IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 26, 2021, accepted November 8, 2021, date of publication November 22, 2021,
date of current version December 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3129775

Research Trends in Network-Based Intrusion Detection Systems: A Review

SATISH KUMAR, SUNANDA GUPTA, AND SAKSHI ARORA
School of Computer Science and Engineering, Shri Mata Vaishno Devi University, Katra, Jammu 182320, India
Corresponding author: Satish Kumar (krsk23@gmail.com)

ABSTRACT Network threats and hazards are evolving at a high-speed rate in recent years. Many mechanisms (such as firewalls, anti-virus, anti-malware, and spam filters) are being used as security tools to protect networks. An intrusion detection system (IDS) is also an effective and powerful network security system to detect unauthorized and abnormal network traffic flow. This article presents a review of the research trends in network-based intrusion detection systems (NIDS), their approaches, and the most common datasets used to evaluate IDS Models. The analysis presented in this paper is based on the number of citations acquired by an article published, the total count of articles published related to intrusion detection in a year, and most cited research articles related to the intrusion detection system in journals and conferences separately. Based on the published articles in the intrusion detection field for the last 15 years, this article also discusses the state-of-the-arts of NIDS, commonly used NIDS, citation-based analysis of benchmark datasets, and NIDS techniques used for intrusion detection. A citation and publication-based comparative analysis to quantify the popularity of various approaches are also presented in this paper. The study in this article may be helpful to the novices and researchers interested in evaluating research trends in NIDS and their related applications.

INDEX TERMS Citation, machine learning, bio-inspired, intrusion detection system, NIDS, datasets.

I. INTRODUCTION

Today's era is of information and communication, and the numbers of host/terminal are continuously increasing in the scenario of computer networking. Vulnerabilities in security systems and unauthorized access to information systems are also growing tremendously. Many techniques, namely firewalls, access control, anti-virus, anti-malware software, application security, behavioral analytic, data loss prevention, distributed denial of service (DDoS) prevention, and network segmentation are commonly used in the computer world to promote internet security mechanisms due to their capabilities of content filtering, blocking data outflow, and alerting and preventing malicious activities. Firewalls and spam filters are generally used with simple rules-based algorithms to allow and denial of the protocols, port, or IP addresses. But the drawback of these firewalls and filters is that sometimes they are unable to control complex attacks of DoS (denial of service) types, and they are also not capable of making the differences between 'good traffic' and 'bad traffic'. An intrusion detection system (IDS) with anti-virus has a significant impact on computer network security mechanisms that provides a more prominent scenario for protecting a computer network from the unauthenticated access. In the perspective of information systems, intrusion refers to any attempt that compromises the integrity, availability, confidentiality, or bypasses the security mechanism in a computer or a network [1].

According to the National Institute of Standard and Technology (NIST), intrusion detection is the process of monitoring events occurring in a computer system or a network and analyze these events for a sign of intrusions. The monitoring processes can be accomplished with software or hardware to secure the system from malicious activity. This also protect integrated policies like firewall port configuration, data encryption, secure sockets layer (SSL) authentication, etc. that are being violated. The IDS performs the intrusion detection process to secure a computer or network. It provides a more prominent scenario for protecting a computer network from the unauthenticated access.

The IDS can be categorized into three categories on the installation basis in the system- Host-based IDS (HIDS), Network-based IDS (NIDS), and Hybrid IDS. HIDS are deployed on a single host. In HIDS, attacks are detected

The associate editor coordinating the review of this manuscript and approving it for publication was Pedro R. M. Inácio.

from a single computer system, and the essential files of the operating system are analyzed. Hence, these types of attacks are usually easy to detect except for some in-filtered malware which is very hard to detect. In a NIDS, malicious information is detected from the diverse interconnection of computers, and NIDS is deployed on routers or switches in a network. Whereas, hybrid IDS can be deployed on hosts as well as on the network. The primary goal of NIDS is to identify malicious or threatened logging information and report to the network manager about this malicious information.

An intrusion detection system usually does not prevent the system from intrusion attack; rather than it merely generates an alarm after detecting an attack in the system in real-time or before the arrival of the attack on the target. It is also equally vital to cause notice of an attack after the happening of that attack in the system because an IDS maintains and updates an intrusion profile in the log. The operating system must also uphold various activities that require excess disc space and central processing unit (CPU) resources for analysis of logs. Managing the logs formats and comparing these formats with identified attack patterns according to security violation issues is also a big challenge in the IDS [2].

The literature regarding intrusion detection systems for a network does not provide the research trend, popularity of the datasets used, evaluation of a NIDS model, and the popularity of different intrusion detection approaches. The research articles taken for research review are usually not based on any qualitative measure. They are chosen arbitrarily. But we took citation as a measure to quantify the popularity and research trend in research articles.

Citation is a measure to identify the popularity of a research article. According to Smith [3], citation is the number of times that the other authors mention a research article in his/her work/s. The citation of an article is a quantitative and qualitative measure to recognize the popularity of a research article and an institution. Citation also provides the trend of research in a specific field.

We want to determine the research trend and popularity in intrusion detection based on various approaches and methodologies. Citation belonging to a published article is used to explore the research trend regarding that article. Citation is a valuable and popular scale to measure the research trend in a field [3]. There are various search engines like Google Scholar, Web Science, Microsoft Academic, Semantic Scholar that record citations of an article. Citation and article publication count regarding an article on a research topic is considered a research trend and popularity related to that particular research topic. Citation to an article is the total number of references that the other articles include the references to an article into their work. But, no search engine or database provides research trends on a search field/topic along with the different used approaches. The current article focuses on research trends in the field of IDS, its related techniques, datasets, total publications, and other citation-related analysis. We considered the related articles to IDS from 2005 to 2020. The major contribution of this article is outlined as follows

1) A comparison of several popular IDS, which are being used commercially for network security.
2) An analysis of the popularity of various datasets to evaluate a NIDS.
3) Find the popularity of different approaches and methodologies used in the intrusion detection system
4) An analysis of most cited published articles by separately tabulating them under conferences and journals fields
5) An analysis of various performance metrics used to evaluate an intrusion detection system

In the past decades, academic search engines and bibliographic databases (ASEBDs) comparison has been widely investigated [4], [5]. A comparative analysis of various academic databases and search engines have also been shown in [6] and [7]. Microsoft Academic Search (MAS) follows semantic search in which the search engine does not only match the keywords to content; instead focuses on their meaning with a broader scope and coverage as compared to Web of Science, Scopus and Google Scholar [4], [5], [8]. It helps searchers by providing some entries and interesting topics when they are unsure about searching string. MAS also supports searching based on journals, conferences, institutions, and authors in different fields for finding the best search results. The total number of publications records by Microsoft Academic [9] is 247,389,875, 261,445,825 authors, 743,427 topics, 4,523 Conferences, 48,974 Journals, and 25,811 Institutions. The total number of estimated citation pairs is 2,390,820,943. In the same vein, a total of 36,765 publications with 592,675 citations are observed in the IDS field. So, Microsoft Academic [9] has been used for taking the records of citations that use a search string to achieve the goal of this paper.

The rest of the paper is organized accordingly. Section II deals with related work in the field of research trends in intrusion detection. Section III discusses the method applied for finding the citation and related articles and other corresponding reviews. Network intrusion detection system, its modules, widespread causes of intrusion in a network, some popular NIDS, and their analysis are presented in Section IV. Section V shows different benchmark datasets and their citation analysis-based records. A study related to various methodologies used for intrusion detection in a network is given in Section VI. Section VII is regarding the performance metrics used to evaluate a network intrusion detection model. Section VIII explores the discussion on the result of our present study. Finally, Section IX presents the conclusion.

For clarity, we explain some abbreviations and their corresponding acronym commonly used in this paper. In the KDD'99 dataset, KDD stands for Knowledge Discovery in Database, 1999. NSL in NSL-KDD dataset stands for Network Security Laboratory. ISCX is the acronym for Information Security Centre of Excellence which is one of the leading institutions in the area of information and

communication security, in collaboration with the Atlantic Canada Opportunities Agency (ACOA). CIDDS is abbreviated for Coburg Intrusion Detection Data Sets. UNSW-NB15 stands for the University of New South Wales Network-Based dataset, 2015. SSENet stands for self-supervised scale equivalent network Dataset. KNN means K-Nearest Neighbors which is a supervised learning algorithm. SVM stands for support vector machine which is also supervised learning that is used for classification as well as regression problems. PCA is denoted for principal component analysis.

II. RELATED WORK

During the last decade, several surveys of intrusion detection have been conducted. One of the earliest was presented by Bishop [10] about trends in vulnerabilities analysis and intrusion detection. Trends in intrusion detection are infrastructure-based protocols and techniques required to design and develop intrusion detection systems.

Another popular survey by Kabiri and Ghorbani [11] presented trends in IDS and also analyzed some problems regarding intrusion detection. Traditional IDS faces challenges like, time consumption, log-file updating, statistical and rule-based analysis, and accuracy.

Zamani and Movahedi [12] presented a review article based on some influential algorithms based on machine learning approaches used in intrusion detection. Zamani explored that using a machine learning approach for intrusion detection enables a high detection rate and low false-positive rate with the capabilities of quick adaptation toward changing intrusive behavior. The analyzed algorithms in this review paper have been categorized into artificial intelligence (AI) and computational intelligence bases.

Agrawal and Agrawal [13] surveyed various data mining techniques for intrusion detection. Various machine learning techniques, individually or in hybrid form have been widely used not only in the field of clustering or classification but also for reducing the dimensionality and feature selection of IDS.

Ahmed *et al.* [14] presented the challenges regarding the datasets which are being used for IDS Model and categories of IDS namely; classification, statistical, information theory, and clustering were also explored.

In current IDS approaches, the statistical method is extended with new methods based on bioinspired approaches. These methods are mainly based on the evolutionary theory or swarm intelligence method [15]. For finding the suitable and best-fit selection of bio-inspired algorithms, various characteristics like Convergence, Intensification, diversification, CPU time, etc. are to be analyzed.

Gendreau and Moorman [16] represented a survey of Intrusion Detection Systems towards an End to End Secure Internet of Things (IoT) and this survey of the IDS use the most recent ideas and methods to propose the present IoT. To understand and illustrate IDS platform differences and the current research trend towards a universal, cross-platform distributed approach has been taken into consideration.

Hamid *et al.* [17] provided a review of the benchmark datasets available for researchers in the field of intrusion detection that are used to train and test their models. The review on various datasets namely; DARPA 98, KDD'99, NSL-KDD, UNM-Dataset, UNSW-NW15, Caida DDoS (Caida Distributed denial of Service) Dataset, ADFA-WD (Australian Defense Force Academy Window Dataset), provided the details of classes, attributes, and instances.

Most recently, Mishra *et al.* [18] also proposed a detailed investigation and analysis using machine learning approaches for intrusion detection. This survey depends on the categorization of the classifiers into four categories viz-a-viz single classifiers with all features in the dataset, the single classifier with selected features of the dataset, multiple classifiers with all features of the dataset, multiple classifiers with selected features of the dataset. This analysis also reveals that a well-performing intrusion detection approach for one type of attack, may not perform well for the other types of attacks.

All the literature discussed so far, does not focus on the research trend and popularity in the NIDS based on some quantitative measure. However, in this article, we analyze various commercially used IDS, the popularity of various benchmark datasets, and the recent trends in the used approaches in intrusion detection. The analysis performed in the article is based on quantitative measures instead of qualitative measures.

III. METHODS

Researchers are more attracted to articles that have a high citation. So, we have taken citations as metrics that provide a standard and validity of a research topic/journal publications in a research area. The string-based searching in Section III took research articles from the year 2005 to 2020.

Following the network intrusion detection model, keywords related to intrusion detection systems, anomaly detection bio-inspired algorithms are used in the 'Microsoft Academic' advance search. At the same time, terminologies related to the intrusion detection system, datasets, methodologies, and issues are utilized. The searching string for datasets and approaches with their sub-classes are tabulated in Table 1. The description for various datasets are described in the Table 3

The publications for various approaches implemented in intrusion detection systems are analyzed based on the searching strings related to intrusion detection systems using filters searched between the years 2005 to 2020. The filters are used as 'intrusion detection system', and, 'oldest first' citations for searching. The searching strings for the performance metrics are also filtered by top topics as 'False positive rate', 'True positive rate', and 'F1 score'. The filters used are the same for both; for the complete publication analysis and the citation analysis of different approaches used in intrusion detection. This analysis is based on the article publication records from the year 2005 to 2020. We are making published article records till 20th December 2020 to avoid day-by-day citation variations.

TABLE 1. The searching string on which research articles from the year 2005 to 2020 are chosen.

Searching base	Searched strings
Datasets	"KDD Cup'99" + "intrusion detection", "NSL-KDD" + "intrusion detection", "Kyoto 2006" + "intrusion detection", "UNSW-NB15" + "intrusion detection", "SSENet" + "intrusion detection", "ISCX" + "intrusion detection", and "CIDDS" + "intrusion detection".
Approaches used in IDS	"Statistical based" + "Intrusion detection", "Knowledge based" + "Intrusion detection", "Machine Learning Based" + "intrusion detection", and "Bio-inspired based" + "Intrusion detection" **Statistical-based NIDS approaches** "Univariate" + "intrusion detection", "Multivariate" + "intrusion detection", "Time series" + "intrusion detection" **Knowledge-based NIDS approaches** "finite state machine" + "intrusion detection" ‖ "FSM" + "intrusion detection", "Description Language" + "Intrusion detection", "Expert System" + "Intrusion detection" **Machine learning-based NIDS approaches** "Linear regression" + "intrusion detection", "Logistic regression" + "intrusion detection", "Decision tree" + "intrusion detection", "K-mean" + "intrusion detection", "Neural network" + "intrusion detection", "KNN" + "intrusion detection", "SVM" + "intrusion detection" ‖ "support vector machine" + "intrusion detection", "Random forest" + "intrusion detection", "Bayesian Network" + "intrusion detection", "Markov Model" + "intrusion detection", "Fuzzy Logic" + "intrusion detection", "Principal component analysis" + "intrusion detection" ‖ "PCA" + "intrusion detection", "AdaBOOST" + "intrusion detection", "Gradient BOOST" + "intrusion detection", "clustering" + "intrusion detection" ‖ "outlier" + "intrusion detection" **Bio-inspired-based NIDS approaches** "Swarm" + "intrusion detection", "ecology-based" + "intrusion detection", "Evolutionary Algorithm" + "intrusion detection", "Genetic Programming" + "intrusion detection", "Genetic Algorithm" + "intrusion detection", "evolution strategy" + "intrusion detection", "ant colony" + "intrusion detection", "partical swarm" + "intrusion detection", "bee colony" + "intrusion detection", "fish swarm" + "intrusion detection", "firefly" + "intrusion detection", "bacterial foraging" + "intrusion detection"
Performance measurements	"Confusion matrix", "Receiver operating characteristic", "Confusion matrix" && "intrusion detection system", "Receiver operating characteristic" && "intrusion detection system", "misclassification rate" && "intrusion detection system", "Accuracy" && "intrusion detection system", "True positive rate" && "intrusion detection system" ‖ "recall" && "intrusion detection system" ‖ "sensitivity" && "intrusion detection system", "true negative rate" && "intrusion detection system" ‖ "Specificity" && "intrusion detection system", "Precision" && "intrusion detection system", "false positive rate" && "intrusion detection system", "Prevalence" && "intrusion detection system" , "F-Score" && "intrusion detection system"

IV. NETWORK INTRUSION DETECTION

The concern about increasing security problems has been expressed by James P. Anderson in a paper [19], published in 1972. After that, in 1980, he outlined an audit base procedure for automated intrusion detection and monitoring processes for hosts [20]. From 1980 to 1990, the US government invested funds for many projects like network audit director and intrusion reporter (NADIR), Haystack, Multics intrusion detection and alerting system (MIDAS), and Discovery, etc. [21].

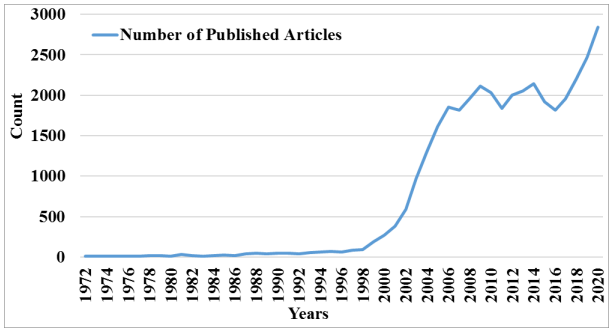

FIGURE 1. Research publications on intrusion detection systems since year 1972 to 2020.

Zuech *et al.* [22] explored that a NIDS helps the forensic process to identify the footprint of breaches. Attacks are conveyed from one computer to another computer in a network through routers and switches. An NIDS observers network traffic data in the OSI layer 3 (network layer) at the routers or switches. Based on pattern-matching of the network traffic data, the NIDS can be further categorized into Anomaly (Unknown)-Based or Misuse (Known) Based IDS. In anomaly detection, pattern base examination of traffic flow is implemented and deviation from normal pattern behavior leads to the inference of intrusive information. On the other hand, parametric examination of features and known signature for an attack is used to compare with a predefined set of rules for the detection of unauthorized action in misuse detection.

A year-wise analysis of the articles published regarding intrusion detection is shown graphically in Figure 1 from 1972 to 20th December 2020. It has been noticed that in the last three-decade, intrusion detection-related publications and research-related articles are continuously growing after the year 1998 with minor crest and troughs.

A NIDS comprises different modules that are shown in Figure 2. These modules perform the detection process for intrusive information in a network. The three modules that comprise a NIDS with their function are shown in Figure 2. The detection machines module helps to detect intrusion or anomalies. The detection software performs detection strategies, and the management machine manages the detection strategies or policies. The other sub-modules of the detection machine module is the data capture module. The intrusion detection module and communication modules capture packets from the network. The second module of a traditional NIDS, Management Machine, is used for managing and maintaining detection policies based on detection strategies. The database is the third module that maintains and stores recorded behavior of intrusion detection based on feature extraction. The most common issues faced by a NIDS are fidelity problems, resource usage, and reliability. Existing intrusion detection systems suffer from at least two of the problems defined by Hoque *et al.* [23]. The various phases of the network intrusion detection model (NIDM) are shown graphically in Figure 2.

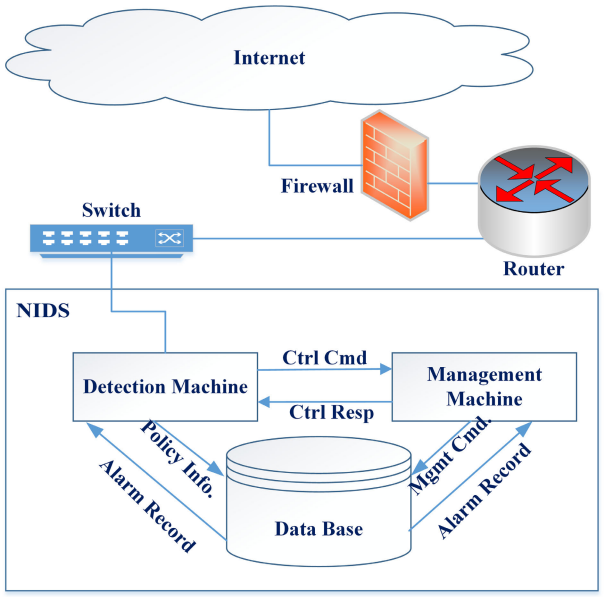

FIGURE 2. A Network-based intrusion detection system (NIDS) with its components.

In Figure 2, Mgmt Cmd stands for management command, Ctrl Resp stands for control response, Ctrl Cmd represents control command, and Policy Info is abbreviated for policy information.

A. CAUSES OF INTRUSION IN NETWORK

Based on Anchugam and Thangadurai [24] and Ghorbani *et al.* [25], we observed some commonly occurred causes of intrusion in a network. These are as follows.

a. Bad packets (produced from corrupt domain name system (DNS) data, software bugs) and local packets may not be detected significantly, which causes high false-alarm rates (false positive).
b. The encrypted packets may cause intrusion, which is not preventive without effective IDS.
c. IDS may not effectively imply the identification and authentication for weak access in the network. When an attacker gains admittance due to a soft authentication mechanism, then IDS is preventive for the misconduct.
d. NIDS systems can be subject to some protocol-based attacks, then hosts in that network may be vulnerable to illegal data, and Transmission Control Protocol/Internet Protocol (TCP/IP) stack attacks may be the reason for the crash of an NIDS.

B. COMPARISONS OF SOME POPULAR NIDS

There are many NIDS that are used commercially for network security purposes. Some popular NIDSs, in Table 2, are tabulated with their comparative analysis.

V. BENCHMARK DATASETS USED IN NIDS MODELS

Various datasets as benchmark datasets have been used to evaluate the intrusion detection model. The work done on the various datasets is to exhibit better classification accuracy, and detection rate [29]. There are many intrusion detection datasets published over the last few years. Finding a relevant dataset to evaluate an intrusion detection model is a tough job. Ring *et al.* [30] explored a survey on existing datasets for network-based intrusion detection along with an analysis of their properties, attack scenarios, and relations between the datasets.

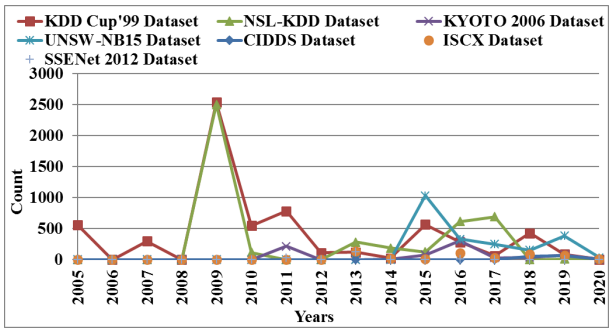

FIGURE 3. Year-wise distribution of the citation for different datasets from the year 2005 to 2020.

Here, regarding the popularity of various datasets, statistical comparison-based citations along with advantages and disadvantages of different benchmark datasets are tabulated in Table 3. Table 3 shows that the KDD Cup'99 dataset has the highest citation as a benchmark dataset since 2005. It means that the highest work has been done using the KDD Cup'99 as a benchmark dataset compared to the other datasets. The second most cited dataset is NSL-KDD, according to Table 3.

A newer dataset containing more modern attacks, such as the UNSW-NB15 dataset generated for the Australian Centre for Cyber Security [40] is also used as a benchmark dataset. This dataset comprises nine sorts of attacks and has a training set with one hundred and seventy five thousand records and a testing set with eighty-two thousand records. The hypertext transfer protocol (HTTP)-based dataset was generated for the CSIC (Consejo Superior de Investigaciones Científicas (Superior Council of Scientific Investigations)), Spanish Research National Council, in 2010 to report the criticisms of KDD'99 [25]. The dataset contains thirty-six thousand of which are 'normal' requests and more than twenty-five thousand anomalous. These datasets may be more applicable for specific cases; however, they are not as ubiquitous as KDD Cup'99 and NSL-KDD datasets. For demonstration of the benchmark datasets, KDD Cup'99 and NSL-KDD are ideal datasets since many papers describe their implementations specifically [41]–[43].

The year-wise distribution of various datasets is presented graphically in Figure 3. This graph shows that the KDD Cup'99 dataset has the highest popularity, followed by the NSL-KDD dataset from 2005 to 2018. Meanwhile, other datasets also came into existence.

TABLE 2. Popular commercially used intrusion detection systems.

S. No.	NIDS	Manufacturer	Approaches Used	Advantage
1	Snort: Created by Martin Roesch, 1998 [26]	Cisco Systems, Sourcefire https://www.snort.org/	Signature-based, network intrusion detection, pattern matching Aho-Corasick algorithm [27].	Free open-source, real-time alerting, and packet logging
2	OSSEC: Open-Source HIDS Security Daniel B. Cid owned the copyrights of the OSSEC project, 2008.	AlienVault® OSSIM™, in 2008 Currently maintained by Atomicorp https://atomicorp.com/about-ossec/	Correlate and analyze logs, log-based intrusion detection	File Integrity Monitoring (FIM), log monitoring, rootkit detection, auditing, export to SIEMs, active response, process monitoring, time-based alerting, and log analysis
3	OSSIM: Open-Source Security Information and Event Management (SIEM)	AlienVault® OSSIM™, in 2008 Currently, AT &T Cybersecurity in 2019	Log processing, correlation directives (rules), behavioral monitoring, SIEM event correlation	Lacks support for Cloud-based servers and applications Reports are heavy and detailed, and tedious to parse through
4	Suricata: Free and open-source, a real-time intrusion detection system	Owned and supported by the Open Information Security Foundation (OISF) www.openinfosecfoundation.org	Signature-based intrusion detection, process multithreading to improve processing speed [28]	Suricata can handle larger volumes of traffic as compared to Snort
5	Bro: An open-source software framework that detect behavioral abnormalities on a network	Initially written by Vern Paxson Later in 2018, Paxson and the project's leadership team gave a new name to this project Zeek for developing the IDS. Like Suricata or Snort, it is also rules-based IDS. https://bricata.com/blog/what-is-bro-ids/	Script interpretation.	Transforms network traffic data into higher-level events. Offers a script interpreter
6	Fragroute/ Fragrouter: A network intrusion detection evasion toolkit	D. Son https://monkey.org/~dugsong/fragroute/	When Fragroute initialize, it deletes the route to the target Intercepts network traffic and modifies the packets before forwarding	Probe packets can be fragmented easily with Fragrout ICMP echo request messages are used by fragtest
7	BASE: Basic Analysis and Security Engine (BASE) offers a web-based front end for examining the alerts produced by Snort.	https://sourceforge.net/projects/secureideas/	It offers a web front-end to query and analysis the alerts produced by a SNORT IDS.	User authentication and role-based system Search interface and Query-builder for identical alerts matching from the alert meta information Packet viewer (decoder)
8	Sguil: Built by a group named network security analysts	https://github.com/bammv/sguil/releases/tag/v0.9.0	Event-driven analysis Network Security Monitoring	Captures raw packet, session data, and Real-time events Compatible on the operating system that supports TCL/TK Receive alerts from OSSEC, Zeek, Suricata, Snort, and other data sources.

VI. APPROACHES USED IN NIDS MODELS

Classical intrusion detection problem-solving methodologies, according to Liu and Lang [2] and Jyothsna *et al.* [44], introduced four branches based on methodologies: statistical-based, knowledge-based, machine learning-based, and bioinspired-based along with their used approaches. Table 4 presents the total number of publications, the highest citation in conferences and journals related to intrusion detection, along with the methodologies. Here, we considered research articles from the year 2005 to 2020. There are many optimization approaches for finding the optimal rating of intrusion detection. Table 5 to Table 8 represent different optimization approaches, their corresponding methodologies, citation, and published articles records for intrusion detection models.

Intrusion detection with machine learning approaches, which have the highest publications, is shown graphically in Figure 4. Comparison of citations related to the articles published in conferences and journals among various methodologies, viz-a-viz statistical-based, knowledge-based, and bioinspired-based is shown graphically in Figure 4. Figure 4 also depicts the most cited articles published in conferences those attained a high count than the articles published in journals.

A. STATISTICAL-BASED NIDS

A statistical-based intrusion detection system (SBIDS) [52] use statistical observation on different variables like the log-in session, resource overflow flags, and timers. The statistical properties like mean, standard deviation, correlation, Analysis of Variance (ANOVA), and statistical tests determine the deviation from the 'normal' behavior of network traffic flow [53].

TABLE 3. Benchmark datasets used in NIDS models.

S. No.	Datasets	Advantages	Disadvantages	Total Number of Citation from the year 2005 to 2020
1	KDD CUP'99 : KDD stands for Knowledge Discovery in Databases https://bit.ly/3wSG5YA	Extensive repository of attack vectors, Large amount of attacks	Obsolete in fixing of many attacks It does not provide real attack data [31]	6382
2	NSL- KDD : NSL-KDD is an updated version of the KDD cup99 data set where NSL stands for Network Security Laboratory http://205.174.165.80/CICDataset/NSL-KDD/	No duplicate data found within the NSL-KDD train dataset and Test set Contains a reasonable number of samples by train and test sets [31]	According to McHugh et al. [32], NSL–KDD may not represent real network flow [33]	4521
3	UNSW-NB15 Dataset : https://bit.ly/2Q1k895 [34]	Separate training and test set 45 Distinct IP addresses Publicly available	UNSW-NB15 datasets contain a limited number of attacks and no attacks related to cloud computing, like SQL injection Imbalanced training and testing classes [35]	2172
4	Kyoto 2006+ : https://bit.ly/3gLf7vo	14 attributes are the same as in KDD CUP 99, besides 10 new attributes Provides real attack data [36]	Limited volume of 'normal' traffic and 'Normal' data is unrealistic that makes it a low backdrop for the attack data Normal traffic records incorporated from a single server with a single domain that includes e-mail and DNS traffic only Limited realistic variety of 'normal' traffic records	731
5	ISCX 2012 Dataset : ISCX stands for Information Security Center of Excellence https://bit.ly/2Rl9o5B	Upto-date dataset compared to the other commonly datasets Representative of real network traffic Dynamic, scalable, reproducible, and labeled benchmark dataset [37]	ISCX-2012 does not comprise any novel traffic attributes or session-based records Due to its unidirectional nature, it is challenging to the buried context in the payload data [36]	472
6	SSENet 2012 Dataset : (Unknown link)	Generated in a real network environment	Unknown	158
7	CIDDS-001 Dataset : The University of Coburg published CIDDS (Coburg Intrusion Detection Data Sets) https://bit.ly/3mLQAYT	Contains detailed metadata for more in-depth investigations Contains modern attacks network traces Multi-class and binary classes. [38]	It includes some biased features such as host IP, destination IP, and Date identified may create biases and not be helpful to detect the attacks [39]	121

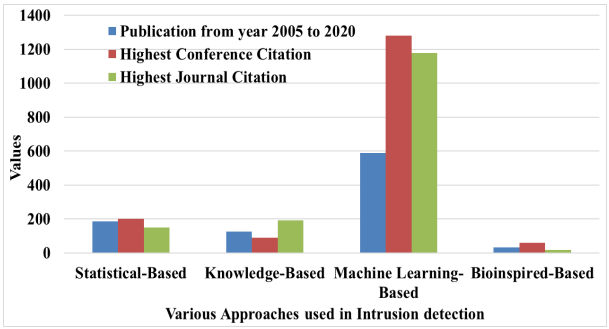

FIGURE 4. The total number of publications, the highest citation in conferences, and the journals for the different approaches of intrusion detection systems between the years 2005 to 2020.

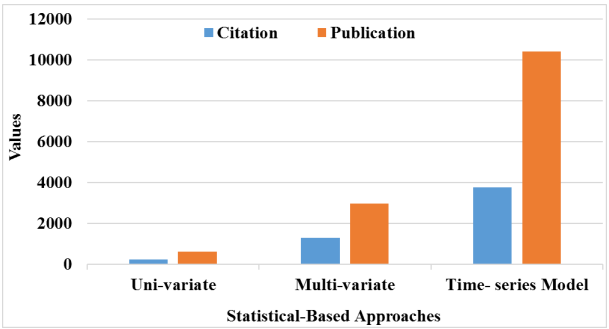

FIGURE 5. Various statistical-based Intrusion detection approaches with the number of publications and the highest citation in conferences and the journal between the years 2005 to 2020.

Articles related to intrusion detection with time-series statistical approach have the highest publication count with the highest citation values than the other statistical approaches, as shown in Figure 5. Table 5 also enlightened the highest cited articles of journals and conferences with citation counts on intrusion detection among different statistical approaches.

Year-wise publications and citation distribution comparisons of articles among various statistical approaches are displayed graphically in Figure 6 and Figure 7 respectively. Articles regarding IDS based on the time-series model represent the highest citations and publications values than the other statistical approaches, shown in Figure 5.

Here, two sharp points are observed. One is that time series-based article publications are highest in counting than the other statistical approaches. Second, the number of

TABLE 4. Highest citation for an article from the Year 2005 to 2020 of the methodology used for network-based intrusion detection models.

Methodology	Description	Total Number of Publication	Conference publication with highest citation count	Journal publication with highest citation count
Statistical-Based	Stochastic behavior and well defined	185	200 Song et al. [45]	149 Zhang et al. [46]
Knowledge-Based	Availability of Prior Knowledge of data/ information	126	89 Midi et al. [47]	193 Ben-Asher and Gonzalez [48]
Machine Learning-Based	Categorization of Patterns	588	1280 Sommer and Paxson [49]	1179 Buczak and Guven [50]
Bioinspired-Based	Bio-inspired computing imparts to machine learning and artificial intelligence	31	59 Liang and Xiao [51]	16 Balasaraswathi et al. [29]

TABLE 5. Total number of publication, citations, and highest citation of articles (conferences and journals) for different approaches of statistical methodologies from the year 2005 to 2020.

Approaches	Total Publication	Total Number of Citation	Conference publication with highest citation count	Journals publication with highest citation count
Time Series	10414	3765	94 Viinikka et al. [54]	88 Viinikka et al. [55]
Multivariate	2972	1286	64 Delgosha and Fekri [56]	25 Sarasamma and Zhu [57]
Univariate	618	227	0	81 Wang et al. [58]

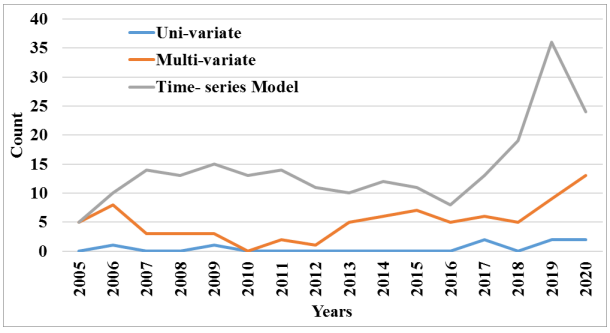

FIGURE 6. Year-wise articles publication distribution for statistical approaches from the year 2005 to 2020.

publications has grown from 5 publications in 2005 to 36 publications in 2019. A year-wise publication analysis among different statistical approaches and articles based on time series model-based intrusion detection also showed the highest publication in 2019 with 36 publications.

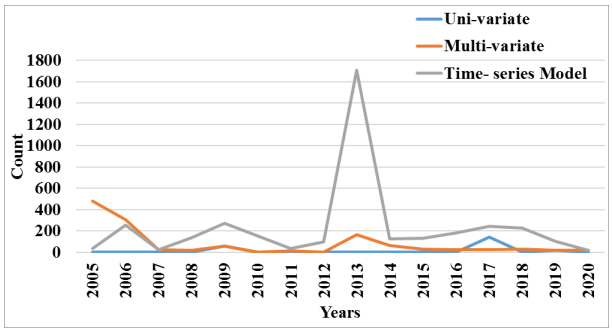

FIGURE 7. Year-wise citation distribution of articles for statistical approaches from the year 2005 to 2020.

TABLE 6. Analysis of research trends in different approaches of Knowledge-based intrusion detection systems.

Approaches	Total Publication	Total Number of Citation	Conference publication with highest citation count	Journals publication with highest citation count
Expert System	3313	3185	108 Treinen and Thurimella [60]	1695 Patcha and Park [59]
FSM	2429	2309	400 Tan & Sherwood [61]	116 Tan et al. [62]
Descriptive Language	49	36	3 Zhu et al. [63]	1 Sourek and Zelezny [64]

Figure 7 shows a year-wise citation count for statistical approaches. The year 2013 has the highest citation score, but citations among other years remain almost the same. It implies that the popularity related to the time-series model-based statistical approach has a steady increase. It also depicts that research on time-series model-based intrusion detection is almost constant from 2005 to 2020. Around 200 citations per year are added in the citation records with 10414 published articles.

B. KNOWLEDGE-BASED NIDS

Knowledge-based IDS (KBIDS) congregate intrusive information about networks and produces less false alarm rate with high accuracy in intrusion detection. But KBIDS requires up to date knowledge repository about network traffic behavior [53]

All knowledge-based techniques with their total number of publications and citations are tabulated in Table 6. The highest cited article based on the expert system is by authors Patcha and Park [59] with the count 1695.

Figure 8 depicts that expert system-based publication count and citation have a higher value than finite state machine (FSM) and descriptive language. As shown in Figure 9, initially, the published article counts have higher values for the FSM technique from 2005 to 2011 as compared with other techniques in the knowledge-based methodologies. The publication count was highest in 2010, with a value of 30 for the FSM technique for the knowledge-based

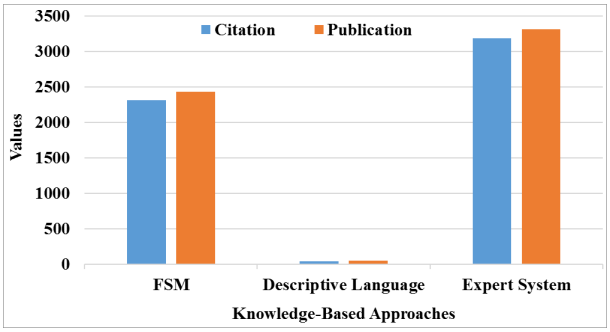

FIGURE 8. Publications versus citations among various knowledge-based approaches used in the intrusion detection system.

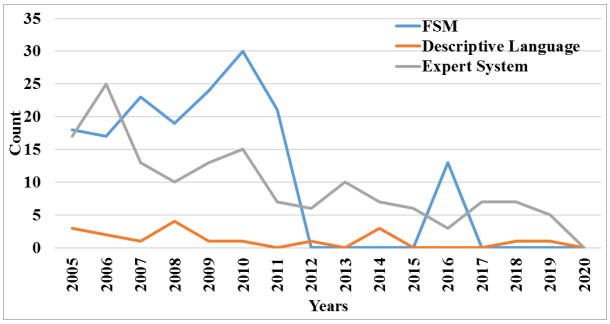

FIGURE 9. Publication analysis for different knowledge-based approaches along with intrusion detection.

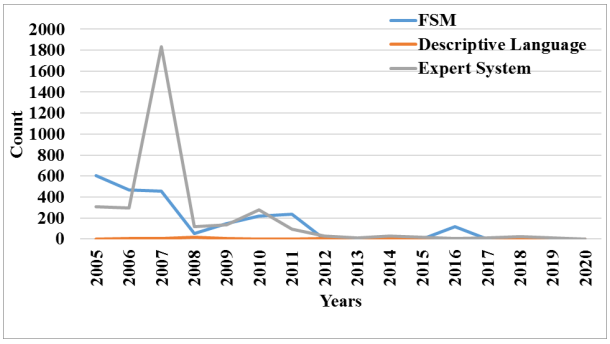

FIGURE 10. Citation analysis for different knowledge-based approaches along with intrusion detection.

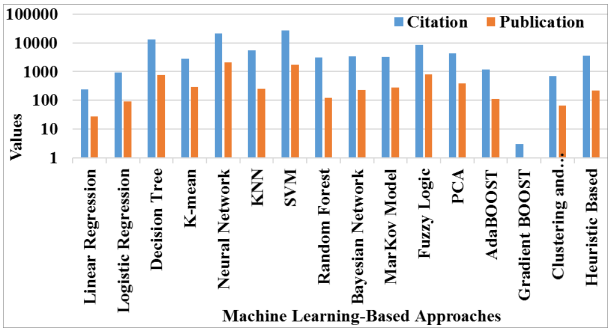

FIGURE 11. Publications versus citations among several machine-learning approaches implemented for intrusion detection.

research articles. But the total published articles based on the expert system are 3313, which is the highest among other research published articles in knowledge-based intrusion detection articles.

The most cited article based on a knowledge-based approach for intrusion detection is [59], written by Patcha and Park with 1695 citations. The work in [59] explores the use of finite state machines as a knowledge-based approach. Figure 10 represents the year-wise citation distribution of knowledge-based research articles in the intrusion detection field. Based on the three curves in Figure 8, expert system-based articles have higher citations from 2005 to 2020. It means that the research trend in the expert system approach of knowledge-based intrusion detection is higher than the other knowledge-based methodologies.

Figure 10 shows that before a decade, expert-system-based research trends had some significant values. But in the present scenario, knowledge-based research for intrusion detection is not a valuable research trend. The research trend based on different knowledge-based techniques gets falls after the years 2008 and 2010.

C. MACHINE LEARNING-BASED NIDS

Traditionally, NIDSs are designed based on high-dimensional network traffic classification into normal or intrusive data. Due to the high dimensionality of network traffic data, intrusive information detection is significantly slower in traditional NIDS. Such traditional NIDSs with a machine learning approach on selected features take comparatively low FPR (false-positive rate) with a high TPR (true positive rate) for predicting the traffic behavior of network [65]. Machine learning-based classifier models trained and fit over on the training sets among selected 'important' features. The 'important' and relevant feature subsets are selected based on which machine learning-based classifier gets trained. Training sets consist of respond classes over which the classifier gets trained and fit over to recolonize network traffic data behavior/classes.

In Table 7, machine learning-based publication and citation counts of articles for IDS are tabulated. According to this table, the SVM is the utmost interested (cited) technique for intrusion detection researchers. The neural network, followed by the decision tree, is also an exciting machine learning-based intrusion detection technique. Table 7 also depicts the total number of publications and citation counts of articles and the most referred articles published in the conferences or the journals.

Articles based on SVM for intrusion detection systems have the highest cited topic in the research. Even though published articles are higher on neural-network-based intrusion detection than the SVM and fuzzy logic, as shown in Figure 11. In contrast, Gao *et al.* [33] explored the drawbacks of the SVM algorithm, which consumes a long time and without gain of accuracy. The Adaboost-based model is not

TABLE 7. Machine learning-based analysis of research trends in intrusion detection.

Approaches	Total Publication	Total Number of Citation	Conference publication with highest citation count	Journals publication with highest citation count
SVM	1716	27329	124 Heba et al. [66]	188 Al-Yaseen et al. [67]
Neural Network	2152	21705	789 Shi et al. [68]	817 Wu and Banzhaf [69]
Decision Tree	757	12798	307 Stein et al. [70]	744 Chebrolu et al. [71]
Fuzzy Logic	820	8415	56 Shanmugam and Idris [72]	108 Fulsoundar V.S et al. [73]
KNN	252	5496	124 Alazab et al. [74]	265 Aburomman and Reaz [75]
PCA	381	4279	125 Wang and Battiti [76]	184 Pajouh et al. [77]
Heuristic-Based	217	3542	204 Fogla and Lee [78]	255 Aydın et al. [79]
Bayesian Network	229	3374	7 Tabia and Leray [80]	744 Chebrolu et al. [71]
Markov Model	284	3271	92 Khanna and Liu [81]	194 Hu et al. [82]
Random Forest	121	3152	241 Zhang et al. [83]	291 Sindhu et al. [84]
K-mean	297	2804	120 Jianliang et al. [85]	265 Tsai and Lin [86]
AdaBOOST	111	1174	400 Hu et al. [87]	133 Hu et al. [88]
Logistic Regression	90	911	72 Gates et al. [89]	102 Y. Wang [90]
Linear Regression	28	237	13 Hassanzadeh and Sadeghian [91]	0
Clustering and Outlier Detection	15	45	32 Mingqiang et al. [92]	10 Jeyannaet al. [93]
Gradient BOOST	1	3	3 Montalbo and Festijo [94]	0

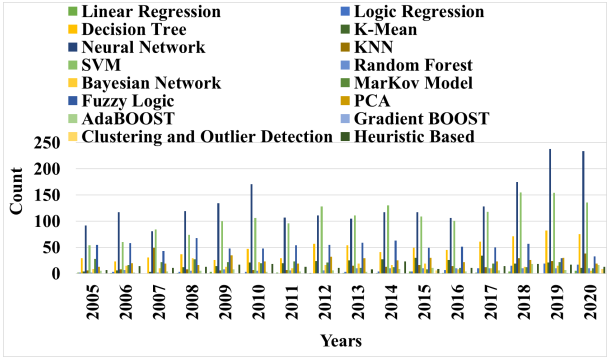

FIGURE 12. Publication analysis among several machine-learning approaches applied in IDS.

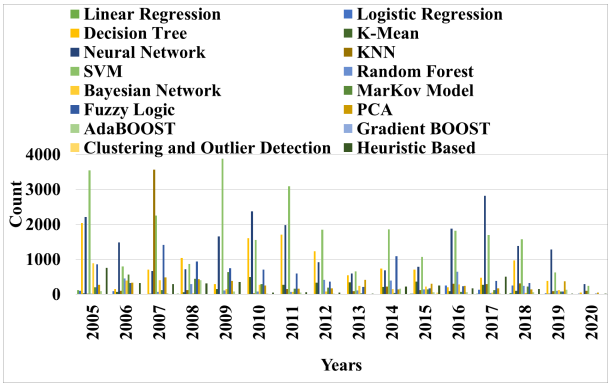

FIGURE 13. Citation analysis for different machine-learning approaches applied in IDS.

ideal whereas, the precision obtained on implementing logistic regression algorithm is not high for intrusion detection.

Figure 12 depicts that a neural network with intrusion detection is the prime choice for authors with 2152 publications from 2005 to 2020. The second most popular technique among authors is SVM, with 1716 publications for intrusion detection after the neural network. Based on 757 publications, the third rank is observed for the decision tree with intrusion detection.

On the other side, the total citation count of published articles is also recorded based on the different machine-learning practices among IDS. Figure 13 represents this citation analysis. The total citation count regarding neural networks is 21705, which is less than the citation count among SVM. SVM has a citation count of 27329. Moreover, SVM has the highest citation count rank, the neural network has the second rank, and the decision tree has the third rank followed by fuzzy logic.

It means that SVM is the most favorite subject among researchers and academicians regarding intrusion detection references.

TABLE 8. Analysis for various bio-inspired approaches along with different algorithms for intrusion detection.

Bio-inspired Approaches	Different bio-inspired Algorithms	Publication	Citation	Conference publication with highest citation count	Journal publication with highest citation count
Evolution-Based Algorithm	Genetic Algorithm	899	9240	212 Gong et al. [95]	27 Pawar et al. [96]
	Genetic Programming	106	2179	97 Hansen et al. [97]	87 Faraoun and Boukelif [98]
	Evolutionary Algorithm	122	1529	35 Gómez et al. [99]	5 Pan and Jiao [100]
	Evolution Strategy	7	17	-	7 Zhang et al. [101]
Swarm-Based	ACO	118	1453	118 Tsang and Kwong [102]	232 Feng et al. [103]
	Bee Colony	49	402	59 Wang et al. [104]	66 Hajimirzaei and Navimipour [105]
	Firefly	23	242	1 Devi and Suganthe [106]	89 Shah and Issac [107]
	Fish Swarm	10	106	15 Liu et al. [108]	78 Hajisalem and Babaie [109]
	PSO	1	0	0 XU Wenbo [110]	-
	BFO	1	0	-	0 Kalaivani and Ganapathy [111]
Ecology-Based		4	25	6 Sakar and Kursun [112]	2 Na et al. [113]

D. BIO-INSPIRED-BASED NIDS

Bio-inspired are popular approaches used for optimization and problem-solving. The requirement for enhancing accuracy and efficiency enforces the use of bio-inspired stochastic algorithms, such as particle swarm optimization (PSO), Genetic algorithm (GA), for solving deterministic problems.

The total published article count, total citation count, most cited/ referred articles concerned with bio-inspired approaches are tabulated in Table 8.

Table 8 along Figure 14 represent the comparison of total article publication and citation count. Genetic programming has the highest published article count with a value of 899 and spikes a citation count with the value 9240. Figure 14 represents a year-wise distribution of the total number of publications and total citation counts, yearly.

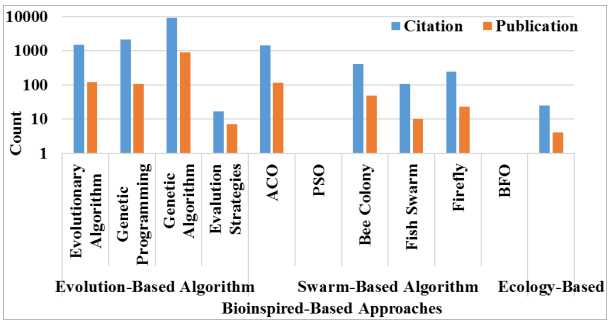

FIGURE 14. Analysis for various bioinspired approaches along with the total number of publications and total citations from the year 2005 to 2020.

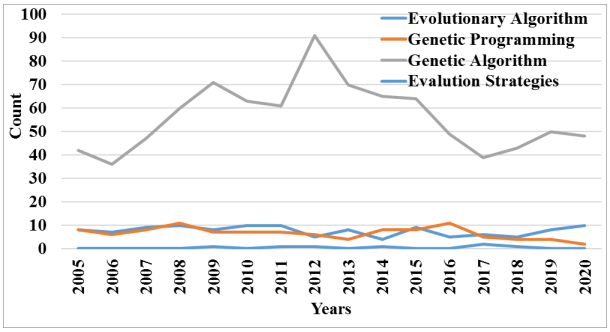

FIGURE 15. Publication analysis for various bio-inspired (Evolutionary) algorithm-based intrusion detection.

Genetic algorithm-based articles are highest published in the category of evolution-based intrusion detection. On the other hand, articles published based on ant colony optimization (ACO) for intrusion detection have the highest value in the swarm-based algorithm category. Published articles count in the ecology-based category has a nominal value with 4 numbers.

Figure 15, Figure 16 and Figure 17 represent a year-wise distribution of published articles count, while Figure 18, Figure 19 and Figure 20 represent the year-wise citation counts for the different bio-inspired approaches used for intrusion detection systems. According to Figure 15 and Figure 18, year-wise published articles count and citation count for genetic algorithm along intrusion detection system have a high distribution in evolution-based category. Similarly, according to Figure 16 and Figure 19, the published articles count and citations for ACO with intrusion detection have the highest year-wise distribution in the swarm-based category.

Hence, ACO in the swarm-based category has the highest research trend in intrusion detection. The genetic algorithm-based IDS has high published article distribution and a high research trend in the evaluation category.

Similarly, ACO in the swarm-based category has the highest research trend for intrusion detection. The genetic algorithm-based IDS has high published article distribution and a high research trend in the evaluation category.

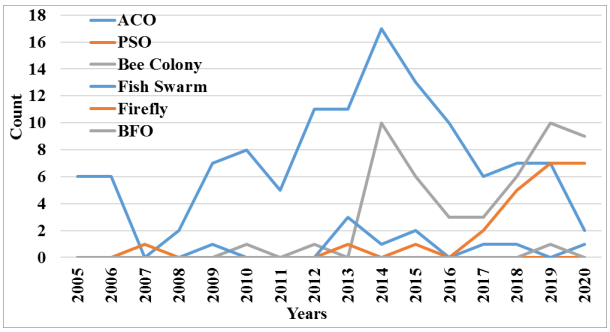

FIGURE 16. Publication analysis for different bioinspired (Swarm)-based intrusion detection.

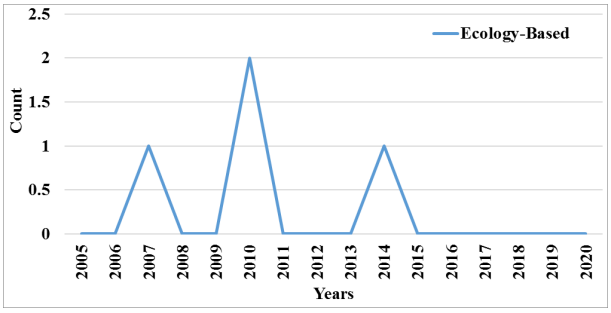

FIGURE 17. Publication analysis for different ecology-based intrusion detection.

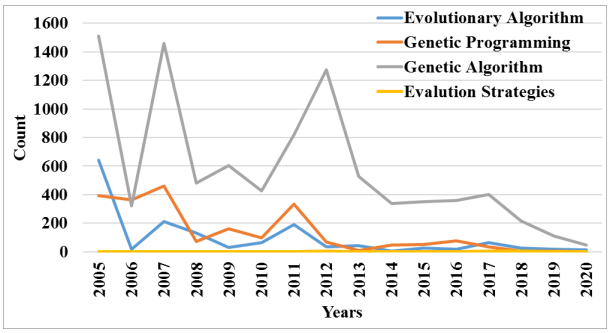

FIGURE 18. Year-wise citation analysis of articles for various evolution-based algorithms in bio-inspired methodology.

E. COMPARISON OF THE MOST CITED APPROACHES USED IN NIDS

Based on the publication count and citation of articles, Table 9 presents a comparison among the most popular methodologies used for intrusion detection. The most cited approaches in intrusion detection are time-series in statistical-based methodologies, expert systems in knowledge-based methodologies, SVM in machine learning-based methodologies, and genetic algorithms in bio-inspired-based methodologies.

In a time series statistical-based intrusion detection system, a series of events are observed within the interval of time. If a new event falls within a specific time, the possibility of being normal is high. Otherwise, the possibility for an event of being normal is very low [114]. Expert systems (ES) are rule-based approaches used in KBIDS, comprising rules,

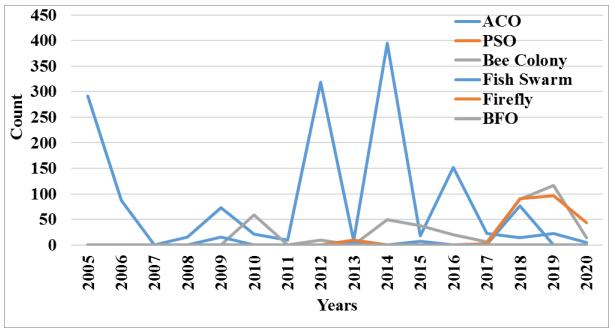

FIGURE 19. Year-wise citation analysis of articles for various swarm-based algorithms in bio-inspired methodology.

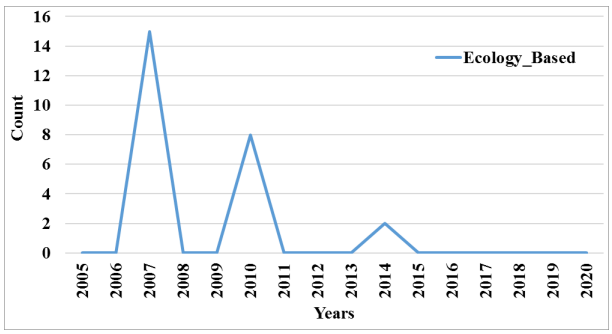

FIGURE 20. Year-wise citation analysis of articles for various ecology-based algorithms in bio-inspired methodology.

facts, and inference methods. Each event is first converted into related facts and rules in an IDS system, and then some inference rule is applied to generate prediction [53]. SVM uses a hyper-plan for differentiating the response classes of the dataset. Genetic algorithm (GA) is an evolutionary algorithm-based approach in which optimization is based on mutation [52]. GA encodes a set of solutions to form a population, and GA evolute this population based on fitness function [65].

VII. PERFORMANCE MEASUREMENTS

The performance of network security can be calculated based on efficiency and effectiveness. Efficiency deals with the resources needed to be allocated to the system, including CPU cycles and main memory. In comparison, effectiveness describes the ability of the system to distinguish between intrusive and non-intrusive activities. In the context of IDS evaluation, researchers generally use metrics to measure the effectiveness quantitatively based on the training and testing of the classifier using benchmark datasets. These metrics measure how well the attack instances are detected against normal instances. The confusion matrix and the receiver operating characteristic (ROC) curve are mainly used to calculate the effectiveness of the IDS.

The total publication count for the confusion matrix and ROC curve is 2045 and 54, respectively. On the other side, the citation count for the confusion matrix and ROC is 75349 and 711, respectively. The searching strings as per Table 1

TABLE 9. A comparative analysis of most cited approaches used in intrusion detection.

S. No	Approaches	Advantages	Disadvantages	Total Number of Citation from the year 2005 to 2020
1	Time series in Statistical-based	Clearly defined procedure, better resource allocations	Computationally expensive for generalization rules from a single study and some appropriate measures within a specific time are not possible	3765
2	The expert system in knowledge-based	Applicable to both anomaly and signature-based IDS	High semantic rules generate complex event data abstraction	3185
3	SVM in machine learning-based	SVM provides a solution for solving the more complex problem using kernel function, less over-fitting, applicable on unstructured and semi-structured, and high dimensional data.	Hyperparameter C and γ are not easy to tune, Slow training on large data	27329
4	Genetic algorithm in bioinspired-based	Easily retrained Systems, easily paralleled, robust for local minima and local maxima	There is less cross rate with a high mutation in intrusion detection. Choosing a fitness function is also complex.	9240

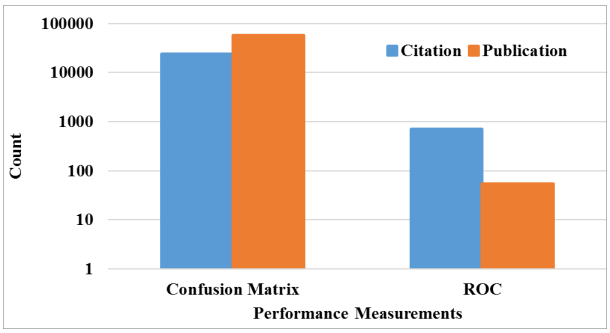

FIGURE 21. Analysis for confusion matrix and ROC based on the total number of publications and total citations from the year 2005 to 2020.

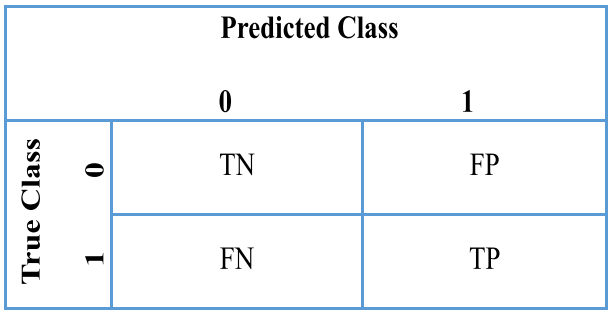

FIGURE 22. Confusion Matrix and Performance Measurement.

with the filters as described in the Section III are used for the selection of publication and citation records. Figure 21 shows a comparison of citation and publication counts regarding confusion and ROC in the intrusion detection field. This figure depicts that the confusion matrix is more popular and having high research trends in intrusion detection for evaluating IDS models.

A. CONFUSION MATRIX

The confusion matrix is an easy and effective way to characterize the classification results of an IDS. The equations of metrics, as shown in (1) to (8), are based on the fundamental measuring parameters of the confusion matrix, as shown in Figure 22.

The fundamental parameters are

- TP: True positive (TP) are the classified instances as 'normal' traffic flow.
- TN: True negatives (TN) are the classified instances as 'attack' traffic flow.
- FP: False positives (FP) are the wrongly classified instances as 'normal' instead of 'attack'
- FN: False negatives (FN) are the wrongly classified instances as 'attack' instead of 'normal'

The performance of the NIDM with data mining classifier is measured based on the following metrics discussed by Almomani [65], and Ferrag *et al.* [115] also.

1. Misclassification Rate (MCR): MCR defines how often is the classifier wrong.

$$MCR = \frac{FP+FN}{TP+TN+FP+FN} \tag{1}$$

2. Accuracy: It gives the total number of correct classifications, i.e., how often is the classifier correct.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} X100\ (\%) \tag{2}$$

3. True Positive Rate: It is also known as Recall or Sensitivity. It gives the total number of correct classifications regarding incorrect classification.

$$TruePositiveRate = \frac{TP}{TP+FN} \tag{3}$$

4. Specificity: Specificity is also known as the true negative rate (TNR). It represents how properly a classifier identifies true negatives. It gives the number of intrusive classifications regarding the total number of intrusive data (i.e., $TN + FP$) during training.

$$Specificity = \frac{TN}{FP+TN} \tag{4}$$

5. Precision (Prec): When it predicts 'normal', how often is it correct.

$$Prec = \frac{TP}{TP+FP} \tag{5}$$

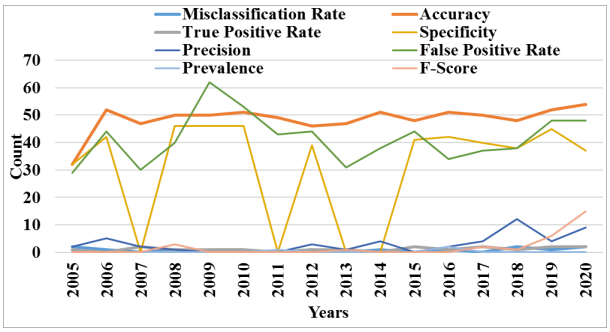

FIGURE 23. Publication distribution analysis of various evaluation metrics used for the intrusion detection system.

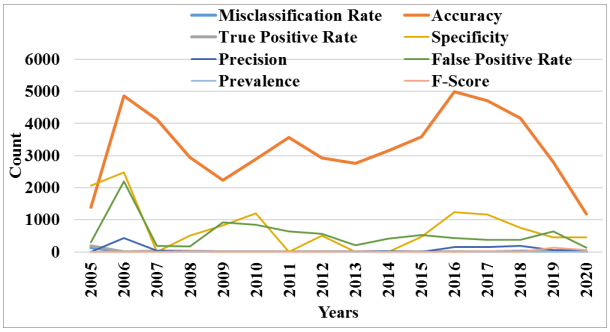

FIGURE 24. Citation analysis of various evaluation metrics used for intrusion detection systems.

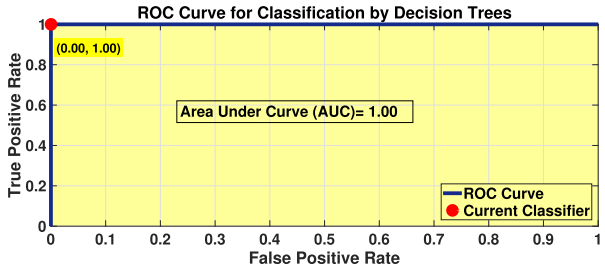

FIGURE 25. ROC curve and AUC for decision tree as a classifier in the intrusion detection model.

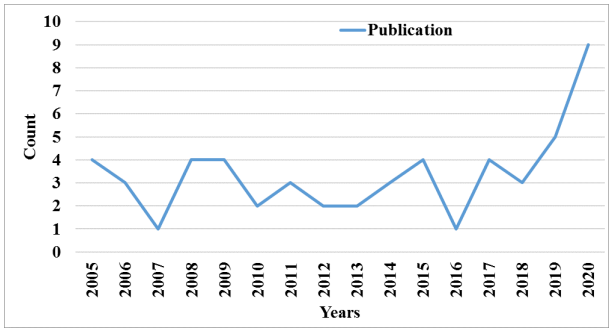

FIGURE 26. Publication distribution analysis for ROC for evaluation of intrusion detection system.

6. False Positive Rate (FPR): When it is actually attacked, how often does it predict normal

$$FPR = \frac{FP}{TP+FN} \tag{6}$$

7. Prevalence: Prevalence tells that how often does the yes condition actually occurs (i.e. total TP+FN) in our sample.

$$Prevalence = \frac{TP+FN}{TP+TN+FP+FN} \tag{7}$$

8. F-Score: It serves as a derived effectiveness measurement.

$$\text{F-Score} = \frac{2 * TP}{2 * TP+FP+FN} \tag{8}$$

The obtained values from these metrics lie between 0 and 1 except accuracy which represents the percentage value.

Figure 23 and Figure 24 presents the year-wise distribution of published article count and citation count respectively. These graphs show that accuracy has the highest popularity followed by specificity and FPR.

B. ROC

The ROC (receiver operating characteristic) curve can also be used to measure the efficiency and efficacy of an IDS. A ROC is termed a performance curve. ROC is a graph between detection accuracy against false alarm rate. Alternatively, it displays the false alarm rate generated by the detector at a specified probability of detection [116]. The area under the curve (AUC) determines the misclassification in an IDS. If AUC is less than or equal to 0.5, it means misclassification is more than 50 percent, and the performance is poor for intrusion detection model [117].

For an illustration, we simulated an intrusion detection model using a decision tree as a classifier in an environment of Intel Core i5 2.60 GHz with 7.88 GB of RAM along with MatLab R2017b. The KDD cup'99 dataset is considered a benchmark dataset. The training set, which consists of 494021 records, is trained on the two response classes which are either of 'intrusion' or 'normal' dataset records. Hence, Figure 25 as ROC is plotted for the simulated IDS model.

In Figure 25, the x-axis signifies the FPR (False Positive Rate) with a value of 0.00.While, the y-axis denotes the TPR (True Positive Rate) with a value of 1.0. Here, the area under curve have a 1.00 value that represents 100% classification accuracy (means 0% misclassification) for the training of the classifier.

Figure 26 and 27 represents the year-wise distribution of publication and citation count, respectively. These two figures depict that articles related to ROC in the field of intrusion detection are growing gradually since 2005 and hence, the popularity and research trend in intrusion detection of ROC to evaluate the IDS model is minimal as compared to accuracy, specificity, and FPR.

The number of published articles, total citation count, highest cited article in conference and journal regarding

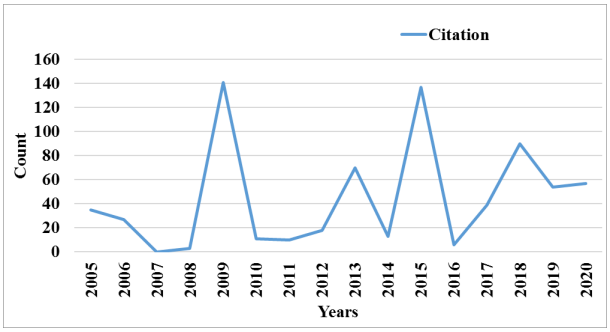

FIGURE 27. Citation analysis for ROC for evaluation in intrusion detection.

TABLE 10. Analysis for performance metrics for intrusion detection.

Performance measurements	Metrics	Publication	Citation	Conference publication with highest citation count	Journals publication with highest citation count
Confusion Matrix	Misclassification Rate	12	190	5 Atli et al. [118]	126 Yun Wang [90]
	Accuracy	778	52324	135 Xu and Wang [119]	331 Sindhu et al. [84]
	True Positive Rate	18	259	22 Xiao and Xiao [120]	190 Pietraszek and Tanner [121]
	Specificity	494	12158	745 Raza et al. [122]	325 Cavusoglu et al. [123]
	Precision	49	1150	23 Penya et al. [124]	66 Wang et al. [125]
	False Positive Rate	663	8988	111 Yu and Frincke [126]	2 Shanqing et al. [127]
	Prevalence	3	14	9 Jan Vykopal [128]	5 Li and Liao [129]
	F-Score	28	266	23 Ullah and Mahmoud [130]	73 Zeng et al. [131]
ROC		54	711	6 Sakar and Kursun [112]	2 Na et al. [113]

performance measurements are tabulated in Table 10. Table 10 along Figure 23, and Figure 26 represents the publication analysis. While, Table 10 along Figure 24, and Figure 27 represents the citation analysis of the different performance evaluation matrix in the field of intrusion detection.

VIII. DISCUSSION

As highlighted in Section I, an IDS provides a prominent mechanism to the computer network security system by generating an alarm on detecting malicious information. The study in this article presents quantification of popularity based on citations of articles related to NIDS. We have analyzed the intrusion detection-related articles categorically in different classes viz-a-viz the commercially used IDS, the datasets to evaluate the NIDS models, the approaches used in NIDS, and different evaluation metrics. Here, Microsoft Academic is used for taking the records of citations of the published articles related to datasets, and various methodologies with their subclasses.

The analysis for research trends in benchmark datasets to evaluate NIDS models is also presented graphically. It is found that the KDD Cup'99 dataset has the highest popularity, followed by the NSL-KDD dataset. But the problem with the KDD'99 dataset is that it is a very old dataset and it does not resemble the modern traffic data flow. Nevertheless, there are other datasets also available, but the research trend in these datasets is very less due to the less popularity of these new datasets among researchers. It is suggested that researchers must be encouraged to the new datasets with richer features according to the modern environment.

Bioinspired-based NIDS, especially swarm-based NIDS, has very limited literature. These approaches often show quick convergence. But, there is a lack of theoretical literature that how these algorithms perform quick convergence. Parameter tuning is also another major issue related to the bioinspired-based approach and there are only a few articles related to parameter tuning for these algorithms for intrusion detection.

The tabular and graphical analysis of this article explores that researchers are more attracted to the field in which a high count of published articles with high citation values are recorded. Furthermore, the year-wise distributions also show that researchers abide by the research field in which publication count and citation have a high value. Unquestionably, the future of IDS is promising. Furthermore, the research trends in IDS research will grow where publication count and citation have high values for the articles. It is evident from the literature review that researchers are required to evaluate algorithms based on bio-inspired approaches for intrusion detection. Machine learning and bioinspired-based new hybrid algorithms can also be evaluated and compared to promote efficient and accurate intrusion detection systems.

A similar type of approach as used in the current article can also be implemented to quantify research trends in other areas such as image processing, cloud computing, data mining, bioinformatics, etc. This type of review will help in finding the most popular as well as averse methodologies in a particular research area. More effort will be made by the research community after finding such popularity comparison in those approaches where less effort have been made.

In the future, we want to implement the less cited bio-inspired approaches that have few publication count values for articles related to network intrusion detection systems for future subsequent work. We want to determine whether the less cited approaches with fewer counts of published

articles are equally applicable to achieve an efficient and effective intrusion detection model.

IX. CONCLUSION

We explored a comprehensive and straightforward analysis for anyone who wants to compare various approaches used to design Network Intrusion Detection models. This review is established based on numerous research papers in different journals/publications between 2005 and 2020. In this article, we took citation as a quantitative measure to review the popularity of the intrusion detection system among various approaches. This paper presents various tables that offer a rapid analysis of different NIDS, research trends, and research scope. A review of diverse datasets with their characteristics, merits, demerits, and citation analysis has also been presented. The various approaches used in the network intrusion detection system are tabulated with their advantages and disadvantages also. A review concerning research trends regarding different techniques in IDS is presented.

The comparative research trend analysis regarding intrusion detection systems for a network is also graphically presented based on citation and number of published article counts. The most cited articles, with their citation count for conferences and journals, are also presented. The popular approaches, with the most cited papers regarding their methodology, are tabulated in different tables. We also observed that articles published in conferences have the highest citation than the articles published in journals.

REFERENCES

[1] R. Heady, G. Luger, A. Maccabe, and M. Servilla, "The architecture of a network level intrusion detection system," Los Alamos Nat. Lab., New Mexico Univ., Albuquerque, NM, USA, Tech. Rep. LA-SUB-93-219 ON: DE97002400; TRN: AHC29703%%44, 1990.
[2] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," *Appl. Sci.*, vol. 9, no. 20, p. 4396, Oct. 2019.
[3] L. C. Smith, "Citation analysis," *Library Trends*, vol. 30, no. 1, pp. 83–106, 1981.
[4] M. Thelwall, "Microsoft academic automatic document searches: Accuracy for journal articles and suitability for citation analysis," *J. Informetrics*, vol. 12, no. 1, pp. 1–9, Feb. 2018.
[5] M. Gusenbauer, "Google scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases," *Scientometrics*, vol. 118, no. 1, pp. 177–214, Nov. 2018.
[6] *The Top List of Academic Search Engines*, Paperpile, Cambridge, MA, USA, Jun. 2021. [Online]. Available: https://paperpile.com/g/academic-search-engines/
[7] *List of Academic Databases and Search Engines*, Wikimedia Found., San Francisco, CA, USA, Jun. 2021.
[8] A. Martín-Martín, M. Thelwall, E. Orduna-Malea, and E. D. López-Cózar, "Google scholar, Microsoft academic, scopus, dimensions, web of science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations," *Scientometrics*, vol. 126, no. 1, pp. 871–906, Jan. 2021.
[9] Microsoft. *Microsoft Academic*. Accessed: Dec. 20, 2020. [Online]. Available: https://academic.microsoft.com
[10] M. Bishop, "Trends in academic research: Vulnerabilities analysis and intrusion detection," *Comput. Secur.*, vol. 21, no. 7, pp. 609–612, Nov. 2002.
[11] P. Kabiri and A. A. Ghorbani, "Research on intrusion detection and response: A survey," *Int. J. Netw. Secur.*, vol. 1, no. 2, pp. 84–102, 2005.
[12] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," 2013, *arXiv:1312.2177*.
[13] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Proc. Comput. Sci.*, vol. 60, no. 1, pp. 708–713, 2015.
[14] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *J. Netw. Comput. Appl.*, vol. 60, pp. 19–31, Jan. 2016.
[15] D. Camacho, "Bio-inspired clustering: Basic features and future trends in the era of big data," in *Proc. IEEE 2nd Int. Conf. Cybern. (CYBCONF)*, Jun. 2015, pp. 1–6.
[16] A. A. Gendreau and M. Moorman, "Survey of intrusion detection systems towards an end to end secure Internet of Things," in *Proc. IEEE 4th Int. Conf. Future Internet Things Cloud (FiCloud)*, Aug. 2016, pp. 84–90.
[17] Y. Hamid, Balasaraswathi, V. Ranganathan, L. Journaux, and M. Sugumaran, "Benchmark datasets for network intrusion detection: A review," *Int. J. Netw. Secur.*, vol. 20, no. 4, pp. 645–654, 2018.
[18] P. Mishra, V. Varadharajan, U. Tupakula, and E. S. Pilli, "A detailed investigation and analysis of using machine learning techniques for intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 686–728, 1st Quart., 2019.
[19] J. P. Anderson, "Information security in a multi-user computer environment," in *Advances in Computers*, vol. 12. Amsterdam, The Netherlands: Elsevier, 1972, pp. 1–36.
[20] J. P. Anderson, "Computer security threat monitoring and surveillance," Anderson Co., Fort Washington, PA, USA, Tech. Rep., 1980. [Online]. Available: https://csrc.nist.gov/csrc/media/publications/conference-paper/1998/10/08/proceedings-of-the-21st-nissc-1998/documents/early-cs-papers/ande80.pdf
[21] G. Bruneau, "The history and evolution of intrusion detection," SANS Inst., Bethesda, MD, USA, Tech. Rep., 2001, vol. 1, doi: 10.13130/b-biagianti_phd2017-02-10.
[22] R. Zuech, T. Khoshgoftaar, and R. Wald, "Intrusion detection and big heterogeneous data: A survey," *J. Big Data*, vol. 2, no. 3, pp. 1–41, Dec. 2015.
[23] M. Sazzadul Hoque, M. A. Mukit, and M. A. N. Bikas, "An implementation of intrusion detection system using genetic algorithm," 2012, *arXiv:1204.1336*.
[24] C. V. Anchugam and K. Thangadurai, "Classification of network attacks and countermeasures of different attacks," in *Network Security Attacks and Countermeasures*. Hershey, PA, USA: IGI Global, 2016, pp. 115–156.
[25] A. A. Ghorbani, W. Lu, and M. Tavallaee, "Network attacks," in *Network Intrusion Detection and Prevention*. Boston, MA, USA: Springer, 2010, pp. 1–25.
[26] M. Pihelgas, "A comparative analysis of open-source intrusion detection systems," Tallinn Univ. Technol., Univ. Tartu, Tallinn, Estonia, Tech. Rep., 2012. [Online]. Available: http://mauno.pihelgas.eu/files/Mauno_Pihelgas-A_Comparative_Analysis_of_OpenSource_Intrusion_Detection_Systems.pdf
[27] M. Norton, "Optimizing pattern matching for intrusion detection," Sourcefire, Columbia, MD, USA, Tech. Rep., 2004. [Online]. Available: https://snort-org-site.s3.amazonaws.com/production/document_files/files/000/000/085/original/OptimizingPathernMatchingForIDS.pdf
[28] S. Patil, P. B. Rane, P. S. Kulkarni, and B. B. Meshram, "Snort, BRO, NetSTAT, emerald and SAX2: A comparison," *Int. J. Adv. Res. Comput. Sci.*, vol. 3, no. 2, pp. 317–323, 2012.
[29] V. R. Balasaraswathi, M. Sugumaran, and Y. Hamid, "Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms," *J. Commun. Inf. Netw.*, vol. 2, no. 4, pp. 107–119, Dec. 2017.
[30] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," *Comput. Secur.*, vol. 86, pp. 147–167, Sep. 2019.
[31] N. Sainis, D. Srivastava, and R. Singh, "Feature classification and outlier detection to increased accuracy in intrusion detection system," *Int. J. Appl. Eng. Res.*, vol. 13, no. 10, pp. 7249–7255, 2018.
[32] J. McHugh, "Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 4, pp. 262–294, Nov. 2000.
[33] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," *IEEE Access*, vol. 7, pp. 82512–82521, 2019.
[34] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, Nov. 2015, pp. 1–6.

[35] A. A. Shah, Y. D. Khan, and M. A. Ashraf, "Attacks analysis of TCP and UDP of UNCW-NB15 dataset," *VAWKUM Trans. Comput. Sci.*, vol. 8, no. 1, pp. 48–54, 2020.
[36] C. Wheelus, T. M. Khoshgoftaar, R. Zuech, and M. M. Najafabadi, "A session based approach for aggregating network traffic data—The SANTA dataset," in *Proc. IEEE Int. Conf. Bioinf. Bioeng.*, Nov. 2014, pp. 369–378.
[37] S. Soheily-Khah, P. Marteau, and N. Béchet, "Intrusion detection in network systems through hybrid supervised and unsupervised mining process—A detailed case study on the ISCX benchmark dataset," in *Proc. 1st Int. Conf. Data Intell. Secur. (ICDIS)*. South Padre Island, TX, USA: IEEE, 2018, pp. 219–226, doi: 10.1109/ICDIS.2018.00043.
[38] A. Verma and V. Ranga, "On evaluation of network intrusion detection systems: Statistical analysis of CIDDS-001 dataset using machine learning techniques," *Pertanika J. Sci. Technol.*, vol. 26, no. 3, pp. 1307–1332, 2018.
[39] N. Thapa, Z. Liu, D. B. Kc, B. Gokaraju, and K. Roy, "Comparison of machine learning and deep learning models for network intrusion detection systems," *Future Internet*, vol. 12, no. 10, p. 167, Sep. 2020.
[40] N. Moustafa and J. Slay. (2016). *The UNSW-NB15 Data Set Description*. Accessed: May 10, 2020. [Online]. Available: https://www.unsw.adfa.edu.au/unsw-canberracyber/cybersecurity/ADFA-NB15-Datasets/
[41] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, "Cost-based modeling for fraud and intrusion detection: Results from the JAM project," in *Proc. DARPA Inf. Survivability Conf. Expo. (DISCEX)*, vol. 2, Jan. 2000, pp. 130–144.
[42] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. IEEE Symp. Comput. Intell. Secur. Defense Appl.*, Jul. 2009, pp. 1–6.
[43] University of New Brunswick. *NSL-KDD Data-Set for Network-Based Intrusion Detection Systems*. [Online]. Available: https://www.unb.ca/cic/datasets/nsl.html
[44] V. Jyothsna, V. R. Prasad, and K. M. Prasad, "A review of anomaly based intrusion detection systems," *Int. J. Comput. Appl.*, vol. 28, no. 7, pp. 26–35, 2011.
[45] J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, and K. Nakao, "Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation," in *Proc. 1st Workshop Building Anal. Datasets Gathering Exper. Returns Secur.*, 2011, pp. 29–36.
[46] J. Zhang, R. Perdisci, W. Lee, U. Sarfraz, and X. Luo, "Detecting stealthy P2P botnets using statistical traffic fingerprints," in *Proc. IEEE/IFIP 41st Int. Conf. Dependable Syst. Netw. (DSN)*, Jun. 2011, pp. 121–132.
[47] D. Midi, A. Rullo, A. Mudgerikar, and E. Bertino, "Kalis—A system for knowledge-driven adaptable intrusion detection for the Internet of Things," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 656–666.
[48] N. Ben-Asher and C. Gonzalez, "Effects of cyber security knowledge on attack detection," *Comput. Hum. Behav.*, vol. 48, pp. 51–61, Jul. 2015.
[49] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symp. Secur. Privacy*, May 2010, pp. 305–316.
[50] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016.
[51] X. Liang and Y. Xiao, "Studying bio-inspired coalition formation of robots for detecting intrusions using game theory," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 683–693, Jun. 2010.
[52] S. S. Soniya and S. M. C. Vigila, "Intrusion detection system: Classification and techniques," in *Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT)*, Mar. 2016, pp. 1–7.
[53] S. Jose, D. Malathi, B. Reddy, and D. Jayaseeli, "A survey on anomaly based host intrusion detection system," *J. Phys., Conf. Ser.*, vol. 1000, Apr. 2018, Art. no. 012049.
[54] J. Viinikka, H. Debar, L. Mé, and R. Séguier, "Time series modeling for IDS alert management," in *Proc. ACM Symp. Inf., Comput. Commun. Secur. (ASIACCS)*, 2006, pp. 102–113.
[55] J. Viinikka, H. Debar, L. Mé, A. Lehikoinen, and M. Tarvainen, "Processing intrusion detection alert aggregates with time series modeling," *Inf. Fusion*, vol. 10, no. 4, pp. 312–324, Oct. 2009.
[56] F. Delgosha and F. Fekri, "Threshold key-establishment in distributed sensor networks using a multivariate scheme," in *Proc. 25th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2006, pp. 1–12.
[57] S. Sarasamma and Q. A. Zhu, "Min-max hyperellipsoidal clustering for anomaly detection in network security," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 36, no. 4, pp. 887–901, Aug. 2006.
[58] H. Wang, J. Gu, and S. Wang, "An effective intrusion detection framework based on SVM with feature augmentation," *Knowl.-Based Syst.*, vol. 136, pp. 130–139, Nov. 2017.
[59] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51, no. 12, pp. 3448–3470, Aug. 2007.
[60] J. J. Treinen and R. Thurimella, "A framework for the application of association rule mining in large intrusion detection infrastructures," in *Proc. 9th Int. Conf. Recent Adv. Intrusion Detection*, 2006, pp. 1–18.
[61] L. Tan and T. Sherwood, "A high throughput string matching architecture for intrusion detection and prevention," in *Proc. 32nd Int. Symp. Comput. Archit. (ISCA)*, vol. 33, 2005, pp. 112–122.
[62] L. Tan, B. Brotherton, and T. Sherwood, "Bit-split string-matching engines for intrusion detection and prevention," *ACM Trans. Archit. Code Optim.*, vol. 3, no. 1, pp. 3–34, Mar. 2006.
[63] J. Zhu, Y. Huang, and H. Wang, "A formal descriptive language and an automated detection method for complex events in RFID," in *Proc. 33rd Annu. IEEE Int. Comput. Softw. Appl. Conf.*, vol. 1, Jul. 2009, pp. 543–552.
[64] G. Sourek and F. Zelezny, "Efficient extraction of network event types from NetFlows," *Secur. Commun. Netw.*, vol. 2019, pp. 1–18, Feb. 2019.
[65] O. Almomani, "A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms," *Symmetry*, vol. 12, no. 6, p. 1046, Jun. 2020.
[66] F. E. Heba, A. Darwish, A. E. Hassanien, and A. Abraham, "Principle components analysis and support vector machine based intrusion detection system," in *Proc. 10th Int. Conf. Intell. Syst. Design Appl.*, Nov. 2010, pp. 363–367.
[67] W. L. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, "Multi-level hybrid support vector machine and extreme learning machine based on modified k-means for intrusion detection system," *Expert Syst. Appl.*, vol. 67, pp. 296–303, Jan. 2017.
[68] Y. Shi, Y. Tian, G. Kou, Y. Peng, and J. Li, "Network intrusion detection," in *Optimization Based Data Mining: Theory and Applications*. Springer, 2011, pp. 237–241, doi: 10.1007/978-0-85729-504-0_15.
[69] S. X. Wu and W. Banzhaf, "Review: The use of computational intelligence in intrusion detection systems: A review," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 1–35, 2010.
[70] G. Stein, B. Chen, A. S. Wu, and K. A. Hua, "Decision tree classifier for network intrusion detection with GA-based feature selection," in *Proc. 43rd Annu. Southeast Regional Conf. (ACM-SE)*, 2005, pp. 136–141.
[71] S. Chebrolu, A. Abraham, and J. P. Thomas, "Feature deduction and ensemble design of intrusion detection systems," *Comput. Secur.*, vol. 24, no. 4, pp. 295–307, Jun. 2005.
[72] B. Shanmugam and N. B. Idris, "Improved intrusion detection system using fuzzy logic for detecting anamoly and misuse type of attacks," in *Proc. Int. Conf. Soft Comput. Pattern Recognit.*, 2009, pp. 212–217.
[73] A. D. Falke, V. S. Fulsoundar, R. S. Pawase, S. B. Wale, and S. J. Ghule, "Network intrusion detection system using fuzzy logic," *Int. J. Sci. Res. Educ.*, vol. 2, no. 4, pp. 101–111, 2014.
[74] M. Alazab, S. Venkatraman, P. Watters, and M. Alazab, "Zero-day malware detection based on supervised learning algorithms of API call signatures," in *Proc. 9th Australas. Data Mining Conf. (AusDM)*, vol. 121, 2011, pp. 171–182.
[75] A. A. Aburomman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," *Appl. Soft Comput.*, vol. 38, pp. 360–372, Jan. 2016.
[76] W. Wang and R. Battiti, "Identifying intrusions in computer networks with principal component analysis," in *Proc. 1st Int. Conf. Availability, Rel. Secur. (ARES)*, 2006, pp. 270–279.
[77] H. H. Pajouh, R. Javidan, R. Khayami, A. Dehghantanha, and K.-K. R. Choo, "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 7, no. 2, pp. 314–323, Apr./Jun. 2019.
[78] P. Fogla and W. Lee, "Evading network anomaly detection systems: Formal reasoning and practical techniques," in *Proc. 13th ACM Conf. Comput. Commun. Secur. (CCS)*, 2006, pp. 59–68.
[79] M. A. Aydın, A. H. Zaim, and K. G. Ceylan, "A hybrid intrusion detection system design for computer network security," *Comput. Elect. Eng.*, vol. 35, no. 3, pp. 517–526, 2009.
[80] K. Tabia and P. Leray, "Bayesian network-based approaches for severe attack prediction and handling IDSs' reliability," in *Proc. Int. Conf. Inf. Process. Manage. Uncertainty Knowl.-Based Syst.*, 2010, pp. 632–642.

[81] R. Khanna and H. Liu, "System approach to intrusion detection using hidden Markov model," in *Proc. Int. Conf. Commun. Mobile Comput. (IWCMC)*, 2006, pp. 349–354.
[82] J. Hu, X. Yu, D. Qiu, and H. H. Chen, "A simple and efficient hidden Markov model scheme for host-based anomaly intrusion detection," *IEEE Netw.*, vol. 23, no. 1, pp. 42–47, Jan. 2009.
[83] J. Zhang, M. Zulkernine, and A. Haque, "Random-forests-based network intrusion detection systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 5, pp. 649–659, Sep. 2008.
[84] S. S. S. Sindhu, S. Geetha, and A. Kannan, "Decision tree based light weight intrusion detection using a wrapper approach," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 129–141, 2012.
[85] M. Jianliang, S. Haikun, and B. Ling, "The application on intrusion detection based on k-means cluster algorithm," in *Proc. Int. Forum Inf. Technol. Appl. (IFITA)*, vol. 1, 2009, pp. 150–152.
[86] C.-F. Tsai and C.-Y. Lin, "A triangle area based nearest neighbors approach to intrusion detection," *Pattern Recognit.*, vol. 43, no. 1, pp. 222–229, 2010.
[87] W. Hu, W. Hu, and S. Maybank, "AdaBoost-based algorithm for network intrusion detection," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 577–583, Apr. 2008.
[88] W. Hu, J. Gao, Y. Wang, O. Wu, and S. Maybank, "Online Adaboost-based parameterized methods for dynamic distributed network intrusion detection," *IEEE Trans. Syst., Man, Cybern.*, vol. 44, no. 1, pp. 66–82, Mar. 2014.
[89] C. Gates, J. J. McNutt, J. B. Kadane, and M. I. Kellner, "Scan detection on very large networks using logistic regression modeling," in *Proc. 11th IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2006, pp. 402–408.
[90] Y. Wang, "A multinomial logistic regression modeling approach for anomaly intrusion detection," *Comput. Secur.*, vol. 24, no. 8, pp. 662–674, Nov. 2005.
[91] A. Hassanzadeh and B. Sadeghian, "Intrusion detection with data correlation relation graph," in *Proc. 3rd Int. Conf. Availability, Rel. Secur.*, Mar. 2008, pp. 982–989.
[92] Z. Mingqiang, H. Hui, and W. Qian, "A graph-based clustering algorithm for anomaly intrusion detection," in *Proc. 7th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Jul. 2012, pp. 1311–1314.
[93] J. A. J. E. Indumathi, "A network intrusion detection system using clustering and outlier detection," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 3, no. 2, pp. 975–982, Mar. 2015.
[94] F. J. P. Montalbo and E. D. Festijo, "Comparative analysis of ensemble learning methods in classifying network intrusions," in *Proc. IEEE 9th Int. Conf. Syst. Eng. Technol. (ICSET)*, Oct. 2019, pp. 431–436.
[95] R. H. Gong, M. Zulkernine, and P. Abolmaesumi, "A software implementation of a genetic algorithm based approach to network intrusion detection," in *Proc. 6th Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput. 1st ACIS Int. Workshop Self-Assembling Wireless Netw. (SNPD/SAWN)*, 2005, pp. 246–253.
[96] S. N. Pawar and R. S. Bichkar, "Genetic algorithm with variable length chromosomes for network intrusion detection," *Int. J. Autom. Comput.*, vol. 12, no. 3, pp. 337–342, Jun. 2015.
[97] J. V. Hansen, P. B. Lowry, R. D. Meservy, and D. M. McDonald, "Genetic programming for prevention of cyberterrorism through dynamic and evolving intrusion detection," *Decis. Support Syst.*, vol. 43, no. 4, pp. 1362–1374, Aug. 2007.
[98] K. Faraoun and A. Boukelif, "Genetic programming approach for multi-category pattern classification applied to network intrusions detection," *Int. Arab J. Inf. Technol.*, vol. 4, pp. 237–246, Mar. 2007.
[99] J. Gómez, C. Gil, R. Baños, A. L. Márquez, F. G. Montoya, and M. G. Montoya, "A Pareto-based multi-objective evolutionary algorithm for automatic rule generation in network intrusion detection systems," *Soft Comput.*, vol. 17, no. 2, pp. 255–263, Feb. 2013.
[100] X. Pan and L. Jiao, "A granular agent evolutionary algorithm for classification," *Appl. Soft Comput.*, vol. 11, no. 3, pp. 3093–3105, Apr. 2011.
[101] B.-C. Zhang, G.-Y. Hu, Z.-J. Zhou, Y.-M. Zhang, P.-L. Qiao, and L.-L. Chang, "Network intrusion detection based on directed acyclic graph and belief rule base," *Electron. Telecommun. Res. Inst. J.*, vol. 39, pp. 592–604, Aug. 2017.
[102] C.-H. Tsang and S. Kwong, "Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction," in *Proc. IEEE Int. Conf. Ind. Technol.*, Dec. 2005, pp. 51–56.
[103] W. Feng, Q. Zhang, G. Hu, and J. X. Huang, "Mining network data for intrusion detection through combining SVMs with ant colony networks," *Future Generat. Comput. Syst.*, vol. 37, pp. 127–140, Jul. 2014.
[104] J. Wang, T. Li, and R. Ren, "A real time IDSs based on artificial bee colony-support vector machine algorithm," in *Proc. 3rd Int. Workshop Adv. Comput. Intell.*, Aug. 2010, pp. 91–96.
[105] B. Hajimirzaei and N. J. Navimipour, "Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm," *ICT Exp.*, vol. 5, no. 1, pp. 56–59, Mar. 2019.
[106] E. M. Roopa Devi and R. C. Suganthe, "Improved relevance vector machine (IRVM) classifier for intrusion detection system," *Soft Comput.*, vol. 23, no. 19, pp. 9111–9119, Oct. 2019.
[107] S. A. R. Shah and B. Issac, "Performance comparison of intrusion detection systems and application of machine learning to Snort system," *Future Generat. Comput. Syst.*, vol. 80, pp. 157–170, Mar. 2018.
[108] T. Liu, A.-L. Qi, Y.-B. Hou, and X.-T. Chang, "Feature optimization based on artificial fish-swarm algorithm in intrusion detections," in *Proc. Int. Conf. Netw. Secur., Wireless Commun. Trusted Comput.*, vol. 1, Apr. 2009, pp. 542–545.
[109] V. Hajisalem and S. Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection," *Comput. Netw.*, vol. 136, pp. 37–50, May 2018.
[110] H.-Q. Zhang, J. Sun, and W.-B. Xu, "Anomaly detection approach based on quantum-behaved partical swarm optimization," *Jisuanji Gongcheng yu Yingyong (Comput. Eng. Appl.)*, vol. 43, no. 8, pp. 129–130, 2007.
[111] S. Kalaivani and G. Ganapathy, "Bacterial foraging optimization for enhancing the security in intrusion detection system," *Int. J. Eng. Res. Technol.*, vol. 8, no. 10, pp. 1–8, 2019.
[112] C. O. Sakar and O. Kursun, "A hybrid method for feature selection based on mutual information and canonical correlation analysis," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 4360–4363.
[113] S. W. Na, S. K. Choi, T. W. Lee, and Y. H. Cho, "Clustering algorithm for efficient energy consumption in wireless sensor networks," *J. Korea Soc. Comput. Inf.*, vol. 19, no. 6, pp. 49–59, 2014.
[114] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, pp. 1–22, Dec. 2019.
[115] M. A. Ferrag, L. Maglaras, A. Ahmim, M. Derdour, and H. Janicke, "RDTIDS: Rules and decision tree-based intrusion detection system for Internet-of-Things networks," *Future Internet*, vol. 12, no. 3, p. 44, Mar. 2020.
[116] J. E. Gaffney and J. W. Ulvila, "Evaluation of intrusion detectors: A decision theory approach," in *Proc. IEEE Symp. Secur. Privacy. (S&P)*, May 2001, pp. 50–61.
[117] P. A. R. Kumar and S. Selvakumar, "Distributed denial of service attack detection using an ensemble of neural classifier," *Comput. Commun.*, vol. 34, no. 11, pp. 1328–1341, 2011.
[118] B. G. Atli, Y. Miche, and A. Jung, "Network intrusion detection using flow statistics," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2018, pp. 70–74.
[119] X. Xu and X. Wang, "An adaptive network intrusion detection method based on PCA and support vector machines," in *Proc. 1st Int. Conf. Adv. Data Mining Appl. (ADMA)*, 2005, pp. 696–703.
[120] M. Xiao and D. Xiao, "Alert verification based on attack classification in collaborative intrusion detection," in *Proc. 8th ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw., Parallel/Distrib. Comput. (SNPD)*, vol. 2, Jul. 2007, pp. 739–744.
[121] T. Pietraszek and A. Tanner, "Data mining and machine learning—Towards reducing false positives in intrusion detection," *Inf. Secur. Tech. Rep.*, vol. 10, no. 3, pp. 169–183, Jan. 2005.
[122] S. Raza, L. Wallgren, and T. Voigt, "SVELTE: Real-time intrusion detection in the Internet of Things," *Ad Hoc Netw.*, vol. 11, no. 8, pp. 2661–2674, May 2013.
[123] H. Cavusoglu, B. Mishra, and S. Raghunathan, "The value of intrusion detection systems in information technology security architecture," *Inf. Syst. Res.*, vol. 16, no. 1, pp. 28–46, Mar. 2005.
[124] Y. K. Penya, P. G. Bringas, and A. Zabala, "Advanced fault prediction in high-precision foundry production," in *Proc. 6th IEEE Int. Conf. Ind. Informat.*, Jul. 2008, pp. 1672–1677.
[125] J. Wang, D. Fangand Z. Yang, H. Jiang, X. Chen, T. Xing, and L. Cai, "E-HIPA: An energy-efficient framework for high-precision multi-target-adaptive device-free localization," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 716–729, Mar. 2017.

[126] D. Yu and D. Frincke, "Alert confidence fusion in intrusion detection systems with extended Dempster–Shafer theory," in *Proc. 43rd Annu. Southeast Regional Conf. (ACM-SE)*, 2005, pp. 142–147.
[127] S.-Q. Guo, X.-L. Yang, Y. P. Zeng, L. Xie, and C. Gao, "Survey of the security alerts correlation algorithms," *Comput. Appl.*, vol. 25, no. 10, pp. 2276–2279, 2005.
[128] J. Vykopal, "A flow-level taxonomy and prevalence of brute force attacks," in *Proc. Int. Conf. Adv. Comput. Commun.*, 2011, pp. 666–675.
[129] C. Li and X. Liao, "The impact of hybrid quarantine strategies and delay factor on viral prevalence in computer networks," *Math. Model. Natural Phenomena*, vol. 11, no. 4, pp. 105–119, 2016.
[130] I. Ullah and Q. H. Mahmoud, "A two-level hybrid model for anomalous activity detection in IoT networks," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2019, pp. 1–6.
[131] Y. Zeng, H. Gu, W. Wei, and Y. Guo, "Deep-full-range: A deep learning based network encrypted traffic classification and intrusion detection framework," *IEEE Access*, vol. 7, pp. 45182–45190, 2019.

SATISH KUMAR received the bachelor's and master's degrees in computer application from IGNOU, Delhi, India, in 2004, and the master's degree in computer science and engineering and the M.Phil. degree in computer science from CDLU, Sirsa, India, in 2006 and 2007, respectively.

Since 2016, he has been a Research Scholar at Shri Mata Vaishno Devi University, Jammu and Kashmir, India. He has authored several national/international research articles along with various presentations at conferences. His research interests are primarily in the areas of network security with emphasis on intrusion detection systems, machine learning, and bio-inspired heuristic. His current research interest includes bio heuristic approaches in the network security systems.

SUNANDA GUPTA received the bachelor's degree in sciences and the master's degree in computer applications from the University of Jammu, and the Ph.D. degree in computer science engineering from the Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, Jammu and Kashmir, India.

She is currently working as an Assistant Professor with the Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, with more than 13 years of teaching experience. She has authored several research articles in international journals of repute. Her research interests include combinatorial optimization problems, genetic algorithms, and image processing. Besides having presented papers in several international national conferences, she has been invited as an expert to various international conferences as a paper reviewer/program technical committee member.

SAKSHI ARORA is currently working as an Assistant Professor with the Department of Computer Science and Engineering, Shri Mata Vaishno Devi University, Jammu and Kashmir, India, with more than 12 years of teaching experience. She has associated herself in guiding several undergraduate and postgraduate students in their projects and is currently providing Ph.D. supervision to four research scholars. She has authored several research articles in international journals of repute. Her research interests include machine learning, soft computing, and image processing.

Dr. Arora is also affiliated with international societies, like IEEE, ACM, ACM Digital Library, and IETE. Besides having presented papers in several international/national conferences, she has been invited as an expert to various international conferences as a paper reviewer/program technical committee member. She has delivered lectures in various institutions/universities and has also participated in various training programs and attended several workshops. She is also actively involved in institutional activities, like organizing conferences/workshops.