A Meta-analytic Review of Intelligent Intrusion Detection Techniques in Cloud Computing Environment

Meghana G Raj, Dr. Santosh Kumar Pani
Department School of Computer Engineering, Kalinga Institute of Industrial Technology
Deemed to be University, Bhubaneswar, India

Abstract—Security and data privacy continue to be major considerations in the selection and study of cloud computing. Organizations are migrating more critical operations to the cloud, resulting in increase in the number of cloud vulnerability incidents. In recent years, there have been several technological advancements for accurate detection of attacks in the cloud. Intrusion Detection Systems (IDS) are used to detect malicious attacks and reinstate network security in the cloud environment. This paper presents a systematic literature review and a meta-analysis to shed light on intelligent approaches for IDS in cloud. This review focuses on three intelligent IDS approaches—Machine Learning Algorithms, Computational Intelligence Algorithms and Hybrid Meta-Heuristic Algorithms. A qualitative review synthesis was carried out on a total of 28 articles published between 2016 and 2021. This study concludes that IDS based on Hybrid Meta-Heuristic Algorithms have increased Accuracy, decreased False Positivity Rate and increased Detection Rate.

Keywords—Intrusion detection system (IDS); machine learning; computational intelligence algorithms; hybrid meta-heuristic algorithms; cloud security; cloud computing

I. INTRODUCTION

Cloud Computing (CC) provides on-demand network access to a group of configurable computing assets like servers, services, applications, storage, and networks that could be rapidly released with lesser management endeavors or service provider interaction. While it offers many benefits, one of the main challenges for organizations looking to adopt cloud-based solutions is security. This is because of the nature of the cloud infrastructure i.e., fully distributed and open, thus making it more vulnerable to threats and attacks. This environment creates incentives for potential intruders to initiate attacks targeting devices having access to data stored on the cloud. The threats due to attacks are to the integrity, confidentiality, and availability of cloud services and resources [56]. For example, a Distributed Denial of Service attack, is one that aims to prevent availability of data stored on the cloud, by choking the network bandwidth through packet flooding. Other potential attack types include IP Spoofing, Domain Naming System (DNS) Poisoning, Man in the Middle Attack, Port Scanning, etc. [50]. Cloud security is an interesting active field of study and various heuristics have evolved and been proposed. Basic security elements such as a firewall that protects the internal network and adoption of message encryption may be employed as initial lines of defense. However, a firewall may not be able to identify an attack initiated by an insider [33]. In order to meet the security challenges effectively, a dedicated Intrusion Prevention System (IPS)/ Intrusion Detection System (IDS) should be integrated within the cloud environment. IDS has become an important and irreplaceable part of the network protection system. An intrusion i.e. an attempt to compromise the availability, confidentiality, and integrity of cloud-based resources can be detected utilizing cloud based IDSs [16]. Traditional network security techniques which are not integrated within the cloud environment may not be effective in meeting the requirements of cloud security. This is due to certain limitations of traditional IDS such as incorrect classification of network anomalies as attack, low rate of detection of attacks and high false positive rate among the detected attacks [27]. Techniques in IDS such as anomaly detection and misuse detection are now relying on machine learning to increase performance effectiveness. Machine learning incorporates meta-heuristic algorithms to enhance the performance, and to identify and classify normal and unusual attacks in the network. The IDS should monitor the potential means and ends for attacks, such as network traffic and audit data in a network/ computer system, and employ different methods for detecting unauthorized activities as intrusion. Fig. 1 provides a summary of IDS in cloud environments [44]. Design of an integrated IDS was described in [19] and [54]. The primary goal of IDS is to identify each intrusion in an effective way [60]. The execution of IDS enables network administrators to detect security objective violations. These security objectives include both securing cloud resources from attacks by external sources who are attempting to get unauthorized access, as well as securing them from attacks by internal sources who are attempting to abuse their access privileges. However, the efficient and effective development of IDS is a complex problem due to meeting the twin requirements of achieving low false positive rate and high true positive rate, while consuming minimal computing resources for these purposes [62]. An IDS with a high false positive rate could potentially generate unwarranted alerts and consume significant cloud resources in response to anomalous network states which were not the result of an attempted intrusion. The application of detection methods could then result in initiation of response events within the cloud environment, which eventually cause an overload in the network. Simultaneously, achieving a high true positive rate through accurate and rapid detection of intrusion is crucial in reducing the potential damage caused by an intrusion or unauthorized access to the cloud resources. Within IDS, Host-based Intrusion Detection
Systems (HIDS) functions on data collected from a computer system, and permits analysis of activities of processes and users in the attack on a specific system. It visualizes the attempted attack’s outcome, access and observe data files directly and the process of the operating system [25]. It identifies the attacks which may not have been detected by Network-based Intrusion Detection System (NIDS), as it observes the events which are local to the computer system. Host-based Intrusion Detection and Prevention (HIDPS) consist of software involves in observation and analysis of events takes place in the computer and information system in identification and stopping harmful incidents in the system, becomes more important as it protects the computer system and its network activities [33].

IDSs use different methods to detect potential malicious activities during an intrusion. One such method is Signature-based method, which attempts to map the current set of system parameters with the previously recorded system parameters patterns which correspond to known attacks or intrusions which have occurred in the past [51]. A second method is Anomaly-based which attempts to detect attacks or intrusions using machine learning and statistics to create simulations, which are then compared with the current anomalies that may have been seen in the cloud environment [51]. The Anomaly-based method has training and testing phases. Learning of normal traffic from data takes place during the training phase, and during the testing phase, tests are performed on previously unseen data. There are two types of IDS approaches- Hybrid and Non-hybrid. Hybrid IDS is the approach which attempts to reduce the limitations of Signature-based and Anomaly-based methods through higher accuracy and detection of known and unknown threats from a large dataset, by combining different intelligent algorithms [29]. Hybrid IDS relies on the reality that it is very difficult to manipulate cyber data without detection to carry out an attack [15]. Non-hybrid IDS is the approach which relies on a single intelligent algorithm to detect potential attacks. Numerous IDS models based on statistical models, machine learning, deep learning (DL), meta-heuristic algorithms, etc. are available in the existing literature. In recent years, hybridization of any of these approaches has been used to enhance intrusion detection performance. However, a comparative review of performance of various IDS approaches, after classifying them into different approach types- Machine Learning Algorithms, Computational Intelligence Algorithms and Hybrid Meta-Heuristic Algorithms- along selected parameters is not available in the literature. This paper provides a review of existing IDS algorithms, particularly developed for the CC environment, with the objective of comparing the performance of IDS approaches along selected parameters. The recent, state of the art IDS techniques consist of both non-hybrid IDS approaches as well as hybrid approaches. The existing IDS algorithms under each approach category have been reviewed and the merits of each algorithm have been identified. In addition, the reviewed algorithms have been compared to one another, based on selected parameters. Finally, the open issues, possible future directions, and limitations of the study have been elaborated.

Riaz, A., et al., [45] conducted a brief analysis of IDS techniques presented for the cloud environment. To attain this goal, at the initial stage, the unique characteristics and limitations of all the techniques were enumerated. Next, a set of criteria were established for evaluating the IDS framework. In this work, a relative analysis of many current IDSs on different dimensions was elaborated. Lastly, the discussion of open issues and drawbacks was provided in detail. Zouhair, C. et al., in [63] presented the review of cloud infrastructure and summary of distinct intrusions in the cloud. In addition, the essential characteristics and challenges of cloud based IDS techniques were identified. Next, the researchers analyzed 24 cloud based IDS regarding their different positions, types, data sources, and detection times. Also, the strengths and limitations of various IDS, to evaluate whether they meet the security requirement of CC infrastructure or not were listed. Mthunzi, S.N. and Benkhelifa, E. in [41] identified security issues that are of catastrophic nature in the cloud environment and listed out a survey of the counter measures for cloud security with bio-inspired approaches and enumerated the advantages and limitations of the approaches. Mishra, P., et al., in [39] provided a comprehensive study of different IDSs presented for cloud infrastructure with analysis of their attack detection abilities. The researchers proposed an attack taxonomy and threat model in the cloud framework, to list out the various vulnerabilities in the cloud environment. The taxonomy of IDS techniques represented an advanced classification and provided an exhaustive literature survey of techniques using their distinct characteristics. Chattopadhyay, M., et al., in [14] examined the limitations in using machine learning techniques to detect intrusions and compared different techniques on several datasets and calculated the performance merits. The best technological solutions have been identified for various usage patterns. Sharma, S. and Kaul, A., in [53] presented a short overview about the different IDSs for a Vehicular Ad-hoc Network (VANET). Proposals were made to develop IDSs which could have potential application in VANET and VANET Cloud. This study aimed to explore open challenges, research directions in the future aspects, and leading trends in the placement of IDS in VANET. Lee, S.W., et al., in [32] focused on the Deep Learning (DL) IDS approach and
investigated how DL networks may be applied with distinct methods in various phases of the IDS, in order to achieve better results. The researchers categorized the surveyed IDS systems with respect to DL networks employed and described their major contributions. As well, in every classification, basic characteristics such as datasets, evaluated metrics, environments, and simulators were enumerated. In addition, a comparison of the results using DL IDS approach was provided, to compare the major approaches employed. Tama, B.A. and Lim, S., in [58] provided a summary of how ensemble learner may be employed in the IDS, through systematic mapping. The researchers analyzed and collected 124 high quality publications and the selected publications were later mapped to various classes like publication venues, years of publication, ensemble methods, IDS techniques, and datasets used. Furthermore, this survey analyzed and reported the experimental research of a novel classifier ensemble method for abnormality based IDSs. Shamshirband, S., et al., in [50] conducted a complete review of IDSs which used Computational Intelligence (CI) techniques in a (mobile) cloud environment. Initially, a summary of CC paradigm and service models was offered. Next, a review of the security risks in this context was provided. Earlier works related to this subject were surveyed critically, highlighting the limitations and advantages of those earlier studies. Next, a taxonomy for Intrusion Detection System was presented CI based techniques were categorized into hybrid and single approaches, for the different classifications of IDSs.

Based on the above overview of the background for this paper, research questions have been formulated, to focus on two broad approaches to intelligent IDS and on hybridization of algorithms from these two broad approaches, to determine whether such hybridization could result in enhanced performance of Intrusion Detection Systems.

### III. RESEARCH METHODOLOGY

A Systematic Literature Review (SLR) denotes evaluation of previous works on a specific set of problems from a critical perspective, with an attempt to list out all relevant studies on the basis of first principles. This study devises a structured method in locating and assembling a body of research studies on IDS in cloud environments [47]. Previous studies state that such methods have overlooked limitations, reduced chance effect and improved data validity process [21]. The SLR structure to review past work on intelligent IDS in cloud computing are presented in this section. This requires an impartial and overall layout of literature in this SLR. First, research questions are proposed as per the objectives of the survey, search query and criteria of inclusion and exclusion bias are illustrated in sections 3.1 and 3.2 with review methodology in 3.3.

**RQ1:** What are the different Computational Intelligence Algorithms- such as Bio-inspired, Swarm Intelligence, Evolutionary Computing- used in intelligent Intrusion Detection Systems in Cloud Computing?

**RQ2:** What are the different Machine Learning Algorithms used in intelligent Intrusion Detection Systems in Cloud Computing?

**RQ3:** What are the performance advantages of hybridization of Computational Intelligence Algorithms and Machine Learning Algorithms, for cloud IDS?

These questions were considered during the process of conducting SLR for intelligent IDS in cloud environments.

### A. Search Terms

Research articles in reference to keywords such as "Intelligent IDS in cloud computing", "Machine learning based intrusion detection systems in Cloud", "hybrid and non-hybrid approaches", "Bio-inspired IDS in Cloud", "Nature-inspired IDS in Cloud", "Swarm intelligence IDS in Cloud" and "Hybrid Meta-Heuristic IDS in Cloud" were searched from online sources including IEEE, Springer, Taylor & Francis, Scopus, Science direct, and Google Scholar. A total of 140 articles were collected based on these keywords, with preference given to the top research articles from renowned journals. Analysis was conducted in an orderly fashion, to initiate the review process, resulting in identification of 28 articles for Meta Analysis. This process has been summarized in Fig. 2.

### B. Inclusion and Exclusion Bias

The search was commenced with journals with an overview of the research presented in: (a) articles listed in the peer-reviewed journals; (b) published in English; (c) related to cloud IDS; (d) published between from 2016 to 2021, from databases. Duplicate articles, Conference Publications, Theoretical Research articles were removed from the initial search. Irrelevant articles were excluded further after reading the Title and Abstract. After reading the remaining full text articles, 28 articles were found to be probable sources for the review.

### C. Qualitative Review Synthesis

These twenty-eight articles were considered probable sources and their contents were streamlined in the SLR. The articles were further categorized into those that were based on machine learning models, computational intelligence (bio-inspired) algorithms and hybrid approaches for IDS. These three approaches were reviewed in terms of Accuracy, False Positive Rate and Detection Rate, as parameters [4].

![Flowchart of Article Selection](image-url)
IV. LITERATURE REVIEW

A detailed review of IDS under each category is given in the following section.

A. Machine Learning based Ids Approaches in Cloud

Machine Learning (ML) is used to address the optimal solution for complex problems which have multiple non-linear constraints, high number of dimensions and time limitations in the field of science and engineering. ML techniques have many features to resolve conflicts in classification of patterns as well as regression, optimization and estimation of functions [23]. ML provides computing input or training data to facilitate the process of learning and improving, without manual programming. The main focus of ML is to develop programs that use data in the discovery process without human intervention. ML algorithms can be classified into Supervised ML algorithms—which enable predictions of output from given data; Unsupervised ML algorithms—which enable inferences to be drawn on structures which are not obvious from unknown data; and Semi-supervised ML algorithms—which enable blending of features of both Supervised and Unsupervised ML algorithms and are mostly used to quantify the training data [8]. A detailed comparison of ML based IDS approaches reviewed in this SLR is given in Table I. A brief summary of these models follows. A novel DDoS attack detection technique in CC platform was developed in [31]. The presented model was defined by the use of an ML model called Voting Extreme Learning Machine (V-ELM). A voting scheme was developed and attack class was allotted to a sample, in case of having many votes. The performance of the V-ELM technique was validated using the NSL-KDD and the ISCX intrusion detection datasets. In [59], the researchers aimed to detect the presence of DDoS attacks in SDN. This method classified the SDN traffic as normal or attack traffic with the use of ML models integrated into Neighborhood Component Analysis (NCA). In addition, a public dataset with 23 attributes was used for experimental validation and the results demonstrated the superior performance of the proposed model with limited features. Sharma, P., et al., in [52] developed a multi-layer IDS to classify different types of attacks using the ExtraTrees, classification model and the Extreme Learning Machine (ELM) model was employed for the detection of individual attacks.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Algorithm Used</th>
<th>Objectives</th>
<th>Accuracy</th>
<th>FPR</th>
<th>DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kashwah et al. 2020[1]</td>
<td>V-ELM</td>
<td>To detect DDoS attacks in cloud</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Tonkal et al. 2021[59]</td>
<td>NCA+ML models</td>
<td>To detect DDoS attacks in SDN</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Sharma et al. 2019 [52]</td>
<td>ExtraTrees + ELM + Softmax</td>
<td>To classify many attacks</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Lopez et al. 2019 [35]</td>
<td>DT + RF + DNN</td>
<td>To detect DDoS attacks</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Sambangi and Gondi 2020 [48]</td>
<td>MLR model</td>
<td>To detect DDoS attacks</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Gumaate and Shinde 2020[26]</td>
<td>Apache Spark</td>
<td>To identify DDoS attacks in OpenStack-based Private Cloud</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Ferrag et al. 2021 [22]</td>
<td>CNN + DNN + RNN</td>
<td>To detect intrusion in agricultural sector</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Kim et al. 2020[30]</td>
<td>CNN + RNN</td>
<td>To detect intrusion using DL models</td>
<td>Low</td>
<td>-</td>
<td>High</td>
</tr>
<tr>
<td>Aborujilah and Musa 2017[2]</td>
<td>Covariance matrix</td>
<td>To detect DDoS HTTP Attacks in cloud</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Shen et al. 2020 [55]</td>
<td>MKELM</td>
<td>To detect intrusions</td>
<td>High</td>
<td>High</td>
<td>-</td>
</tr>
<tr>
<td>Somasundaram 2021[57]</td>
<td>Resource Scaling</td>
<td>To mitigate DDoS attack</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>
Shen, Y., et al., in [55] proposed a Mixed Kernel Extreme Learning Machine (MKELM) method integrating the ReliefF algorithm with nature inspired algorithms, for IDS. The MKELMs were developed to predict attacks, with the ReliefF algorithm providing inputs to the MKELM for selecting a suitable feature. The nature inspired algorithm determined the fitness function on the basis of kernel alignment, which was then used to build an optimum composite kernel in the MKELM. In [57] a novel approach was presented for evaluating resource consumption through ‘scaling down’ the resource i.e., through an improvement of the ‘scale inside out’ approaches. The presented approach utilized two modules-authentication model and elastic load balancing- to detect and mitigate DDoS attacks.

B. Computational Intelligence based IDS Approaches in Cloud

Computational Intelligence approaches reviewed in this study include Bio-inspired algorithms, Evolutionary Computation algorithms and Swarm Intelligence algorithms. A detailed comparison of Computational Intelligence approaches reviewed in this SLR is given in Table II. A brief summary of these models follows. Bio-inspired algorithms aim to mimic natural biological patterns and behavior to develop novel ways to solve complex optimization problems [17]. Bio-inspired algorithms have been used to address major problems due to their features of adaptability, to attempt achievement of optimal solutions in cloud computing [12]. Bio inspired algorithms have been previously used to meet requirements of the cloud environment such as load balancing, provisioning of resources, and performance improvements, and may prove to be useful for adoption in IDS as well. Comparison of Bio-inspired algorithms for purposes such as sentiment analysis was described in [61]. Evolutionary Computation algorithms have been derived from biological evolution, and essentially aim ‘to evolve’ from an initial set of solutions to arrive at a best fit solution [20]. It is an approach in which different solutions adapt to different environments through processes similar to natural selection and breeding, so that only those which are truly fit and effective will survive the environment. Those which are not effective will not survive, but extreme conditions may result in mutations, similar to the biological analogy. Through iterations of this process, the best fitting solution to the problem is determined [28]. The population of potential solutions is first initialized randomly and the selection of solutions with the best fit through either survival or mutation mechanisms are devised; the rest are terminated. Evolutionary Computation has also been defined as the probable search performed for test data to be executed for a specific number of times by optimization algorithm based on Charles Darwin’s theory of evolution [18]. It works on a potential solution with a permissible value for the variables coded for optimization problems, and is especially known for robustness and suited for complex domains of large numbers of variables [34]. The initialization of a population of solutions for a problem is first set at random, fitness of each individual solution in the population is calculated, and the algorithm is run until optimization as initially defined is achieved or any of the defined stop conditions are achieved. The results are graded from very poor to good. Then, selection of pairs of individual solutions from the population results in recombination, with the resulting progeny subjected to mutation to maintain diversity. The resulting new generation solutions are evaluated for fitness, and a reinsertion process replaces the older generation solutions with fitness values which are lower than those of the new generation [62]. Swarm Intelligence algorithms emerged from observation of the behavior of social organisms, such as ants, wasps, bees and termites. Swarm Intelligence algorithms aim to mimic natural swarm behavior of organisms to forage for food or resources, to construct nests and to move in their environments. Swarm Intelligence algorithms follow five principles- proximity, quality, diverse response, stability and adaptability. Each possible solution to a problem is analogous to an organism in the swarm and has autonomy in behavior; the resulting emergence of self-organization in the swarm of solutions leads to adaptability to address the problem. The basis of self-organization includes amplification (positive feedback with the use of more resources) as well as stabilization (negative feedback to achieve counter balancing stability), random errors and multiple iterations of interaction between solutions in the swarm. Swarm Intelligence algorithms begin with in initialization phase to set the values of parameters, and continue to execute until defined stop conditions are achieved or stop is executed. Fitness function is evaluated for each solution and the Swarm Intelligence algorithm is updated mathematically based on the results. The fitness functions for each solution or search agent in the swarm leads to proposal of taxonomy and identification of the best fit solution to the problem. Swarm Intelligence algorithms have been used in optimization problems such as Agent Swarm Optimization (ASO) with the coexistence of different agents and their interaction, to ensure problem specificity, facilitation for testing and application to real-life problems [13]. One of the concepts significantly used in cloud computing is virtualization, as it enables higher resource utilization and lower operating costs. During virtualization, Computational Intelligence based optimization algorithms can play a vital role during the process of Virtual Machine Placement (VMP) scheduling. Such algorithms may be adopted for the purpose of IDS as well. Computational Intelligence algorithms are divided into two categories- Single-objective optimization algorithms and Multi-objective optimization algorithms. Examples for Single-objective algorithms include Ant Colony Optimization, Crow Search, Cuckoo Search, Fire Fly, Genetic, Grey Wolf Optimizer, Imperialist Competitive, Memetic, Particle Swarm Optimization, Simulated Annealing, and Whale Optimization Algorithm. Examples for Multi-objective algorithms include Biogeography-based Optimization, Krill Herd, Multi-Objective Evolutionary Algorithm, and Non Dominated Sorting Genetic Algorithm [37]. Taxonomy of Computational Intelligence intrusion detection techniques in mobile cloud computing environments was described in [49].
C. Review of Hybrid Meta-Heuristic IDS Approaches in Cloud

Hybridization combines the benefits of different algorithms to form a hybrid algorithm with increased profitable synergy and minimization of disadvantages from the combination. This usually results in improved performance in terms of parameters such as computational speed, storage space and accuracy in detection of attacks. Hybrid algorithms may be classified into two types—Unified purpose hybrid algorithms, where the component algorithms are used to solve the same problem with each used at different stages; and Multiple purpose hybrid algorithms, where one primary component algorithm is used to solve the problem, while other component algorithms are used to alter the parameters of the primary algorithm. Hybrid algorithms may also be categorized as collaborative hybrid, involving a combination of two or more component algorithms run sequentially, or in parallel. These sequential or parallel runs can either comprise a single stage or have multiple stages. Another type of Hybrid algorithm is integrative hybrid, where one algorithm is considered as a subordinate embedded into a master algorithm. It involves incorporation of operators manipulated by the subordinate algorithm into the master algorithm. The process of hybridization creates additional components but usually increases computational speed [33]. Two different algorithms could be hybridized by optimizing the parameters of both the algorithms to produce the best result. Different hybrid combinations are created and tested, in order to obtain overall best performance through experimentation. Because of the limitations of any standalone ML/DL method or Computational Intelligence algorithm, accomplishing optimum intrusion detection performance in a cloud environment requires hybridization. Since every approach has its merits and demerits, in this view, several authors have integrated the merits of two or more techniques in various aspects. For designing an effective hybrid IDS technique, the concept of mixing algorithms is essential. In this section, the hybrid IDS approaches developed for cloud environments are reviewed and a comparison is made in Table III.

### TABLE II. COMPUTATIONAL INTELLIGENCE BASED IDS IN CLOUD

<table>
<thead>
<tr>
<th>Paper</th>
<th>Algorithm Used</th>
<th>Objectives</th>
<th>Accuracy</th>
<th>FPR</th>
<th>DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reddy et al., 2021</td>
<td>Crow Search Algorithm</td>
<td>To detect DDoS attacks</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Ahmad et al., 2018</td>
<td>Dendritic Cell Algorithm</td>
<td>To detect co-residency attack</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prathyusha et al., 2021</td>
<td>Artificial immune system</td>
<td>To mitigate DDoS attacks</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Alharbi et al., 2021</td>
<td>Local-Global best BAT</td>
<td>To detect botnet attacks</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alamiedy et al., 2020</td>
<td>Grey wolf optimization</td>
<td>To detect anomaly-based intrusions.</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Alharbi et al., 2021</td>
<td>Cuckoo Algorithm</td>
<td>To achieve intrusion detection</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Niemiec et al., 2021</td>
<td>Multivariable heuristic</td>
<td>To detect intrusions using flag and entropy values</td>
<td>High</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE III. HYBRID METAHEURISTIC INTELLIGENT IDS APPROACHES IN CLOUD

<table>
<thead>
<tr>
<th>Paper</th>
<th>Algorithm Used</th>
<th>Objectives</th>
<th>Accuracy</th>
<th>FPR</th>
<th>DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meghanian et al., 2020</td>
<td>ANN+GOA</td>
<td>To detect network intrusion patterns</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Ali et al., 2018</td>
<td>ABC+BPNN</td>
<td>To detect DDoS attacks in cloud</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Oasaniye et al., 2016</td>
<td>Ensemble-based multi-filter FS</td>
<td>To design ensemble of four filter approaches</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghanem et al., 2020</td>
<td>ABC+DA+MLP</td>
<td>To detect intrusions by optimal training of MLP</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghosh et al., 2021</td>
<td>Modified Firefly algorithm</td>
<td>To design IDS using feature selection approach</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Mazini et al., 2019</td>
<td>ABC+AdaBoost</td>
<td>To develop a A-NIDS technique</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alharbi et al., 2021</td>
<td>Enhanced BA</td>
<td>To detect botnets in IIoT</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Bojović et al., 2019</td>
<td>Feature-based and volume-based detection</td>
<td>To design DDoS detection model</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Lv et al., 2020</td>
<td>KPCA-DEGSA-HKELM</td>
<td>To design IDS based on attack signatures</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Aaslhi-Shahri et al., 2016</td>
<td>SVM+GA</td>
<td>To identify anomalies using hybrid method</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>
In [27], a hybrid approach was presented which used an Artificial Neural Network (ANN) approach as a learning approach while a Swarm Intelligence algorithm- Grasshopper Optimization Algorithm- was used to reduce IDS errors. Ali et al. [28] presented a hybrid approach using a combination of Ant Colony Optimization (ACO) and Back Propagation Neural Network (BPNN). This hybrid approach was employed to detect DDoS attacks in the CC environment. Osanaiye, O., et al., in [43] proposed an ensemble based multifilter feature selection approach which integrated the output of 4 filter approaches to achieve optimal selection. The presented approach has been evaluated using standard datasets such as NSL-KDD. Ghanem, et al., [23] proposed a novel binary classification method for detecting intrusions, depending on the hybridization of Artificial Bee Colony (ABC) algorithm and Dragonfly algorithm to train an ANN and thereby increase the classification performance for non-malicious and malicious traffic in the network. The hybrid approach sets the initial parameters and appropriate weights for the ABC and Dragonfly algorithms. Ghosh, P., et al., in [24], proposed an IDS which provides security based on the concept of feature selection using Modified Firefly algorithm. The developed hybrid approach was evaluated on the NSL-KDD dataset and was found to consume lesser storage space due to the decreased number of dimensions from feature selection, and also require lower training time, thereby improving classification performance. A meta-heuristic algorithm based feature selection and recurrent neural network for DoS attack detection was proposed in [46]. Mazini, M., et al., in [38] proposed a novel hybrid approach for an Anomaly Network IDS, using ABC and AdaBoost algorithms to obtain higher detection rate and lower false positive rate. The ABC algorithm was used for feature selection and the AdaBoost algorithm was used for evaluation and classification of features. In [42], a novel multi-variable heuristic IDS was proposed, depending on distinct kinds of flags and values of entropy. The organizations distributed the data to improve the efficacy of IDS. Alharbi, A., in [6] proposed a Local Global Best Bat Algorithm with Neural Network (LGBB-NN) for selecting hyper parameters and feature subsets for effective detection of botnet attacks. The presented hybrid approach adapted the inertia weights from the LGBB algorithm to update the parameters of the solution in the swarm. In order to address the swarm diversity for the solutions, a Gaussian distribution was employed during the initialization of the population. Bojović, P.D., et al., in [13] introduced a hybrid approach for detecting DDoS attacks, which combined volume and feature based detections. This method was dependent on an exponential moving average approach to make decisions, used on entropy values and packet number time sequences. Lv, L., Wang, et al., in [36] presented an approach for detecting several attacks on the basis of Hybrid Kernel Extreme Learning Machine (HKELM) model. This hybrid approach integrated the Gravitational Search Algorithm (GSA) and Differential Evolution (DE) algorithm to optimize the parameter of HKELM, which in turn enhanced its local and global optimization capabilities at the time of predictive attack. A Kernel Principal Component Analysis (KPCA) method was presented for feature selection and reduction of number dimensions for the IDS. Aslahi-Shahri, B.M., in [11], presented a hybrid approach of Support Vector Machine (SVM) and Genetic Algorithm (GA) for execution of IDS. The presented hybrid approach was used to decreasing the number of features from forty-five to ten. The features were classified on the basis of priority, using the Genetic Algorithm.

V. Meta Analysis

Statistical analysis was performed for the above findings, using three performance metrics- Accuracy, False Positive Rate (FPR) and Detection Rate (DR), by classifying the performance of each algorithm as High, Medium or Low. Fig. 3 shows the ML based IDS approaches used in cloud environment, with the performance of the algorithms specified in terms of the three performance parameters. Algorithms employed in IDS must aim to decrease the FPR, since it represents the number of false alerts or alarms generated by the IDS, which may have a detrimental impact on cloud performance due to increase in computation time and storage space requirements of the IDS due to situations which are not really intrusions [9]. For this reason, having Low FPR is the most significant and important parameter for an effective IDS in the cloud environment. At the same time, algorithms must aim to increase Accuracy and Detection Rate, for obvious reasons.

Fig. 3 shows the comparison of ML based IDSs. First, 7 out of 11 ML based algorithms showed High Accuracy. Secondly, 4 out of the 11 ML based algorithms evaluated showed Low FPR. Thirdly, 4 out of the 11 ML based algorithms showed High Detection Rate. Incidence Rates for the desirable states of these parameters can be computed accordingly.

Fig. 4 shows the comparison of CI based IDSs. First, all 7 CI based algorithms evaluated showed High Accuracy. Secondly, 4 out of 7 CI based algorithms showed Low FPR. Thirdly, 5 out of 7 CI based algorithms showed High Detection Rate. Incidence Rates for the desirable states of these parameters can be computed accordingly.

Fig. 5 shows the comparison of Hybrid Meta-heuristic intelligent IDS approaches. First, 8 out of 10 hybrid approaches showed High Accuracy. Secondly, 8 out of 10 hybrid approaches showed Low FPR. Thirdly, 6 out of 10 hybrid approaches showed High Detection Rate. Incidence Rates for the desirable states of these parameters can be computed accordingly.

Among the three types of IDS- ML based, CI based and Hybrid Meta-heuristic- the Hybrid Meta-heuristic based IDS appear to have the highest overall incidence and scope for achieving the combination of High Accuracy, Low FPR and High Detection Rate.
Fig. 3. Performance of Machine Learning based IDSs in Cloud.

Fig. 4. Performance of Computational Intelligence based IDSs in Cloud.

Fig. 5. Performance of Hybrid Meta-heuristic Intelligent IDS Approaches.
Computation of incidence rates for desirable states of the three parameters (High Accuracy, Low FPR and High Detection Rate) across algorithms within each IDS was carried out, and using these incidence rates and associated weights assigned to parameters, a Weighted Score computed for each type of IDS. The results are presented in Table IV and plotted in Fig. 6. While the performance of Hybrid Meta-heuristic intelligent algorithms in terms of Detection Rate and Accuracy is marginally lower than that seen in Computational Intelligence algorithms, the performance is higher in terms of False Positive Rate, which is the parameter which has been assigned higher weight in determining the overall weighted performance score across the three parameters. From the results, the significance of potential adoption of Hybrid Meta-heuristic intelligent algorithms in IDS, to achieve the desirable performance states of High Accuracy, Low FPR and High Detection Rate is apparent.

VI. OPEN CHALLENGES

This section discusses the major challenges that exist in the reviewed IDS models in the cloud environment.

- Existing IDS in the cloud environment only detect security-based events and are unable to block it. If many such events occur simultaneously, they could overwhelm the system. If the number of attacks or data breaches are very high in number, a valid threat may not be detected on time. Existing studies have not focused much on intrusion prevention.

- One of the major challenges in network-based IDS is that it cannot detect encrypted traffic without interception and decryption. In some organizations, the IDS decrypts traffic as it flows into the network. However, it will not be able to decrypt the traffic if the attacker uses a key to initiate encryption and decrypts the host.

- A high False Positive Rate is a significant and material open challenge to the execution of successful IDS. A system that generates high number of false alerts can potentially create serious business challenges for organizations, which may outweigh the benefits of even implementing the IDS itself.

- From the reviewed papers, it is evident that most of the IDS models have been developed using Machine Learning (ML) models and only a few works have concentrated on Deep Learning (DL) models. In addition, the experimental validation of the reviewed approaches should be extended to larger and real time datasets.

- In addition, the high number of dimensions of the IDS data-set should be carefully reviewed to reduce and discard irrelevant, and repetitive features. Feature selection techniques should be highly focused on minimizing the count of features in the data-set, to retain only essential features. Feature reduction and cluster techniques can be designed to boost the detection performance in large scale IDS datasets.
• Furthermore, the reviewed approaches are not designed for multi-objective formulation or multiple attacks detection, which needs to be further explored. Lastly, the offline IDS process needs to be extended to real time intrusion detection.

• The hybrid IDS technique should incorporate the combination of improved meta-heuristic optimization algorithms, to utilize their benefits.

• Most of the studies have focused on DoS and DDoS attacks. In future, IDS techniques should be designed to handle new and emerging types of attacks. The CC makes use of wireless networks for communication with the user system. Owing to few features of wireless networks such as resource limitations, mobility, and restricted bandwidth, issues related to network management and security need to be addressed. Based on the reviewer’s works, hybrid methods may be employed for the detection of anomaly and signature based IDS in cloud environments.

• The choice of appraising classification model and feature selection is a major challenging issue in IDS. Therefore, it is important to design a rapid and precise IDS with minimal false positives and maximum true positives in a cloud environment.

• The choice of parameters has a significant influence in comparative analysis of various IDS approaches. Therefore, it is important to take into account any other parameter apart from the 3 parameters used in this study, which is considered significant and material for a particular IDS.

• On the other hand, the description of the implementation on setup can be a serious challenging problem for the cloud environment to accomplish security. In some cases, the models developed to improve the outcome of the IDS might be ineffective, resulting in false alarms owing to the inappropriate choice of evaluation criteria. The data integrity and security of information handled by cloud providers and probable susceptibilities which may result in data breaches need to be addressed in future. Based on the open issues and possible future directions, an effective IDS can be designed with respect to the consideration of the dimensions and features of the cloud environment.

VII. LIMITATIONS OF THE STUDY

The authors utilize Google scholar as a reliable electronic database that recommends highly relevant and effective studies depending upon the previous empirical works. But it could not be guaranteed that all selection is applicable studies. There is a chance that few significant works are not considered in the article selection process. Although this literature will provide an overall understanding utilization of an intelligent Intrusion Detection System in cloud environments and could be applied practically, this review article and its findings are theoretical only, which is one of the limitations of this study as it could not be reproduced in terms of practical implications. There is also a limitation arising due to the 3 performance parameters selected, since some of the IDS approaches may theoretically perform better if other performance parameters were to be considered. Practical implementations are required to prove the benefits of the study.

VIII. CONCLUSION

This paper conducted SLR and Meta Analysis in order to evaluate the efficacy of Hybrid Meta-heuristic based IDSs in the cloud environment along three performance parameters, compared to two other types of IDS- ML based and CI based. The significance of various recent studies was summarized, and the performance of different algorithms/approaches within each type of IDS was reviewed along the parameters of Accuracy, FPR and Detection Rate. The reviewed approaches were briefly explained, along with the merits and demerits. The open research issues which have to be addressed in future study have been discussed. The highlight of the study is the significance of potential adoption of Hybrid Meta-heuristic intelligent algorithms in IDS, to achieve High Accuracy, Low FPR and High Detection Rate. We strongly believe the outcome of this review study will be helpful to design new hybrid IDS approaches for cloud environments, particularly utilizing Meta-heuristic techniques.

REFERENCES


