Li Ming given how to identify aggressive conduct uneven CC areas, and the Wang Yichuan kept a model for identifying the attacks within CC environment, integrating game theory principle. In the year 2016, suggested the enhanced algorithm and method that might take detection of intrusion on the basis of outdated BP algorithm. The Xu Yang showed particular model for dealing with the detection of DDoS attack from the web-application perspective of service layer

Nevertheless, the contemporary research is simply regarding the implementation of detection models towards environment of cloud. Many of these models require training samples before and need not deliberate the performance of detection when faced through huge IDS data [20]. While there are voluminous data, it is intricate for conducting practical identification with the algorithms and it might have less accuracy of detection and rates of coverage. Hence, the IDS for CC need to possess self-learning capability, the ability of identifying anomaly intrusion, false negative, high speed, less false-positive for massive data.

II. RELATED WORKS

The work [21] presents that other effective and rapid secure IDS is proposed along with HIDS & NIDS. Here, in this system, there is IDS cloud which capture from the packets of network and examine them, later forward the reports towards administrator of cloud based on analysis of using hybrid classifier called KNN-NN. In respect to train and test, this data set called NSLKDD need to be utilized & once report is attained from IDS-cloud, the service provider of cloud needs to produce another novel alert aimed at this consumer. Further handle list of logs for assuring entire stored IP addresses which are malicious. This method is suggested for managing definite huge flow of data packets and also produces the reports on the basis of analysis.

The work [22] presents that other anomaly detection system is suggested in the hypervisor layer called hypervisor detector utilizing 1 or several novel hybrid algorithms which is mixture of FCM clustering algorithm besides with ANN, which is a cause for enhancing detection systems accuracy. This suggested method is applied and the DARPA's KDD cup & its 1999 dataset are utilized for simulations. Based on such theoretical and its analysis of performance, it is shown that this definite suggested model might identify entire anomalies utilizing greater accuracy and lesser rate of false alarm and might also surpass such classifiers called Naive Bayes and ANN.

The work [23] presents that other "Cascade of the ensemble-based ANN for a multi-class Intrusion Detection (CANID)" is utilized in network traffic of computer. This suggested method will learn further that several NN are linked to 1 cascade for every of such kind of networks which are trained by utilizing small training samples. This proposed novel cascade infrastructure is capable for delivering the small training samples which need to be utilized with learning algorithm based on boosting is employed in optimal set learning of NN factors for every such following partition. Here, the outcomes of this simulation envisioned that suggested contribution will now be capable for detecting effectively the entire diverse cyber-attacks in the networks of computer.

The work [24] presents that HIDS algorithm is proposed that was utilized aimed at private environment of cloud, which was deliberated as effective for the cause of performance and security. The contemporary IDS is capable for providing fine picture of system, but is not capable for identifying intrusion in productive way.

The AI (artificial intelligence) is now included in this research contribution for identifying any kind of intrusion in instance of private cloud and the inclusion of AI method resulted in the IDA- self adaptive and is tested utilizing practical data by collecting the data at same time.

Here, this algorithm is applied in instance of private cloud which is highly secured constructed for the cause of the military and here the sector of banking for observing the network actions.

There are several algorithms of meta-heuristic implemented aimed at dealing with the scheduling issues. The work [25] presents other in-depth examination of PSO.

Its task and workflow scheduling strategies are suggested aimed at the environment of cloud in this review. Here, it offers proposed strategies classifications based on PSO is implemented and finally further directions of research is outlined.The work [26] presents that further suggested parallel designs and realization model for optimized PSO-BP, is the NN on the basis of Map reduce on the platform of Hadoop through parallel designs & PSO. This is utilized for BP-NN optimization and its primary weights and the thresholds aimed at enhancing the classification of algorithm and its precision. The parallel programming method based on the Map reduce is utilized for attaining parallel processing of BP aimed at solving communication overhead & hardware on the basis of issue during BP and big data is addressed by NN. Here algorithm is proposed further for this network and is showing greater classification accuracy and enhanced effectiveness of time can depict the prominent enhancement attaining from parallel processing towards intelligent algorithm aimed at big-data. Better solutions are detected by Hyper-heuristic algorithms (HHA) for CC scheduling and also for further enhancing such outcomes of scheduling leads for make-span. The work [27] proposed other new “multi objective PSO task scheduling (MO-PSO)” and GA on the basis of HHA aimed at resource scheduling that was hybrid-algorithm. Here, algorithm performance is assessed utilizing Sim-toolkit of cloud. They compared the algorithm of hybrid scheduling possessing contemporary common heuristic & scheduled algorithms. Their outcomes show better execution than contemporary algorithms through lessening cost and enhancing the make-span. The suggested method has presented enhanced resources utilization, throughput & make-span.

III. METHODS AND MATERIALS

The proposed CFID is meant to define Calibration Factors to estimate the intrusion scope of network transaction of cloud computing. Since the proposal is a machine learning method, it is being trained by set of network transactions that are labeled as biased (prone to intrusion) or unbiased (fair enough network transaction). A cloud computing network transaction is said to be the values representing the set of sequence of attributes [28]. In order to discover the calibration factors, the initial phase of the method discovers all possible unique subsets of the values representing the corresponding subset of attributes. Here after, these unique subsets of values are being referred as features.

In order to assess each feature effect(s) with respect to patterns discovered from unbiased and biased training sets using a graph strategy. Hence the associated effects in this process are meant to define and derive the Scale of Calibration Factors within the biased and unbiased labels.

The main objective of the present CFID method is to design a Scale of Calibration Factors using the knowledge of cached transactions that are labeled biased Remote-to-Local (R2L), port scanning (PS), and virtual machine trapping (VMT) or unbiased. To do this, the given labeled data is partitioned into respective biased labels called Remote-To-Local (R2L), port scanning (PS), and virtual machine trapping (VMT) as well as unbiased (healthy person blood samples). The given training corpuses of records with labels are classified into their respective categories based on its label. Because each record contains many numbers of attributes and a huge chunk of them might be insignificant to the respective category of biased. Therefore, the first step in the present method is conducting the feature optimization process for the elimination of the features that are insignificant. Further, the confidence assessment for every category of cloud network transaction data is performed. A new procedure is described here for the assessment of every cloud network transaction confidence against the features of every category. In addition, the confidence estimated for every feature of the cloud network transaction of the respective category shall be used as input for defining Scale of Calibration Factors for the estimation of the scope of Remote-To-Local, port scanning, as well as virtual machine trapping. The whole process of the proposed method is described in the following Algorithm.

Algorithm for CFID:

Step 1. Elimination of insignificant features using Feature optimization Process

Step 2. Confidence Assessment for features and records using ANOVA methods

Step 3. Identification of Scale of Calibration Factors to predict R2L, PS, VMT, and Unbiased categories of the cloud network transaction.

Step 4. Testing and Experimental study.

The details of each step are discussed in the subsequent sections.

3.1 Elimination of insignificant features using Feature optimization Process

Let dataset $D_i = \{e(i)_1, e(i)_2, ...e(i)_{|D_i|}\}$ of size $|D_i|$ for each record i, shall be taken into consideration for training towards defining Scale of Calibration Factors. Every cloud network transaction is generally represented by the sequence of attributes, which are chosen for the respective attack prone context. The description is binding to all cloud network transactions that are labeled as R2L, PS, and VMT.

Let $D_n = \{r_1, r_2, ..., r_{|D_n|}\}$ be the set of cloud network transactions of unbiased labels, similarly, the set $D_m = \{r_1, r_2, ..., r_{|D_m|}\}$ be the cloud network transactions labeled as malevolent.

The sets $F_i = \{f(i)_1, f(i)_2, ..., f(i)_{|F_i|}\}$ and $F_n = \{f(n)_1, f(n)_2, ..., f(n)_{|F_n|}\}$ are the feature sets of cloud network transactions which are represented by D_i and D_n in that order.

The attribute set $G(i)_j = \{g(ij)_1, g(ij)_2, ...g(ij)_{|G(i)_j|}\}$ be the set of attributes as values observed for feature $f(i)_j$ of cloud network transactions represented by D_i . On the same note,

the attribute set $G(n)_j = \{g(nj)_1, g(nj)_2, ...g(nj)_{|G(n)_j|}\}$ be the set of attributes as values observed for feature $f(n)_j$ of cloud network transactions represented by D_n . The following procedure is used to eliminate the insignificant features and thereby reduce the process complexity for the subsequent steps.

Feature $f(i)_j$ of F_i is generally said to be an insignificant feature, when attributes $G(i)_j$ of $f(i)_j$ are almost same as the attributes $G(n)_j$ of feature $f(n)_j$ of F_n . Therefore, the identification of insignificant features requires the adoption of hamming distance that applied on attributes of every feature as vectors from every attack prone and normal cases. It is also worth pointing out that the hamming distance with zero or less than the stipulated threshold points out that respective feature is generally insignificant. The computation of hamming distance process is explained below:

Let $CX = \{cx_1, cx_2, ..., cx_n\}$ and $CY = \{cy_1, cy_2, ..., cy_m\}$ be two vectors of size n and m respectively and let $CZ \leftarrow \phi$ is a vector of size 0. $CZ\{i\}$ is the i^{th} component of the vector CZ and $|CZ|$ is the vector CZ 's size. $hd_{CX \leftrightarrow CY}$ is the hamming distance between CX and CY , Then, the computation of Hamming distance between CX and CY is described below:

```
Step 1
 $foreach\ \{i \exists i = 1, 2, 3, .....\max(n.m)\}$
 $if\ (\{cx_i \exists cx_i \in CX\} - \{cy_i \exists cy_i \in CY\}) \equiv 0\ then$
 $CZ \leftarrow \{cx_i \exists cx_i \in CX\} - \{cy_i \exists cy_i \in CY\}$
Else
 $CZ \leftarrow 1$
      Step 2.
```

The hamming distance $hd_{CX \leftrightarrow CY}$ is given by

$$hd_{CX \leftrightarrow CY} = \sum_{j=1}^{|CZ|} CZ\{i\}$$

In the cloud network transaction, if the distance between the attribute features of the attack prone and normal is more, then the feature is to be considered as optimal feature otherwise the feature is ignored. In this way the dataset is optimized. The optimized dataset is used in the confidence assessment discussed in section 3.2.

3.2 Attribute and Cloud network transaction Confidence Assessment

The attributes found for all the optimal feature of the respective cloud network transaction dataset, as well as the cloud network transactions of the dataset shall be used further for the assessment of the attribute and the cloud network transaction confidence. The whole process of the proposed method is described in the following Algorithm.

3.2.1 Identification of Attribute Pairs

Define the attribute pairs in such a way that every attribute representing a different feature of the similar dataset. To do this attribute pair correlation is computed by using the following procedure

Let P_i be the set and contains every possible unique attribute pairs from the respective dataset D_i .

For every cloud network transaction $e(i)_j$ of the respective dataset D_i , find every possible unique pairs of attributes and add to P_i . They are: $\{g_k \exists g_k \in p_j\}$ and $\{g_l \exists g_l \in p_j\}$ be the two attributes paired as $\{p_j \exists p_j \in P_i\}$.

3.2.2 Evaluate the associativity support of every pair of an attribute

When a pair of attributes is present in a cloud network transaction with different features of the dataset then the associativity support will be measured by taking the ratio of the number of such pairs with total number of attributes present in the cloud network transaction of the dataset. The ratio $s(p_i)$ is the correlation of every pair of the attributes in the cloud network transactions. The correlation of every pair $\{p_j \exists p_j \in P_i\}$ as follows.

Let $\{g_k \exists g_k \in p_j\}$ and $\{g_l \exists g_l \in p_j\}$ be the two attributes paired as $\{p_j \exists p_j \in P_i\}$, and then the correlation $s(p_j)$ of the pair p_j is given in the following equation.

$$s(p_i) = \frac{1}{|D_i|} \sum_{v=1}^{|D_i|} \{1 \exists \{g_k, g_l\} \subseteq e(i)_v\}$$

This correlation is used in the assessment of attribute and cloud network transaction confidence of the respective cloud network transaction datasets of port scanning, Remote-To-Local, virtual machine trapping, and the unbiased cases. The details are discussed in section 3.3.

3.2.3 Assessment of Attribute and Transaction Confidence

The assessment of the confidence of the attributes as well as the cloud network transactions of the respective cloud network transaction dataset D_i , a mutual relation graph shall be formed between the cloud network transactions, as well as the attributes of respective D_i .

There shall be an edge between an attribute and cloud network transaction when the chosen attribute exists in the given cloud network transaction.

Then every edge between attribute and cloud network transaction is weighted as indicated below:

```
$\forall_{j=1}^{|G(i)|}\{g_j \exists g_j \in G(i)\}$ Begin
    $\forall_{k=1}^{|D_i|}\{e(l)_k \exists g_j \in e(l)_k\}$ Begin
        $w_{g_j} = 0$
        $\forall_{m=1}^{|e(l)_k|}\{g_m \exists g_m \in e(l)_k \wedge g_j \neq g_m\}$ Begin
            $p_m = \{g_j, g_m\}$
            $w(g_j) += s(p_m)$
        End
        $w_{g_j \leftrightarrow e(l)_k} = \frac{1}{|e(l)_k| - 1} w(g_j)$
    End
End
```

The weights that are obtained for the edges between attributes and cloud network transactions in the mutual graph are further utilized in the assessment of the attribute and cloud network transaction confidence towards the respective Remote-To-Local (R2L), port scanning (PS), and virtual machine trapping (VMT) as well as unbiased datasets.

Additionally, the measurement c_{gj} is the feature confidence of the cloud network transaction dataset D_i . This can be found by aggregating the weight of attribute g_j of every cloud network transaction $e(i)_k$ of the respective dataset D_i . Similarly the respective attribute confidence towards dataset D_i is also estimated by the same formula.

$\forall_{j=1}^{|G(i)|}\{g_j \exists G(i) \ni g_j\}$ Begin

$$c_{g_j \Rightarrow D_i} = \sum_{k=1}^{|D_i|}\{w(g_j) \exists e(i)_k \ni g_j \wedge D_i \ni e(i)_k\}$$

End

Similarly, every respective cloud network transaction confidence for every cloud network transaction dataset D_i is measured by computing the sum of the product of every attribute weight as well as the respective attribute confidence. The attribute is existing in selective cloud network transaction is the confidence of the given cloud network transaction. The confidence measures are given in the following equation.

$\forall_{j=1}^{|D_i|}\{e(i)_j \exists D_i \ni e(i)_j\}$ Begin

$$c_{e(i)_j \Rightarrow D_i} = \sum_{k=1}^{|G(i)|}\{w(g_k) \otimes c_{g_k \Rightarrow D_i} \exists e(i)_j \ni g_k \wedge D_i \ni e(i)_j\}$$

End

The confidence measures of the attributes and the cloud network transactions of every respective cloud network transaction datasets are used for all R2L, PS, VMT, and unbiased cases.

3.4 Measuring Calibration Factors to find the Scope of R2L, PS, VMT, and Unbiased Cases

The confidence of records prone to divergent attacks datasets D_{R2L} , D_{PS} , D_{VMT} used to calculate the Aggregate mean of the respective cloud network transactions confidence of the cloud network transaction dataset D_{R2L} having records biased as Remote-To-Local attack is defined below:

$$m_{R2L} = \frac{1}{|D_{R2L}|}\sum_{i=1}^{|D_{R2L}|}\{c_{e(R2L)_i \Rightarrow D_{R2L}} \exists D_{R2L} \ni e(R2L)_i\}$$

So as to identify the upper and the lower bounds of m_{R2L} , mean absolute distance of D_{R2L} is evaluated as indicated in following equation,

$$m_{R2L}^{ad} = \frac{1}{|D_{R2L}|}\sum_{i=1}^{|D_{R2L}|}\sqrt{\left(m_{R2L} - c_{e(R2L)_i \Rightarrow D_{R2L}}\right)^2}$$

Then the upper and lower bounds of m_{R2L} is measured in following equations:

Lower bound of m_{R2L} is: $ml_{R2L} = m_{R2L} - m_{R2L}^{ad}$

Upper bound of m_{R2L} is: $mu_{R2L} = m_{R2L} + m_{R2L}^{ad}$

In addition, Meta-heuristics for VMT (virtual machine trapping), PS (port scanning), as well as the unbiased (healthy) scope. Aggregate mean of the respective cloud network transactions confidence of port scanning cloud network transaction dataset D_{PS} is:

$$m_{PS} = \frac{1}{|D_{PS}|}\sum_{i=1}^{|D_{PS}|}\{c_{e(PS)_i \Rightarrow D_{PS}} \exists D_{PS} \ni e(PS)_i\}$$

The mean absolute distance of D_{PS} is followed in below equation:

$$m_{PS}^{ad} = \frac{1}{|D_{PS}|}\sum_{i=1}^{|D_{PS}|}\sqrt{\left(m_{PS} - c_{e(PS)_i \Rightarrow D_{PS}}\right)^2}$$

Then the lower and upper bounds of m_{PS} is evaluated in following equations:

Lower bound of m_{PS} is: $ml_{PS} = m_{PS} - m_{PS}^{ad}$

Upper bound of m_{PS} is: $mu_{PS} = m_{PS} + m_{PS}^{ad}$

Aggregate mean of the respective cloud network transactions confidence of virtual machine trapping cloud network transaction dataset D_{VMT} is:

$$m_{VMT} = \frac{1}{|D_{VMT}|} \sum_{i=1}^{|D_{VMT}|} \{c_{e(VMT)_i \Rightarrow D_{VMT}} \exists D_{VMT} \ni e(VMT)_i\}$$

The mean absolute distance of D_{VMT} is followed in below equation:

$$m_{VMT}^{ad} = \frac{1}{|D_{VMT}|} \sum_{i=1}^{|D_{VMT}|} \sqrt{\left(m_{VMT} - c_{e(VMT)_i \Rightarrow D_{VMT}}\right)^2}$$

Then the lower and upper bounds of m_{VMT} is assessed in following equations:

Lower bound of m_{VMT} is: $ml_{VMT} = m_{VMT} - m_{VMT}^{ad}$

Upper bound of m_{VMT} is: $mu_{VMT} = m_{VMT} + m_{VMT}^{ad}$

Aggregate mean of the respective unbiased cloud network transactions confidence of dataset D_U is:

$$m_U = \frac{1}{D_U} \sum_{i=1}^{|D_U|} \{c_{e(U)_i \Rightarrow D_U} \exists D_U \ni e(U)_i\}$$

The mean absolute distance of D_U is denoted in below equation,

$$m_U^{ad} = \frac{1}{|D_U|} \sum_{i=1}^{|D_U|} \sqrt{\left(m_U - c_{e(U)_i \Rightarrow D_U}\right)^2}$$

Then the lower and upper bounds of m_U is assessed as followed in below equations:

Lower bound of m_U is: $ml_U = m_U - m_U^{ad}$

Upper bound of m_U is: $mu_U = m_U + m_U^{ad}$

3.5 Predicting the state of cloud network transaction:

The meta-heuristics will be used further to assess the R2L, PS, and VMT scope of a given cloud network transaction e. The confidence of given cloud network transaction.

The aggregate of the product of each attribute confidence and weight of that exists in $G(D_{R2L})$ and e, which divides by the aggregate of confidence of all attributes exists in $G(D_{R2L})$ is given below:

$$c_{e \Rightarrow R2L} = \frac{1}{\sum_{j=1}^{|G(D_{R2L})|} \left\{c_{g_j \Rightarrow R2L} \otimes w(g_j) \exists g_j \in G(D_{R2L})\right\}} \sum_{i=1}^{|G(D_{R2L}}$$

Further the confidence of e towards D_{PS}, D_{VMT} and D_U assessed by finding the aggregate of the product of every attribute confidence and the weight that exists in $G(D_{PS})$ and e. The aggregate of confidence of all attributes exists in $G(D_{PS})$ is given by:

$$c_{e \Rightarrow PS} = \frac{1}{\sum_{j=1}^{|G(PS)|} \left\{c_{g_j \Rightarrow PS} \otimes w(g_j) \exists g_j \in G(PS)\right\}} \sum_{i=1}^{|G(PS)|} \left\{c_{g_i \Rightarrow PS} \otimes w(g_i) \exists g_i \in G(PS) \wedge e \ni g_i\right\}$$

The aggregate of the product of every attribute confidence and weight that exists in $G(D_{VMT})$ and e, that divides by the aggregate of confidence of all attributes exists in $G(D_{VMT})$ is:

$$c_{e \Rightarrow VMT} = \frac{1}{\sum_{j=1}^{|G(VMT)|} \left\{c_{g_j \Rightarrow VMT} \otimes w(g_j) \exists g_j \in G(VMT)\right\}} \sum_{i=1}^{|G(VMT)|} \left\{c_{g_i \Rightarrow VMT} \otimes w(g_i) \exists g_i \in G(VMT) \wedge e \ni g_i\right\}$$

The aggregate of the product of every attribute confidence and weight that exists in $G(D_U)$ and e, that is divides by the aggregate of confidence of all attributes exists in $G(D_U)$ is:

$$c_{e \Rightarrow U} = \frac{1}{\sum_{j=1}^{|G(U)|} \left\{c_{g_j \Rightarrow U} \otimes w(g_j) \exists g_j \in G(U)\right\}} \sum_{i=1}^{|G(U)|} \left\{c_{g_i \Rightarrow U} \otimes w(g_i) \exists g_i \in G(U) \wedge e \ni g_i\right\}$$

Then the confidence values of the cloud network transaction e with regards to PS, R2L, VMT, and U shall be used for the estimation of the given expression state is unbiased, prone to Remote-To-Local, port scanning or virtual machine trapping according to the following conditions.

Confidence values of the cloud network transaction e used for the estimation of the given expression state is unbiased, prone to Remote-To-Local, port scanning or virtual machine trapping according to the following conditions.

Confidence values Condition	Attack Prone Prediction
$(c_{e \Rightarrow R2L} \geq mu_{R2L}) \vee (c_{e \Rightarrow PS} \geq mu_{PS}) \vee (c_{e \Rightarrow VMT} \geq mu_{VMT})$	Remote-To-Local Confirmed (highly prone to either of three attack prone conditions)
$(c_{e \Rightarrow R2L} \geq m_{R2L}) \wedge (c_{e \Rightarrow PS} \geq ml_{PS}) \wedge (c_{e \Rightarrow VMT} \geq ml_{VMT})$	Remote-To-Local Confirmed (prone to R2L and either or both of the PS, and VMT)
$(c_{e \Rightarrow R2L} \geq ml_{R2L}) \wedge (c_{e \Rightarrow PS} \geq ml_{PS}) \wedge (c_{e \Rightarrow VMT} \geq ml_{VMT}) \wedge (c_{e \Rightarrow U} < m_{VMT})$	Prone to Remote-To-Local
$(c_{e \Rightarrow R2L} < ml_{R2L}) \wedge (c_{e \Rightarrow PS} < ml_{PS}) \wedge (c_{e \Rightarrow VMT} < ml_{VMT}) \wedge (c_{e \Rightarrow U} > m_{VMT})$	Unbiased state Confirmed
$(c_{e \Rightarrow R2L} < m_{R2L}) \wedge (c_{e \Rightarrow PS} < m_{PS}) \wedge (c_{e \Rightarrow VMT} < m_{VMT}) \wedge (c_{e \Rightarrow U} \geq mu_{VMT})$	Prone to Unbiased state

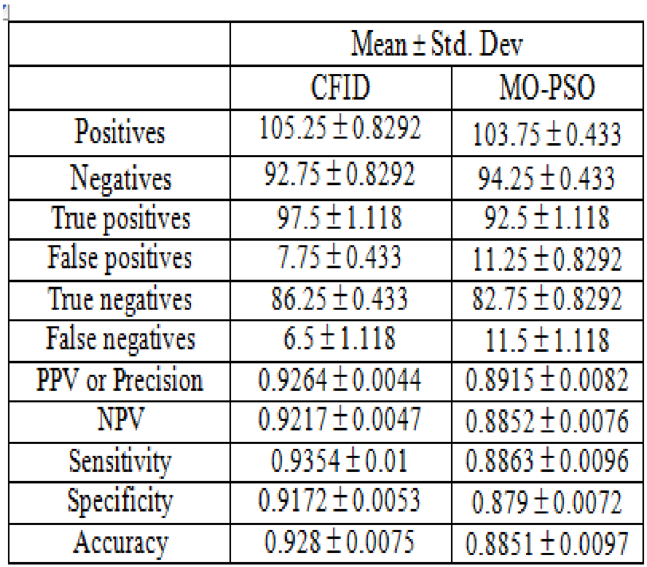
IV. EXPERIMENTAL RESULTS

The dataset CIDDS [29] is utilized in the simulation. The number of records in the involved dataset were 1979 and are labeled as intrude-1043 (R2L: 301, PS: 418, VMT: 324) & benevolent-936, which are deliberated for the simulation study. Here, in respect to estimate the execution of proposal, the 4-fold classification scheme is modified.

The specified records are segregated into 4 folds so that every repetition of simulation, the records of 3-folds are utilized for determining the measure and other fold is utilized for testing. Scale assessment statistics are detailed in Table 1.

Proposal execution is measured by comparing to existing method known as "multi objective PSO task scheduling (MO-PSO)" [27].

Table 1: The statistics noticed for performance metrics in the form of total performance (mean & standard deviation for 4-folds)



	Mean ± Std. Dev	
	CFID	MO-PSO
Positives	105.25 ± 0.8292	103.75 ± 0.433
Negatives	92.75 ± 0.8292	94.25 ± 0.433
True positives	97.5 ± 1.118	92.5 ± 1.118
False positives	7.75 ± 0.433	11.25 ± 0.8292
True negatives	86.25 ± 0.433	82.75 ± 0.8292
False negatives	6.5 ± 1.118	11.5 ± 1.118
PPV or Precision	0.9264 ± 0.0044	0.8915 ± 0.0082
NPV	0.9217 ± 0.0047	0.8852 ± 0.0076
Sensitivity	0.9354 ± 0.01	0.8863 ± 0.0096
Specificity	0.9172 ± 0.0053	0.879 ± 0.0072
Accuracy	0.928 ± 0.0075	0.8851 ± 0.0097

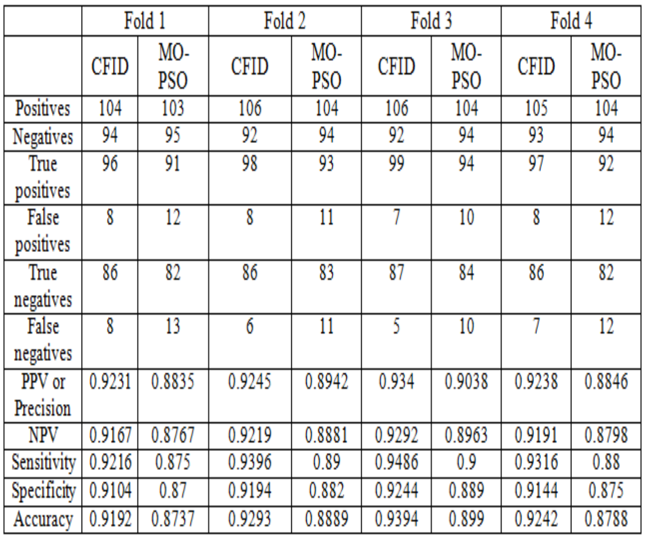
In the experimental study, the statistics noticed for the performance metrics in the form of the overall performance is presented in Table 1. Here, the mean & standard deviation of 4-folds is explored in table. The various performance metrics used for observing the overall performance, and the metrics such as positives, negatives, true positives and many more as depicted in table. The mean & standard deviation for CFID & MO-PSO at various performance metrics are represented. The number of positives for the proposed method CFID is 105.25 ± 0.8292 and, the number of positives for the MO-PSO method is 103.75 ± 0.433. The number of negatives for CFID is 92.75 ± 0.8292 and for MO-PSO is 94.25 ± 0.433. Similarly from the table, we can observe that the number of TPs for CFID is higher while it is compared to MO-PSO. The number of FPs for CFID is less than MO-PSO. The amount of true-negatives of CFID is more than MO-PSO. The amount of FNs for CFID is less than MO-PSO. The amount of PPV for CFID is more than MO-PSO. The amount of NPV for CFID is more than MO-PSO. The sensitivity, specificity and accuracy for CFID are higher while it is compared to MO-PSO.

Performance Analysis

In the empirical study, from the given inputs in testing stage, the correctly labeled benign records are 82.75 ± 0.8292, 86.25 ± 0.433 and correctly labeled intruded records were 92.5 ± 1.118, 97.5 ± 1.118 of Contemporary method MO-PSO and proposed method CFID respectively.

The sensitivity noticed for the contemporary MO-PSO method & proposed CFID method in corresponding sequence are 0.8863 ± 0.0096 & 0.9354 ± 0.01. Identically, specificity is noticed in the same sequence as 0.879 ± 0.0072 & 0.9172 ± 0.0053. These 2 metrics denotes the CFID performance advantage over MO-PSO in respect to the intrusion-detection that is because of specificity & sensitivity (recall) of MO-PSO is less than CFID. The truly identified intruded records in contradiction to entire records detected in the form of intruded indicates the prediction value of intrusion, which often indicates as PPV (positive predictive value) identified to be 0.8915 ± 0.0082 & 0.9264 ± 0.0044 of MO-PSO & CFID in corresponding sequence. Identically, truly identified normal records in averse to entire records detected in the form of normal records would be predictive value of normal record, which often indicates NPV (Negative predictive value) found as 0.8852 ± 0.0076 & 0.9217 ± 0.0047 of MO-PSO & CFID in corresponding order. The NPV & PPV values indicate that CFID are more prominent than existing MO-PSO methods. The accuracy metric indicates ratio of truly-predicted records in contradiction to entire records specified in the form of input aimed at the stage of testing that is often noticed as 0.8851 ± 0.0097 & 0.928 ± 0.0075 for MO-PSO & CFID in corresponding order. Here, accuracy values from simulation study show that proposal is more prominent than contemporary method. These statistics showing that proposed heuristics for measuring scope of benign & intrusion transactions were prominent for discriminating network traffic as benign & intrude though an accuracy, which is much greater than existing method. The represented recall for the suggested CFID method signifies that miss-rate is lower than existing methods. Nonetheless, the comprehensive statistics of every fold of simulation are depicted in Table 2 and Figure 1 - Figure 5.

Table 2: Statistics noticed for the performance metrics under 4-fold schemes



	Fold 1		Fold 2		Fold 3		Fold 4	
	CFID	MO-PSO	CFID	MO-PSO	CFID	MO-PSO	CFID	MO-PSO
Positives	104	103	106	104	106	104	105	104
Negatives	94	95	92	94	92	94	93	94
True positives	96	91	98	93	99	94	97	92
False positives	8	12	8	11	7	10	8	12
True negatives	86	82	86	83	87	84	86	82
False negatives	8	13	6	11	5	10	7	12
PPV or Precision	0.9231	0.8835	0.9245	0.8942	0.934	0.9038	0.9238	0.8846
NPV	0.9167	0.8767	0.9219	0.8881	0.9292	0.8963	0.9191	0.8798
Sensitivity	0.9216	0.875	0.9396	0.89	0.9486	0.9	0.9316	0.88
Specificity	0.9104	0.87	0.9194	0.882	0.9244	0.889	0.9144	0.875
Accuracy	0.9192	0.8737	0.9293	0.8889	0.9394	0.899	0.9242	0.8788

The statistics noticed for the metrics of performance under 4-fold schemes for the proposed method CFID and contemporary method MO-PSO is depicted in Table 2. At fold-1, the performance of CFID and MO-PSO at various performance metrics is presented, where the number of positives for CFID is 104 and MO-PSO is 103.

The amount of negatives for CFID is higher while it is compared with MO-PSO. Similarly, the performance of CFID & MO-PSO at various metrics is depicted in the above Table 2. At Fold-2, the number of positives for CFID is 106 and MO-PSO is 104, whereas the amount of negatives for CFID is higher while it is compared to contemporary method MO-PSO.

Similarly, the performances of CFID and MO-PSO at fold-3, fold-4 over various metrics are presented in Table 2.

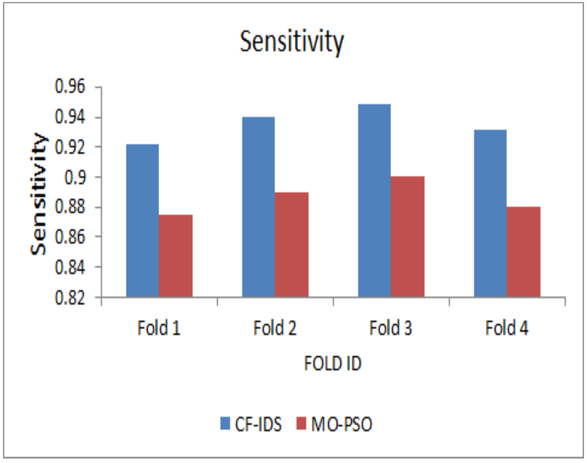


Figure 1: The value of Sensitivity noticed for CFID and MO-PSO

The sensitivity value noticed for the proposed method CFID & contemporary method MO-PSO is depicted in Figure 1. From the figure, it is observed that graph is drawn among sensitivity and four folds. The sensitivity of CFID at fold-1 is greater while it is compared to MO-PSO. As observed, it is clear that the recall of proposed method CFID is more than contemporary method MO-PSO.

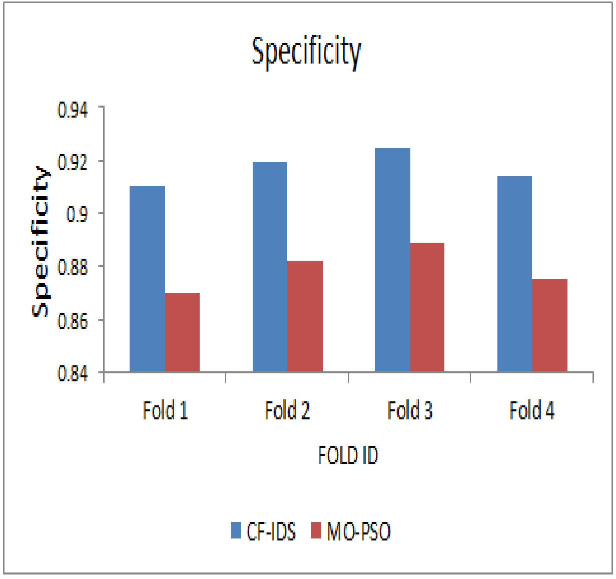


Figure 2: The value of specificity noticed for CFID and MO-PSO

The specificity value noticed for the proposed method CFID & contemporary method MO-PSO is depicted in Figure 2. From the figure, it is observed that graph is drawn among specificity and 4-folds.

The specificity of CFID at fold-1 is greater, while it is compared to MO-PSO. As observed, it is clear that proposed method CFID specificity value is higher than contemporary method MO-PSO.

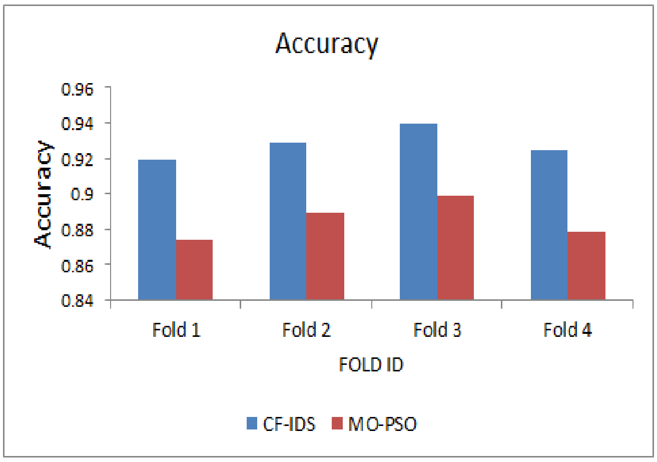


Figure 3: The value of accuracy noticed for CFID and MO-PSO

The Accuracy value noticed for the proposed method CFID & contemporary method MO-PSO is depicted in Figure 3. From the figure, it is observed that graph is drawn among Accuracy & 4-folds. The Accuracy of CFID at fold-1 is greater while it is compared to MO-PSO. As observed, it is clear that the Accuracy of proposed method CFID is higher than contemporary method MO-PSO.

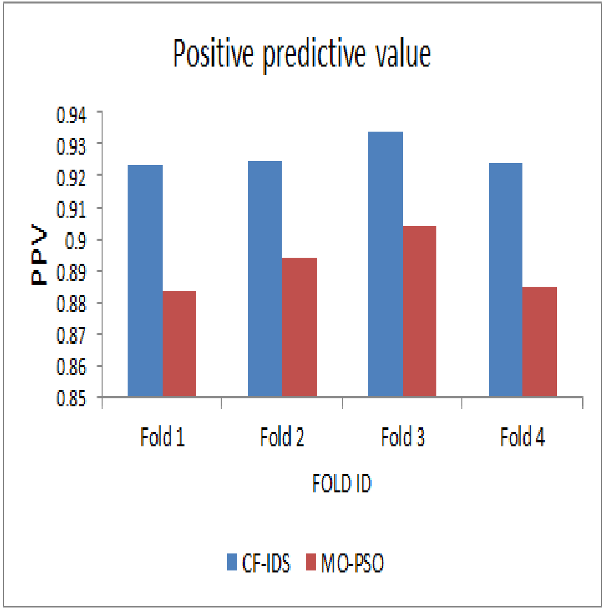


Figure 4: The value of PPV (precision) noticed for CFID and MO-PSO

The PPV value noticed for the proposed method CFID & contemporary method MO-PSO is depicted in Figure 4. From the figure, it is observed that graph could be drawn among PPV and 4-folds.

The PPV of CFID at fold-1 is greater while it is compared to MO-PSO.

As observed, it is clear that the PPV of proposed method CFID is higher than contemporary method MO-PSO.

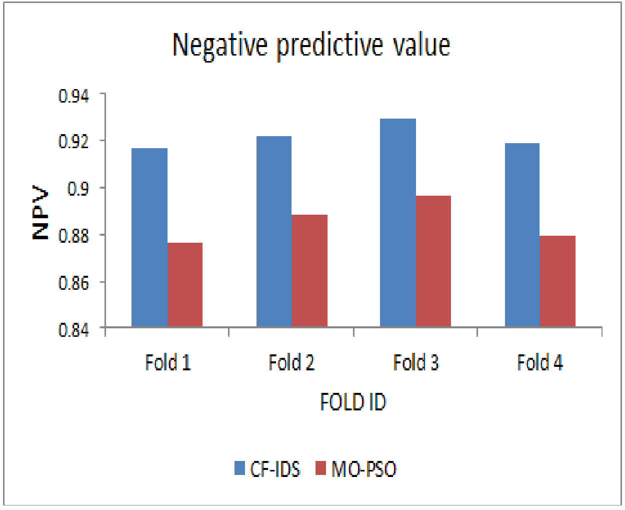


Figure 5: The value of NPV noticed for CFID and MO-PSO

The value of NPV noticed for the proposed method CFID & contemporary method MO-PSO is depicted in Figure 5. From the figure, it is observed that graph is plotted among NPV & 4-folds. The NPV of CFID at fold-1 is greater while it is compared to MO-PSO. As observed, it is clear that the NPV of proposed method CFID is higher than contemporary method MO-PSO.

V. CONCLUSION

This contribution portrayed a model for defending internet protocol based intrusion activities on distributed cloud computing that labeled as “Calibration Factors based Intrusion Detection (CFID)”. This method derives calibration factors from the given training corpus that contains the cloud computing network transactions labeled as biased (R2L, PS or VMT) or unbiased. These calibration factors of biased and unbiased scope are used further to estimate the attack scope of the new cloud computing network transaction. Experimental study portrayed the significance of the proposal, which has been scaled by comparing with other contemporary model built on particle swarm optimization technique. This method can be escalated to next level of detection accuracy in future research that deals with other possible attacks The other dimension may evince the scope to use the projected calibration factors as fitness function to the evolutionary methods.

REFERENCES

1. Mell, Peter, and Timothy Grance. "The NIST Definition of Cloud Computing.[Online] Available at: http://csrc. nist. gov/publications/nistpubs/800-145." SP800-145. pdf (2011).
2. “Google apps engine.” [Online]. Available: URLhttp://code.google.com/appengine.
3. Amazon web services. [Online]. Available: http://aws.amazon.com
4. Google apps. [Online]. Available: http://www.google.com/apps/business
5. Opennebula. [Online]. Available: http://www.opennebula.org
6. Eucalyptus. [Online]. Available: http://eucalyptus.cs.ucsb.edu/.
7. Azure services platform. [Online]. Available:http://www.microsoft.com/azure
8. International Data Corporation. [Online]. Available:http://blogs.idc.com/ie/wpcontent/uploads/2009/12/idc_cloud_challenges_2009.jpg, 2009.
9. Lockheed Martin White Paper: Available:http://www.lockheedmartin.com/data/assets/isgs/documents/CloudComputingWhitePaper.pdf
10. C. Brooks. Amazon EC2 Attack Prompts Customer Support Changes.Tech Target. [Online]. Available:http://searchcloudcomputing.techtarget.com/news/article/0,289142,sid201_gci1371090,00.html
11. Chen, Yao, and Radu Sion. "On securing untrusted clouds with cryptography." Proceedings of the 9th annual ACM workshop on Privacy in the electronic society. ACM, 2010.
12. Mazzariello, Claudio, Roberto Bifulco, and Roberto Canonico. "Integrating a network ids into an open source cloud computing environment." 2010 Sixth International Conference on Information Assurance and Security. IEEE, 2010.
13. Schulter, K. "Intrusion detection for grid and cloud computing." IEEE IT Professional Journal 7 (2010).
14. Kumar, Pardeep, et al. "A novel approach for security in cloud computing using hidden markov model and clustering." 2011 World Congress on Information and Communication Technologies. IEEE, 2011.
15. Modi, Chirag N., et al. "Bayesian Classifier and Snort based network intrusion detection system in cloud computing." 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12). IEEE, 2012.
16. Kannan, Anand. "Performance evaluation of security mechanisms in Cloud Networks." (2012).
17. Ficco, Massimo, Luca Tasquier, and Rocco Aversa. "Intrusion detection in cloud computing." 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing. IEEE, 2013.
18. Xiong, Wei, et al. "Anomaly secure detection methods by analyzing dynamic characteristics of the network traffic in cloud communications." Information Sciences 258 (2014): 403-415.
19. Kene, Snehal G., and Deepti P. Theng. "A review on intrusion detection techniques for cloud computing and security challenges." 2015 2nd International Conference on Electronics and Communication Systems (ICECS). IEEE, 2015.
20. Ilgun, Koral, Richard A. Kemmerer, and Phillip A. Porras. "State transition analysis: A rule-based intrusion detection approach." IEEE transactions on software engineering 3 (1995): 181-199.
21. Ghosh, Partha, Abhay Kumar Mandal, and Rupesh Kumar. "An efficient cloud network intrusion detection system." Information systems design and intelligent applications. Springer, New Delhi, 2015. 91-99.
22. Pandeeswari, N., and Ganesh Kumar. "Anomaly detection system in cloud environment using fuzzy clustering based ANN." Mobile Networks and Applications 21.3 (2016): 494-505.
23. Baig, Mirza M., Mian M. Awais, and El-Sayed M. El-Alfy. "A multiclass cascade of artificial neural network for network intrusion detection." Journal of Intelligent & Fuzzy Systems 32.4 (2017): 2875-2883.
24. Rajendran, Praveen Kumar, M. Rajesh, and R. Abhilash. "Hybrid intrusion detection algorithm for private cloud." Indian Journal of Science and Technology 8.35 (2015): 1-10.
25. Masdari, Mohammad, et al. "A survey of PSO-based scheduling algorithms in cloud computing." Journal of Network and Systems Management 25.1 (2017): 122-158.
26. Cao, Jianfang, et al. "Big data: A parallel particle swarm optimization-back-propagation neural network algorithm based on MapReduce." PloS one 11.6 (2016): e0157551.
27. Kumari, K. Raja, P. Sengottuvelan, and J. Shanthini. "A hybrid approach of genetic algorithm and multi objective PSO task scheduling in cloud computing." Asian Journal of Research in Social Sciences and Humanities 7.3 (2017): 1260-1271.

28. Aldribi, Abdulaziz, Issa Traore, and Belaid Moa. "Data Sources and Datasets for Cloud Intrusion Detection Modeling and Evaluation." Cloud Computing for Optimization: Foundations, Applications, and Challenges. Springer, Cham, 2018. 333-366.

29. Kholidy, Hisham A., and Fabrizio Baiardi. "Cidd: A cloud intrusion detection dataset for cloud computing and masquerade attacks." 2012 Ninth International Conference on Information Technology-New Generations. IEEE, 2012.