IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 8, 2018, accepted September 7, 2018, date of publication September 28, 2018, date of current version October 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872784

Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review

FADI SALO[1], MOHAMMADNOOR INJADAT[1], ALI BOU NASSIF[1,2], ABDALLAH SHAMI[1], AND ALEKSANDER ESSEX[1]
[1]Department of Electrical and Computer Engineering, Western University, London, ON N6A 5B9, Canada
[2]Electrical and Computer Engineering Department, University of Sharjah, Sharjah 27272, United Arab Emirates

Corresponding author: Fadi Salo (fsalo@uwo.ca)

ABSTRACT The continued ability to detect malicious network intrusions has become an exercise in scalability, in which data mining techniques are playing an increasingly important role. We survey and categorize the fields of data mining and intrusion detection systems, providing a systematic treatment of methodologies and techniques. We apply a criterion-based approach to select 95 relevant articles from 2007 to 2017. We identified 19 separate data mining techniques used for intrusion detection, and our analysis encompasses rich information for future research based on the strengths and weaknesses of these techniques. Furthermore, we observed a research gap in establishing the effectiveness of classifiers to identify intrusions in modern network traffic when trained with aging data sets. Our review points to the need for more empirical experiments addressing real-time solutions for big data against contemporary attacks.

INDEX TERMS Intrusion detection system, real-time detection, data mining, network security.

I. INTRODUCTION

Detecting malicious network intrusions has been a subject of study for decades. As data scientists can appreciate, however, when the scale of a problem grows by an order of magnitude, existing approaches often are no longer effective; the problem is sufficiently different that it requires a new solution. As the volume of network traffic has grown through orders of magnitude, the field of intrusion detection has had to re-invent itself around big data techniques.

An intrusion detection system (IDS) monitors either networks or other systems for malicious or anomalous behaviors. Complementing preventative technologies such as firewalls, strong authentication, and user privilege [1], IDSs have become an essential part of enterprise IT security management [2]. They are typically classified as either *misuse-based* or *anomaly-based* systems [3]. Data Mining (DM) techniques are increasingly being used to identify attacks, anomalies or intrusions in a protected network environment [4]. DM can be defined as ‘‘the process of discovering interesting patterns in databases that are useful in decision making’’ [5]. On the other hand, machine learning is the attempt to ‘‘automate the process of knowledge acquisition from examples’’ [5].

Despite a large DM component in the IDS literature, we noticed few papers deployed IDSs in the context of online (real-time) detection, suggesting more research is needed to improve their performance. As shown in Table 5 of the Appendix which lists the selected articles in this survey (P# is a unique identification number for each paper), many online IDSs classify network traffic into two categories only: *normal* and *attack* (P1, P4, P26, and P56). Some are able to classify one attack type (P36, P45, P65, P74, P78, and P86), or two (P20, P25, P29, P64 and P66), while others require manual applications of expert knowledge to classify attacks. Some used only one detection method, such as accuracy or detection rate (P5, P6, P25, and P68), and did not factor in the testing time as an evaluation metric, which is a critical feature of an IDS, but especially so in the big data context.

Several surveys have been conducted in related domains. Khalilian *et al.* [6] conducted a survey of IDS shortcomings, challenges, and solutions. Denatious and John [7] presented a survey identifying DM techniques applied in IDSs to classify the known and unknown attack patterns. Similarly, Injadat textitet al. [8], discussed DM techniques in the social media context. A systematic literature review was conducted to identify distributed denial of service (DDoS) attacks threatening the existence of cloud-assisted WBANs [9]. Subaira and Anitha [10] compared the advantages and disadvantages of the implemented DM techniques in IDSs. The use of similarity and distance measures within the network intrusion anomaly detection research was presented by Weller-Fahy *et al.* [11]. Few surveys, however, have been conducted in this area without giving full justification for using DM techniques in IDSs.

To the best of our knowledge, there is no systematic literature review (SLR) concentrating on DM techniques actually implemented in the IDS literature, which has motivated this work. Articles were carefully considered and selected with regards to: (i) the DM techniques used in intrusion detection, (ii) attack types that different IDS implementations detect, (iii) evaluation metrics used in the empirical studies of the IDS, (iv) characteristics of the datasets used to train and evaluate IDSs, and (v) the strengths and weaknesses of DM techniques used.

The remainder of this paper is divided into five sections: Section II describes our methodology. Section III lists and illustrates the results. Section IV addresses the limitations of the methodology, and Section V contains a discussion and suggestions for future work.

II. METHODOLOGY

Our methodology was informed by guidelines proposed by Kitchenham and Stuart [12], which consists of three primary phases: *planning*, *conducting*, and *reporting*. Each phase has specific and distinct steps. A crucial step of the planning phase is to create a review protocol. The protocol should: (I) identify research questions, (II) create a search strategy, (III) define the study selection criteria, (IV) develop quality assessment rules, (V) state the data extraction strategies that will be used, and (VI) determine how the extracted data will be synthesized. Fig. 1 illustrates this research methodology. The following subsections (II-A–II-F) present a detailed description of the proposed protocol.

A. RESEARCH QUESTIONS & SEARCH STRATEGY

1) RESEARCH QUESTIONS

Our main objective is to analyze the DM techniques and implementations that were used in intrusion detection systems literature from 2007 to 2017 inclusive. With that in mind, the following research questions were developed:

1) Which classical or novel DM techniques have been used in the IDS research?
2) What types of attacks do various IDS implementations detect?
3) What are the most common evaluation metrics used in IDS literature?
4) What are the characteristics of the datasets/data sources most commonly used in IDS research?
5) What are the strengths or weaknesses being addressed in the implemented DM techniques in IDSs?

2) SEARCH STRATEGY

Because there are numerous papers related to this research area, we adopted the following guidelines to narrow our search:

- Search terms should be derived from our research questions.
- Search for DM techniques were chosen from the top 10 most common techniques [13].
- Additional search terms were created as a workaround to the synonyms or spelling variants problem.
- Boolean logic was added in the form of search operators (AND, OR, quotations, parentheses) to make the search results more relevant.

We performed numerous searches, but the search criteria that yielded the most relevant results were:

- ‘‘Data mining’’ AND (‘‘Intrusion detection’’ OR ‘‘IDS’’)
- (‘‘Intrusion detection’’ OR ‘‘IDS’’) AND ((‘‘C4.5’’ OR ‘‘C5’’) OR ‘‘AdaBoost’’ OR (‘‘k-Means’’ OR (‘‘SVM’’ OR ‘‘support vector machine’’) OR ‘‘Apriori’’ OR ‘‘Naive Bayes’’ OR ‘‘CART’’ OR ‘‘Expectation Maximization’’ OR ‘‘PageRank’’ OR ‘‘kNN’’ OR ‘‘k-nearest neighbours ’’ OR ‘‘Fuzzy’’)

3) DIGITAL RESOURCES

An essential step in any SLR is to identify the sources/digital libraries that will be used to retrieve the related articles. The digital libraries were chosen in this work include: Google Scholar (Springer, Elsevier,etc.), the Institute of Electrical and Electronics Engineers (IEEE) Library, and the Association for Computing Machinery (ACM) Digital Library. Table 6 of the Appendix summarizes the number of papers selected from each resource.

B. STUDY SELECTION

Our initial search returned 931 articles. Because many articles were duplicated, were of insufficient quality or were not related to our research questions, we performed additional filtering as shown in Fig. 1. Filtration was conducted concurrently and independently by the lead authors to reduce potential bias and identify discrepancies. Filtration accomplished the following:

1) Any duplicate documents were removed.
2) Inclusion and exclusion criteria were applied to determine related articles and discard irrelevant ones.
3) Quality assessment was performed to ensure that we included only high-quality papers.

Our criteria for inclusion included:

1) Articles utilizing DM techniques in IDSs.
2) The most recent version/edition of a paper.
3) Articles published between January 2007 and September 2017.

Our criteria for exclusion included:

1) Articles that involve DM, but were not related to IDSs.
2) Articles that involve IDSs, but are not related to DM.
3) Papers that are not categorized as peer-reviewed journal or conference papers.

After this filtration, 315 journal and conference papers remained as candidates. This number was further reduced by a quality assessment step outlined in the following section.

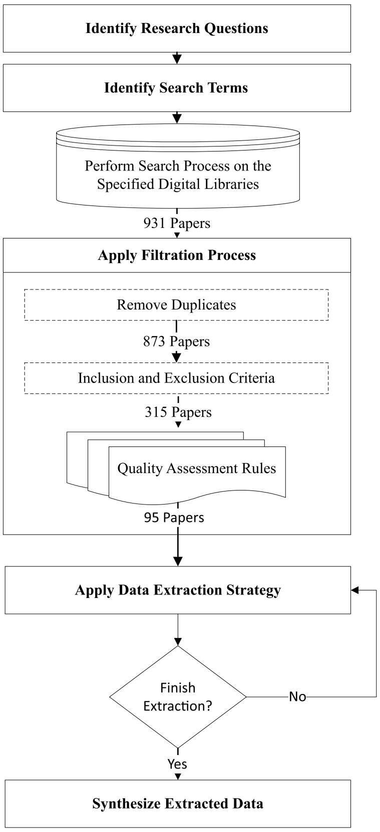

Identify Research Questions
Identify Search Terms
Perform Search Process on the Specified Digital Libraries
931 Papers
Apply Filtration Process
Remove Duplicates
873 Papers
Inclusion and Exclusion Criteria
315 Papers
Quality Assessment Rules
95 Papers
Apply Data Extraction Strategy
Finish Extraction?
No
Yes
Synthesize Extracted Data

FIGURE 1. Research methodology.

C. QUALITY ASSESSMENT (QA)

To evaluate the quality of the initially selected articles, we developed six QA rules to determine how relevant the articles were to our research, in which each QA indicator had a weight of 1. Each indicator was scored as follows: "fully answered" = 1, "above average" = 0.75, "average" = 0.5, "below average" = 0.25, "not answered" = 0. An overall quality score ranging from 0 to 6 was assigned to each article by summing the individual indicator scores. An over quality score of 3 or above was considered a passing grade, and papers meeting or exceeding this threshold were included in this review. The quality indicators were as follows:

1) Are the objectives of the research outlined in sufficient detail?
2) Are any DM approaches explained with sufficient detail?
3) Is the experiment well suited to the scope of this literature review?
4) Is the IDS experiment conducted on datasets/data sources of sufficient size and quality?
5) Are the results generated by the respective DM techniques properly measured and evaluated?
6) Are the experiment' s results and findings clearly reported?

D. FILTRATION PROCESS OUTPUT

After applying these criteria, a total of 95 articles were selected. Table 5 of the Appendix shows the selected articles and the research questions they answer. Table 7 of the Appendix shows the QA scores assigned to the selected articles.

E. DATA EXTRACTION PROCESS

The objective of this step is to provide an answer to the research questions for each paper in a semi-structured way. We used a coordinated effort between the two lead authors to extract and check the data independently, with a comparison made between respective results. In case of disagreement, an in-person meeting was conducted to reach consensus. The data extraction form used in this SLR is given in Table 1.

TABLE 1. Data extraction form.

Data Item	Value
Paper ID	
Extractor	
Checker	
Journal/Conference Name	
Publisher	
Paper Title	
Research Question 1	
Research Question 2	
Research Question 3	
Research Question 4	
Research Question 5	
Additional Notes	

Extraction posed a challenge due to the fundamental unstructured nature of the data. For example, researchers would use different terminologies for related techniques, such as "J48" or "C4.5." Some articles used different names for the same DM technique, such as *Nearest Neighbor*, *k-NN*, and *KNN* when referring to the k-nearest neighbor method, or *Naive Bayes*, *Naïve Bayes*, and *Bayesian Network* when referring to the Naïve Bayes classifier.

To simplify the tracing procedure of the data extraction process results, a binary model was used to show whether each paper answers the research questions or not. This is shown in Table 5 of the Appendix.

F. SYNTHESIS OF THE EXTRACTED DATA

Various methods were discussed in [12] to synthesize the data extracted from the selected articles. To answer our research questions (RQs) given in the following section, we used the *narrative synthesis* method for RQs 1 and 2, the *binary outcome* method for RQs 3 and 4, and the *reciprocal translation* method to answer RQ 5. *Narrative synthesis* is the process of tabulating the results with respect to the research question and visualizing them using techniques such as bar charts and pie charts to enhance the result presentation [12]. On the other hand, *reciprocal translation* refers to the process of providing additive summaries through the translation of similar concepts into a unifying concept [14]. In this SLR, we translate the strength/weakness of a DM or machine learning technique retrieved from different studies with similar meanings into a unified description of the strength/weakness.

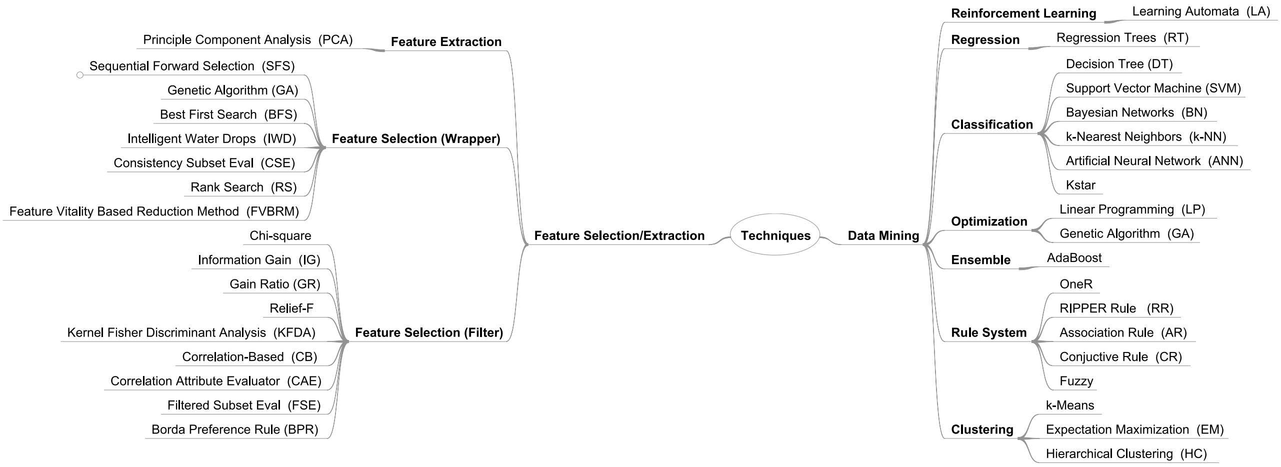

Techniques
Feature Selection/Extraction
Feature Extraction
Principle Component Analysis (PCA)
Feature Selection (Wrapper)
Sequential Forward Selection (SFS)
Genetic Algorithm (GA)
Best First Search (BFS)
Intelligent Water Drops (IWD)
Consistency Subset Eval (CSE)
Rank Search (RS)
Feature Vitality Based Reduction Method (FVBRM)
Feature Selection (Filter)
Chi-square
Information Gain (IG)
Gain Ratio (GR)
Relief-F
Kernel Fisher Discriminant Analysis (KFDA)
Correlation-Based (CB)
Correlation Attribute Evaluator (CAE)
Filtered Subset Eval (FSE)
Borda Preference Rule (BPR)
Data Mining
Reinforcement Learning
Learning Automata (LA)
Regression
Regression Trees (RT)
Classification
Decision Tree (DT)
Support Vector Machine (SVM)
Bayesian Networks (BN)
k-Nearest Neighbors (k-NN)
Artificial Neural Network (ANN)
Kstar
Optimization
Linear Programming (LP)
Genetic Algorithm (GA)
Ensemble
AdaBoost
Rule System
OneR
RIPPER Rule (RR)
Association Rule (AR)
Conjuctive Rule (CR)
Fuzzy
Clustering
k-Means
Expectation Maximization (EM)
Hierarchical Clustering (HC)

FIGURE 2. Taxonomy of techniques observed in the literature.

III. RESULTS AND DISCUSSION

In the following subsections, the findings of this SLR will be presented and discussed for each RQ.

A. RQ1: WHICH CLASSICAL OR NOVEL DM TECHNIQUES HAVE BEEN USED IN THE IDS RESEARCH?

As a fundamental question of this body of literature, we sought to identify the most commonly used DM techniques in the IDS literature. We reviewed DM techniques implemented at all stages of an experiment, such as during the feature selection stage, extraction stage, etc. The review resulted in 92 papers out of the 95 considered to have addressed this research question.

As shown in the right hand side of Fig. 2, we identified 19 different DM techniques that were applied in the development of IDS models based on their application. These techniques can be decomposed into seven categories including:

- Reinforcement learning: determine current action according to past experience.
- Regression: predict a numeric/continuous value.
- Classification: predict a category based on a given dataset.
- Optimization: determine an optimum or a satisfactory solution based in various solutions executed iteratively.
- Ensemble: combine a set of classifiers' predictions into a single decision based on their weighted vote.
- Rule system: use a set of *if-then* rules for classification.
- Clustering: group a set of data into a set of meaningful sub-classes (clusters).

Fig. 3 shows the prevalence of each technique. The most prevalent DM techniques were SVM (48 papers), DT (44 papers), BN (32 papers), and ANN (26 papers). It is worth mentioning that many of the papers considered within this work proposed an ensemble learning approach that combines many of the discussed DM techniques. In such cases, the base techniques used are identified. More findings about the implemented techniques are discussed in Section III-E.

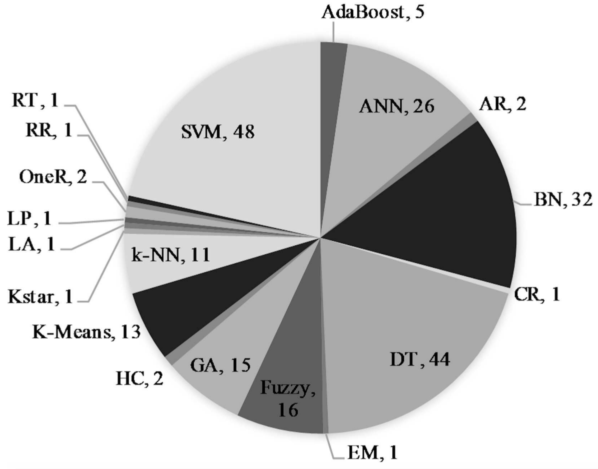


FIGURE 3. Comparisons of DM techniques.

Table 8 of the Appendix provides additional detail about the distribution of the DM techniques by publication year.

Feature selection/extraction was broadly found in the literature, and is an important step to discard irrelevant information, which, in turn, increases the detection accuracy and computational efficiency of the proposed models. Fig. 4 demonstrates 18 different feature selection/extraction techniques were applied based on the data extracted for this RQ. According to the figure, GA, IG, PCA, and CB were the most popular techniques with frequencies of 9,8,7, and 4, respectively. In addition to the reported techniques, this RQ summarized some novel feature selection/extraction techniques and frameworks (some of which incorporate hybrid techniques) extracted from 8 studies, see Table 2. Even though feature selection/extraction is a crucial step, a surprising number of papers did not address it in a robust way. For instance, P5, P7, P55, P65, P68, and P94 applied feature selection/extraction in their methodology; however, their techniques were not discussed. Similarly, P20, P85, and P87 did not use a formal

TABLE 2. Novel feature selection/extraction techniques/frameworks.

Technique	Main Features	Year	Ref
NA	Genetic Wrapper-based: • Reduce dimensionality by applying genetic feature selection wrapper. • Explore network traffic data to pursuit near-optimal feature subset. • Improve classification accuracy, reduce computational complexity, and improve the generalization capability	2007	P3
GFR	Gradually Feature Removal (GFR): • Improve wrapper-based feature reduction. • Remove features gradually and evaluate the new features' combination by a classifier to determine the effectiveness of the remaining features.	2012	P23
FASVFG	Four-Angle-Star Based Visualized Feature Generation Approach (FASVFG): • Allows users to identify each class to which the data belongs visually. • Improves the feature generation rule from visual graphs.	2014	P34
CANN	Center and Nearest Neighbor (CANN): • Transform original feature into dimensional distance-based feature • The distance calculated for each data sample based on cluster center and nearest neighbor.	2015	P53
TVCPSO	Time-Varying Chaos Particle Swarm Optimization (TVCPSO): • Perform parameter setting and feature selection simultaneously.	2016	P75
NA	k-Means based: • Based on a local search algorithm to overcome the computationally hard optimization problems to obtain optimal feature subsets.	2016	P78
RSHGT	Rough Set Theory and Hypergraph (RSHGT): • Apply hypergraph approach to identify the optimum order relations between the features. • Reduce computational complexity and improve the classification accuracy.	2017	P80
NA	Entropy Measure-based: • Select important features based on entropy a measure in order to find the uncertainty/randomness of features.	2017	P89

NA: name not available

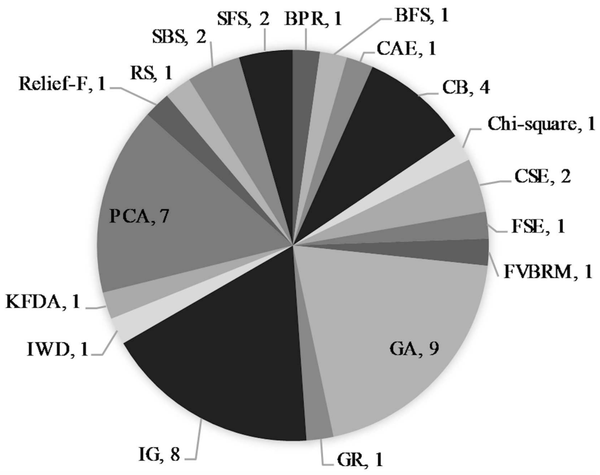


FIGURE 4. Feature selection techniques.

methodology, relying instead on either expert knowledge, or on trial and error. Surprisingly, we found 49 articles had not considered this step at all. Note that this work focused on the base feature selection techniques used as part of the hybrid feature selection approaches proposed by many of the discussed works.

B. RQ2: WHAT TYPES OF ATTACKS DO VARIOUS IDS IMPLEMENTATIONS DETECT?

In this RQ, we attempt to enumerate the various attack types that the DM techniques in IDS is designed to (80 papers in total). We identified twenty-two differing attack types in the literature, which fall into four main categories [15]:

1) *Denial of Service Attacks (DoS)*: Denial of service attacks render a computational or network resource inaccessible to its intended users through flooding the victim with excessive, useless requests. Examples include *Smurf*, *Neptune* (SYN floods), *Ping of death*, *Mail bombs*, and *UDP storms*.
2) *Remote to User Attacks (R2L)*: An attacker attempts to remotely exploit vulnerabilities in a target system to obtain privileges of a local user. Examples include *Xlock, Xnsoop, Sendmail, Phf,* and *Ftp-write*.
3) *User to Root Attacks (U2R)*: The attacker begins with some basic level of privilege on the target system (usually a guest) and attempts to escalate that privilege to that of the root user through a kernel or application vulnerability. Examples include *Eject*, *Ffbconfig*, *Fdformat*, and *Xterm*.
4) *Probing*: Probing involves the scanning of a network or machine to gather information for the purposes of discovering potential vulnerabilities. Probing is not an attack *per se*, but can be an indicator thereof. Examples include *Ipsweep*, *Saint*, *MScan*, and *Nmap*.

Based on the data extracted for this RQ, we found that the majority of IDSs in the selected studies were trained and designed using virtualized/benchmark datasets to detect attacks in an off-line mode. Few considered IDSs in the context of online (real-time) detection. Given the fact that network traffic is increasing dramatically each year, real-time network monitoring remains an elusive computational challenge for researchers. The distribution of attack types based on the utilized datasets is shown in Fig. 5. We also observed in this figure that most studies were able to distinguish the four major attack types in an off-line environment. In contrast, few experiments were implemented for real-time detection mode. Of those operating in an online mode, the attack type was not fully classified, rather individual

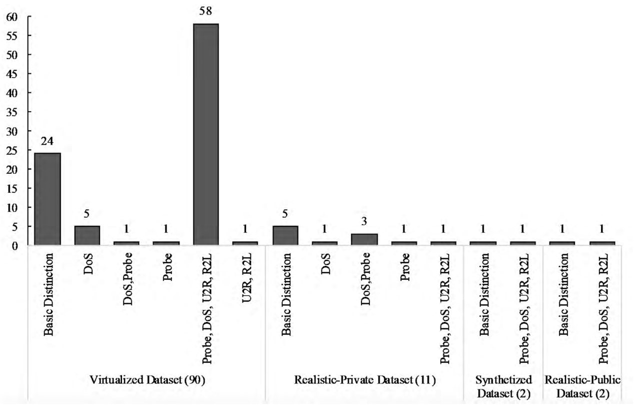


FIGURE 5. Attack classification.

packets were classified as being basic distinction (either normal or malicious).

C. RQ3: WHAT ARE THE MOST COMMON EVALUATION METRICS USED IN IDS LITERATURE?

We enumerate the most commonly used evaluation metrics in IDSs. We found 17 different existing evaluation metrics, and one novel technique in the data extracted for this RQ. These metrics were categorized by detection efficiency and computational performance. A total of 59 papers discussed the evaluation metrics.

1) DETECTION PERFORMANCE/EFFICIENCY METRICS

This category encompasses all measurements used by researchers to validate the results obtained in their DM models in terms of malicious or normal action. Each mathematical representation of these measurements involves any of: true positive (*TP*), true negative (*TN*), false positive (*FP*), or false negative (*FN*) [3]. In total, 13 metrics belong to this group:

1) *Confusion Matrix (CM):* A type of matrix allowing easy comparison between predicted and actual classes. Table 3 demonstrates an example of 2 × 2 CM.

2) *Accuracy (Acc)*: The percentage of correctly classified/ predicted instances in the testing dataset, calculated by

$$Acc = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{1}$$

3) *Error Rate (ER)*: The percentage of all predictions that were incorrectly classified:

$$ER = 1 - Acc \tag{2}$$

4) *True Positive Rate (TPR)*: Also known as sensitivity, detection rate, or *recall*, it is the intrusions that were correctly classified as an attack, given by:

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

TABLE 3. Confusion matrix (2 × 2 dimensions).

		Prediction	
		Intrusion	Legitimate
Actual	Intrusion	*TP*	*FN*
	Legitimate	*FP*	*TN*

5) *False Positive Rate (FPR)*: Also known as false alarm rate. It is the normal patterns that were falsely classified as an attack, given by:

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

6) *True Negative Rate (TNR)*: Also known as specificity, is the normal patterns that were correctly detected as normal, given by:

$$TNR = 1 - FPR \tag{5}$$

7) *False Negative Rate (FNR)*: It is the intrusions that were falsely detected as normal, given by:

$$FNR = 1 - TPR \tag{6}$$

8) *Precision*: It is the ration of actions correctly classified as *attack*, given by

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

9) *F-measure (FM)*: The harmonic means of recall and precision also known as f-value or f-score:

$$FM = 2 \times \frac{precision \times recall}{precision + recall} \tag{8}$$

10) *Kappa*: Measures the chance of agreement between the predicted and the real classes, given by:

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{9}$$

where p_0 is the observed agreement, and p_e is the expected agreement.

11) *Matthews Correlation Coefficient (MCC)*: Measures the correlation between the predicted results and the real data. MCC value is between [-1,1] as shown in (10), if $MCC = +1$, indicates that the prediction was 100 accurate, where $MCC = -1$, means that the prediction was totally wrong.

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \tag{10}$$

12) *ROC Curve (ROC)*: It is a graphical representation tool that shows the intrusion detection accuracy against the false positive rate. *ROC* is considered one of the powerful metrics used to evaluate the performance of IDSs effectively.

13) *NP Ratio (NPR)*: is a novel evaluation metric proposed by [16], where "*N*" denotes *TN* and "*P*" denotes *TP*. *NPR* pays more attention towards classes distribution in the test dataset to detect any negligibly bias due to

imbalanced issue in the given data. The mathematical representation of *NPR* is calculated as follows:

$$NPR = \frac{TN}{TP} \tag{11}$$

Of course, any individual metric is not an adequate performance indicator of detection efficiency. For instance, accuracy in some cases, skewed dataset, can lead to biased results in the performance indicator [17].

The problem is that if one class has a very small percentage in comparison to the others, the classifier might fail to classify the minority class, although it shows high accuracy [18]. Most of the included papers evaluated their models using several evaluation metrics (see Fig. 6). However, not many of them did it effectively. For instance, (P5, P6, P9, P23, P34, P48, P49, P54, P72, P74, P81, P85, and P90) only relied on the accuracy rate, which is no longer a proper measure for classification with the class imbalance problem [19]. Likewise, (P25 and P68) considered only detection rate as a performance indicator. Therefore, a lot of papers and experiments are bound to have inaccurate results. As shown in Fig. 6, the TPR (85), FPR (69), and accuracy (56) are the most commonly metrics used to measure IDS detection effectiveness.

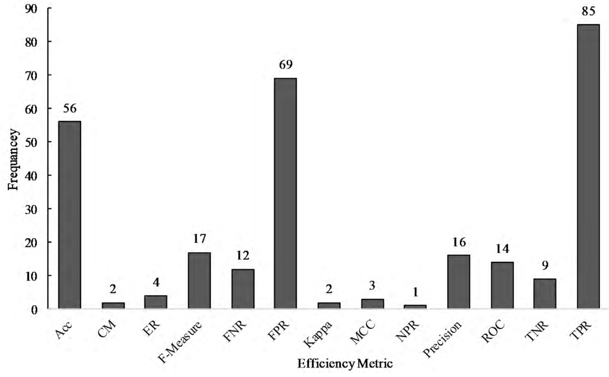


FIGURE 6. Histogram of detection efficiency metrics used in literature.

2) COMPUTATIONAL PERFORMANCE METRICS

This group incorporates metrics utilized to measure the computational performance of IDSs. Overall, 5 different measurements were identified based on the data extracted in this RQ:

1) *Computational Time (CT)*: Also known as worst-case running time, execution time, processing time, or computational complexity, is the time needed to complete a certain task that classifies an action as a normal or malicious.
2) *CPU Utilization (CPU-U)*: Also known as CPU or processor usage, it is the percentage of the CPU load taken by a certain IDS task.
3) *Energy Consumption (EC)*: Also known as power consumption, it is the measurement of the required additional energy to perform a certain IDS task.
4) *Training Time (TRT)*: The time taken to build and train a model to classify an IDS task.
5) *Testing Time (TST)*: The time taken to classify an IDS action as normal or malicious.

Although some of the aforementioned performance metrics are considered critical to ensuring the effectiveness of the IDS [20], we found 63 papers did not adopt any of these metrics in their research. As seen in Fig. 7, only 16 articles considered the testing time in their experiments, and only 6 consider computational time in their systems.

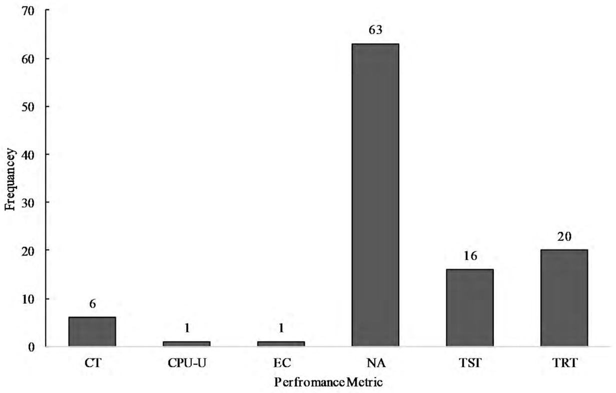


FIGURE 7. Histogram of computational performance metrics used in literature.

To investigate how these evaluation metrics improved over time, we tracked the metrics to measure IDS efficiency. We noticed the lack of consideration given to some metrics in the last decade including, in ascending order, *CM*, *Kappa*, *MCC*, and *ER*. In contrast, as shown in Fig. 8, *TPR*, *FPR*, and *Accuracy* remain the most popular metrics with the number of relevant papers amounting to 85, 69, and 56, respectively. Furthermore, we found that the popularity of *F-Measure*, *precision*, *ROC*, and *TNR* increased substantially after 2012.

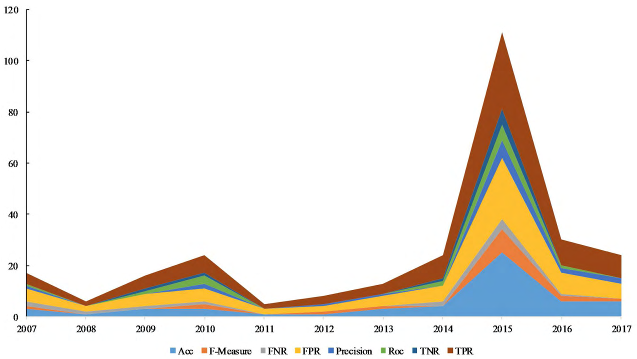


FIGURE 8. Popularity of efficiency metrics over time.

D. RQ4: WHAT ARE THE CHARACTERISTICS OF THE DATASETS/DATA SOURCES MOST COMMONLY USED IN IDS RESEARCH?

We examine the source and characteristics of the data source used to develop intrusion detection in the related papers. Overall, we identified 25 different public and private datasets used in IDSs within the 91 papers addressing this research question. The classification of these datasets is categorized into three categories as follows:

1) Virtualized: datasets artificially generated to meet a certain task, in which many of its features and methods are virtual or abstract. DARPA dataset [21] is a good example used in the field of IDS.
2) Synthesized: dataset generated to satisfy certain needs or specific conditions which may not be available in the available realistic data, i.e. ISCX 2012 [22]. The popularity of these synthetic datasets arises due to the fact that realistic datasets can cause privacy concerns.
3) Realistic: datasets collected from real-world traffic, which can be categorized into public (i.e. Kyoto 2006+ [23]) or private (i.e. blogs, profiles, and network traffic).

Despite the fact that 25 different datasets were identified in this RQ, Fig. 9 shows around 79% of the experiments were adopted DARPA dataset. KDD Cup 1999 [24] and NSL-KDD are both derived from the original DARPA dataset and they account for 56% and 20% of the papers, respectively.

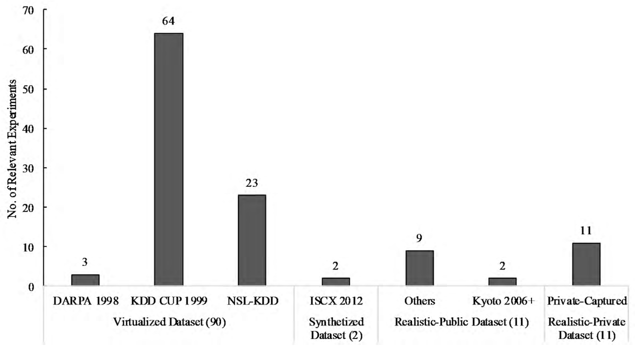


FIGURE 9. Utilized datasets per category.

Other intrusion detection datasets, as shown in the figure, represent approximately 8% of the total experiments, namely A1a, A3a, A4a, Satimage, Vehicle, German, Segment, Svmguide3, Wisconsin-wdbc were used to validate the quality of the proposed intrusion detection approach, additional details about these datasets are presented in [25]. It is worth mentioning that the majority of the experiments are based on the outdated DARPA dataset, which also suffers from inconsistency with regard to the distribution of attacks types [26].

E. RQ5: WHAT ARE THE STRENGTHS OR WEAKNESSES BEING ADDRESSED IN THE IMPLEMENTED DM TECHNIQUES IN IDSs?

We extract the pros and cons of the implemented DM techniques from 36 papers that discussed them. Under this scenario, each relevant observation and conclusion about DM techniques on the relevant papers was included in a single combined list. Afterwards, similar (or equal) observations were grouped, and a table listing techniques' strengths and weaknesses (with their respective sources) was created. The more independent sources an observation or statement had, the more credible or reliable it was scored. Table 4 condenses the knowledge and experiments of numerous authors and studies, providing a helpful cross-section of which DM techniques are applicable to which IDS.

TABLE 4. Strength & weakness of the implemented DM techniques.

Technique		Notes	Ref
SVM	Strength	Most used and overall best classifier.	P1, P21, P30, P33, P86, P91
		Over-fitting can be solved relatively easy by chaining with other algorithms.	P18, P22, P86
		Flexibility when selecting parameters.	P16
		Process feature vectors of high dimensions	P31, P57, P86
	Weakness	Size and dimensionality of dataset affects training complexity considerably.	P18, P91
		Higher false positive than other algorithms make it difficult to use	P1, P31
		Very sensitive to noise	P8, P18
ANN	Strength	Learning effectiveness greatly improves when possessing prior knowledge	P2, P48
		Better pattern recognition than other algorithms	P2, P13, P26
		Self-organizing maps (SOM) do not require labels at input.	P2
	Weakness	Training takes more time in average.	P48
		Lower detection precision for low frequent attacks	P13
DT	Strength	Better in accuracy and false positive rate compared to the static models	P25, P27, P32, P33
		Require less training time	P33
		C5.0 quite robust when missing data, and large numbers of input fields.	P47
BN	Strength	Best algorithm when training data is scarce.	P17, P23, P48
		Average one dependence estimators' algorithm (enhancement version of BN) is useful for large dataset.	P92
		High accuracy and simple.	P46
		Reduce time complexity	P35
GA	Strength	Algorithm's search method is flexible and effective.	P48
		Best algorithm when solution scope is huge.	P4
		Learning classifier system (LCS) can easily adapt to environmental changes.	P10
	Weakness	High resource demand.	P48
k-NN	Strength	Easy to implement	P53
	Weakness	Suffers from the high dimensionality and overfitting.	P79
Fuzzy	Strength	Can be trained effectively with noisy data.	P5
k-Means	Strength	Simple algorithm. Small computational complexity. Suitable for real-time IDS.	P28
	Weakness	Need to specify the number of cluster before starting. Need to input the value of k in advance.	P28
AdaBoost	Strength	Faster than others. Simple Implementation.	P7
	Weakness	Not amenable to incremental learning	P7
HC	Strength	Agglomerative clustering: is a child of HC and do not need know the number of clusters	P83
LA	Strength	Intelligent and adaptive in a random environment.	P87

IV. LIMITATIONS

This study is restricted to journal and conference papers in the field of DM and in IDSs. By applying our search strategy at the first stages of the review, we gained (and therefore excluded) a large number of non-relevant articles. While this ensured that the selected articles suited the research requirements, having additional sources would have further enriched the review. The same idea applies to the quality assessment: because we applied a rigorous QA.

V. CONCLUSIONS

We performed an SLR to investigate DM techniques in the field of IDS. Our motivation was on the relevant empirical studies reported in journals and conferences between the desired period. We manually explored some 873 unique papers returned by the initial search. Overall, 95 relevant papers were chosen after applying our selection criteria. We employed the findings of this SLR to provide an integrated and unified view of DM techniques in IDSs. In addition, our results allow researchers to identify, compare, and evaluate their methods for IDSs from different perspectives including attack types, operation mode, evaluation criteria, datasets, and strength and weakness of the utilized DM techniques. Our conclusions are summarized as follows:

- RQ1: We found the most commonly used DM techniques in IDSs are SVM, DT, BN, and ANN among the 19 different techniques in the related literature. We further enumerated 18 feature selection/extraction techniques, as well as 8 novel techniques from the data extracted in this RQ.
- RQ2: We counted the various attack types, in which the IDSs are designed to prevent, namely DoS, R2L, U2R, and Probe. We also observed that there is a lack of research dealing with real-time IDS as most of the papers designed using virtualized/benchmark datasets to detect attacks in an off-line environment. Of those operating in an online mode, the attack type was not fully classified, rather individual packets were classified as being normal or malicious.
- RQ3: We enumerated 17 different evaluation metrics, as well as a novel technique used as performance indicators in IDSs. The most popular metrics during the last decade are TPR, FPR, and Accuracy. While, on the other hand, the utilization of F-Measure, precision, ROC, and TNR have increased considerably after 2012 in the holistic studies in this SLR. The classification of these metrics categorized based on the detection efficiency metrics to validate the IDS results, and the computational performance metrics to express the empirical measurements used in the context of the computational performance.
- RQ4: We identified 25 different public and private datasets used in IDS. The classification of these datasets were grouped into virtualized, synthesized, and realistic datasets. We noticed that the majority of the proposed IDSs are based on outdated datasets, which suffer from an inconsistency issue with the distribution of the attacks and the lack of the contemporary attack scenarios.

TABLE 5. Applicability of the selected papers to our research questions.

ID	Research questions					Year	Ref	ID	Research questions					Year	Ref
	1	2	3	4	5				1	2	3	4	5		
P1	1	0	1	1	1	2007	[27]	P49	1	1	0	1	0	2015	[28]
P2	1	1	0	1	1	2007	[29]	P50	1	0	0	1	1	2015	[30]
P3	1	1	1	1	0	2007	[31]	P51	1	0	1	1	1	2015	[16]
P4	1	1	0	1	0	2007	[32]	P52	1	1	1	1	1	2015	[33]
P5	1	1	1	1	0	2007	[34]	P53	1	1	1	1	1	2015	[35]
P6	1	1	1	1	0	2007	[36]	P54	1	1	0	1	0	2015	[37]
P7	1	0	1	1	1	2008	[38]	P55	1	1	0	1	0	2015	[39]
P8	1	1	1	1	0	2008	[40]	P56	1	1	0	1	0	2015	[41]
P9	1	1	0	1	0	2008	[42]	P57	1	1	1	1	0	2015	[43]
P10	1	1	0	1	0	2009	[44]	P58	1	1	1	1	1	2015	[45]
P11	1	1	1	1	0	2009	[46]	P59	1	1	1	1	1	2015	[47]
P12	1	1	1	1	0	2009	[48]	P60	1	0	1	1	0	2015	[49]
P13	1	1	1	1	0	2009	[50]	P61	1	1	1	0	0	2015	[51]
P14	1	1	1	1	1	2010	[52]	P62	1	1	0	1	0	2015	[53]
P15	1	1	1	1	0	2010	[54]	P63	1	1	1	1	0	2015	[55]
P16	1	0	0	1	1	2010	[56]	P64	1	1	1	1	0	2015	[57]
P17	1	1	1	1	0	2010	[58]	P65	1	1	1	1	0	2015	[59]
P18	1	1	1	1	0	2010	[60]	P66	1	0	0	1	0	2015	[61]
P19	1	1	0	1	1	2010	[62]	P67	1	0	1	1	0	2015	[63]
P20	1	1	1	1	1	2011	[64]	P68	1	1	1	1	1	2015	[64]
P21	1	1	0	1	0	2011	[65]	P69	1	1	1	1	0	2015	[66]
P22	1	1	1	1	0	2012	[67]	P70	1	1	0	0	1	2015	[68]
P23	1	1	0	1	1	2012	[69]	P71	1	1	1	1	0	2015	[70]
P24	1	1	1	1	1	2012	[71]	P72	1	1	1	1	1	2015	[72]
P25	1	1	1	1	0	2012	[73]	P73	1	1	0	1	0	2015	[74]
P26	1	1	1	0	0	2013	[75]	P74	1	1	1	1	0	2015	[76]
P27	1	1	1	1	1	2013	[77]	P75	1	1	1	1	1	2016	[78]
P28	1	1	1	1	1	2013	[79]	P76	1	1	1	1	0	2016	[79]
P29	1	1	1	1	0	2013	[80]	P77	1	1	0	1	0	2016	[81]
P30	1	1	1	1	0	2013	[82]	P78	1	1	0	1	0	2016	[83]
P31	1	1	1	1	0	2014	[84]	P79	1	0	0	1	1	2017	[85]
P32	1	1	0	1	1	2014	[86]	P80	1	1	1	1	0	2017	[87]
P33	1	1	1	1	1	2014	[88]	P81	1	0	1	1	0	2017	[89]
P34	1	1	0	1	0	2014	[90]	P82	1	1	1	1	1	2017	[91]
P35	1	1	0	1	0	2014	[92]	P83	1	0	1	1	1	2017	[93]
P36	1	1	0	1	0	2015	[94]	P84	1	1	0	1	0	2017	[95]
P37	1	1	0	1	1	2014	[96]	P85	1	1	1	1	0	2017	[97]
P38	1	1	0	1	0	2014	[98]	P86	1	1	0	0	1	2017	[99]
P39	1	1	0	1	0	2014	[100]	P87	1	1	1	1	1	2017	[101]
P40	1	1	1	1	0	2014	[102]	P88	1	0	1	1	0	2017	[103]
P41	1	1	0	1	0	2015	[49]	P89	1	1	1	1	0	2017	[104]
P42	1	1	1	1	1	2015	[25]	P90	1	1	0	1	0	2016	[105]
P43	0	1	1	1	0	2015	[106]	P91	1	0	0	1	1	2016	[107]
P44	0	1	1	1	1	2015	[108]	P92	1	1	0	1	1	2016	[109]
P45	0	1	0	1	0	2015	[110]	P93	1	1	0	1	1	2016	[111]
P46	1	1	1	1	1	2015	[112]	P94	1	1	0	1	0	2016	[113]
P47	1	1	1	1	1	2015	[114]	P95	1	0	1	1	0	2016	[115]
P48	1	0	1	1	0	2015	[116]	**Total**	**92**	**80**	**59**	**91**	**36**		

- RQ5: We summarized the strengths and weaknesses of the utilized DM techniques based on the researchers' experience. This information is rich in breadth and depth, in which it provides a solid foundation that researchers can extend for future research in IDS domain.

Based on the SLR outcomes, we identify some interesting opportunities for future work.

Given that many researchers did not appear to consider or outline their feature selection/extraction strategy,

TABLE 6. Number of papers selected per digital resource.

Publisher	Papers Selected
IEEE	44
Elsevier	26
Springer	10
International Press Corporation (INPRESSCO)	1
MECS	1
International Journal of Scientific & Engineering Research (IJSER)	1
ACM	1
Academy Publisher	1
Binary Information Press	1
International Journal of Computer Science and Network Security, South Korea (IJCSNS)	1
P&R Publishing	1
Iranian Society of Cryptology	1
Scientific Research Publishing	1
Research India Publications	1
International Digital Organization for Scientific Information (IDOSI)	1
Global Journals Inc.	1
Academy Publisher	1
Eoryx Publications	1
Total	**95**

TABLE 7. Quality assurance score of the selected papers.

Score	No. of papers	Paper ID
3	8	P19, P22, P40, P45, P47, P48, P49, P68
3.25	10	P2, P11, P50, P51, P54, P66, P85, P86, P93, P95
3.5	12	P9, P21, P29, P37, P43, P55, P62, P72, P79, P91, P92, P94
3.75	10	P13, P30, P31, P36, P39, P44, P60, P61, P78, P90
4	16	P4, P10, P17, P18, P25, P33, P46, P52, P56, P59, P63, P74, P80, P81, P83, P84
4.25	7	P5, P12, P24, P27, P35, P69, P82
4.5	11	P6, P23, P32, P53, P64, P67, P73, P75, P77, P8, P87
4.75	11	P7, P16, P26, P34, P38, P41, P57, P58, P65, P71, P89
5	9	P1, P14, P15, P20, P28, P42, P70, P76, P88
5.5	1	P3

greater transparency and articulation of research methodology will be an important area for improvement. More empirical experiments should be performed to address the need of real-time solutions that can operate effectively in the big data era. Similarly, some researchers reported their results across a single evaluation metric only (e.g., accuracy), which is not an effective strategy when considering imbalanced datasets. Additionally, very few researchers considered computational effort as an evaluation metric for their systems, although this is considered a very critical issue, especially with real-time IDSs. Finally, the implication of our findings is that there is a general lack studies that explore algorithms that can be effective classifying traffic in contemporary networks, when using antiquated datasets. Simply put, research with new dataset is strong needed. In light of such a scenario, it is significant to generate modern intrusion detection datasets to resolve the current datasets issues, which, in turn, allow researchers to support the findings and validity of their new approaches.

APPENDIX

See Tables 5–8.

TABLE 8. Data mining techniques iteration per year.

Year	DM Technique	Freq.	Year	DM Technique	Freq.
2007	k-NN, BN	1	2015	OneR, AR	1
	Fuzzy	2		Fuzzy, GA, AdaBoost	3
	DT, ANN	3		K-Means, k-NN	5
	GA, SVM	4		ANN	9
2008	ANN, AdaBoost, GA, Fuzzy	1		BN	14
	DT, SVM	2		SVM	16
2009	Fuzzy, GA, DT, k-NN	1		DT	21
	ANN	2	2016	CR, GA, LP, k-NN	1
	SVM	3		Fuzzy, DT, ANN, K-Means	2
2010	K-Means, GA	1		BN	3
	Fuzzy, k-NN, ANN	2		SVM	4
	DT, BN	3	2017	HC, RT, OneR, k-NN,	1
	SVM	5		GA, Kstar, EM, LA	
2011	HC	1		ANN	2
	SVM	2		Fuzzy	3
2012	ANN, BN, DT, SVM	2		K-Means	4
2013	BN, RR, K-Means, AR	1		DT	5
	ANN	2		SVM	6
	DT	3		BN	7
2014	ANN, AdaBoost, BN	1			
	DT, Fuzzy	2			
	GA	3			
	SVM	4			

REFERENCES

[1] P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, "Practical real-time intrusion detection using machine learning approaches," *Comput. Commun.*, vol. 34, no. 18, pp. 2227–2235, Dec. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014036641100209X

[2] S. Axelsson, "The base-rate fallacy and the difficulty of intrusion detection," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 3, pp. 186–205, Aug. 2000. [Online]. Available: http://doi.acm.org/10.1145/357830.357849

[3] A. G. C. de Sá, A. C. M. Pereira, and G. L. Pappa, "A customized classification algorithm for credit card fraud detection," *Eng. Appl. Artif. Intell.*, vol. 72, pp. 21–29, Jun. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0952197618300605

[4] W. Lee, S. J. Stolfo, and K. W. Mok, "A data mining framework for building intrusion detection models," in *Proc. IEEE Symp. Secur. Privacy*, May 1999, pp. 120–132.

[5] I. Bose and R. K. Mahapatra, "Business data mining—A machine learning perspective," *Inf. Manage.*, vol. 39, no. 3, pp. 211–225, Dec. 2001.

[6] M. Khalilian, N. Mustapha, M. N. Sulaiman, and A. Mamat, "Intrusion detection system with data mining approach: A review," *Global J. Comput. Sci. Technol.*, vol. 11, no. 5, pp. 28–34, Apr. 2011. [Online]. Available: https://computerresearch.org/index.php/computer/article/view/714

[7] D. K. Denatious and A. John, "Survey on data mining techniques to enhance intrusion detection," in *Proc. Int. Conf. Comput. Commun. Informat.*, Jan. 2012, pp. 1–5.

[8] M. Injadat, F. Salo, and A. B. Nassif, "Data mining techniques in social media: A survey," *Neurocomputing*, vol. 214, pp. 654–670, Nov. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092523121630683X

[9] R. Latif, H. Abbas, and S. Assar, "Distributed denial of service (DDoS) attack in cloud-assisted wireless body area networks: A systematic literature review," *J. Med. Syst.*, vol. 38, no. 11, p. 128, Sep. 2014, doi: 10.1007/s10916-014-0128-8.

[10] A. S. Subaira and P. Anitha, "Efficient classification mechanism for network intrusion detection system based on data mining techniques: A survey," in *Proc. IEEE 8th Int. Conf. Intell. Syst. Control (ISCO)*, Jan. 2014, pp. 274–280.

[11] D. J. Weller-Fahy, B. J. Borghetti, and A. A. Sodemann, "A survey of distance and similarity measures used within network intrusion anomaly detection," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 70–91, 1st Quart., 2015.

[12] K. Barbara and C. Stuart, *Guideline for Performing Systematic Literature Reviews in Software Engineering (Version 2.3.)*. Staffordshire, U.K.: Univ. of Keele, 2007.

[13] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Jun. 2008, doi: 10.1007/s10115-007-0114-2.

[14] G. W. Noblit and R. D. Hare, *Meta-Ethnography: Synthesizing Qualitative Studies*, vol. 11. Newbury Park, CA, USA: Sage, 1988.

[15] N. Ben Amor, S. Benferhat, and Z. Elouedi, "Naive Bayes vs decision trees in intrusion detection systems," in *Proc. ACM Symp. Appl. Comput. (SAC)*. New York, NY, USA: ACM, 2004, pp. 420–424. [Online]. Available: http://doi.acm.org/10.1145/967900.967989

[16] P. Aggarwal and S. K. Sharma, "A new metric for proficient performance evaluation of intrusion detection system," in *Proc. Int. Joint Conf.*, Á. Herrero, B. Baruque, J. Sedano, H. Quintián, and E. Corchado, Eds. Cham, Switzerland: Springer, 2015, pp. 321–331.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[18] C. G. Weng and J. Poon, "A new evaluation measure for imbalanced datasets," in *Proc. 7th Austral. Data Mining Conf. (AusDM)*, vol. 87. Darlinghurst, Australia: Australian Computer Society, Inc., 2008, pp. 27–32. [Online]. Available: http://dl.acm.org/citation.cfm?id=2449288.2449295

[19] Q. Gu, Z. Cai, L. Zhu, and B. Huang, "Data mining on imbalanced data sets," in *Proc. Int. Conf. Adv. Comput. Theory Eng.*, Dec. 2008, pp. 1020–1024.

[20] F. Idrees, M. Rajarajan, M. Conti, T. M. Chen, and Y. Rahulamathavan, "PIndroid: A novel Android malware detection system using ensemble learning methods," *Comput. Secur.*, vol. 68, pp. 36–46, Jul. 2017.

[21] S. T. Brugger and J. Chow, "An assessment of the DARPA IDS evaluation dataset using snort," Dept. Comput. Sci., UCDAVIS, Tech. Rep. CSE-2007-1, 2007, vol. 1, p. 22.

[22] G. Maciá-Fernández, J. Camacho, R. Magán-Carrión, P. García-Teodoro, and R. Therón, "UGR'16: A new dataset for the evaluation of cyclostationarity-based network IDSs," *Comput. Secur.*, vol. 73, pp. 411–424, Mar. 2018.

[23] R. Chitrakar and C. Huang, "Selection of candidate support vectors in incremental SVM for network intrusion detection," *Comput. Secur.*, vol. 45, pp. 231–241, Sep. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404814000996

[24] V. Bolón-Canedo, N. Sanchez-Maroño, and A. Alonso-Betanzos, "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5947–5957, May 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417410012650

[25] J. M. Fossaceca, T. A. Mazzuchi, and S. Sarkani, "MARK-ELM: Application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection," *Expert Syst. Appl.*, vol. 42, no. 8, pp. 4062–4080, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414008197

[26] K. Shafi and H. A. Abbass, "Evaluation of an adaptive genetic-based signature extraction system for network intrusion detection," *Pattern Anal. Appl.*, vol. 16, no. 4, pp. 549–566, Nov. 2013, doi: 10.1007/s10044-011-0255-5.

[27] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Inf. Sci.*, vol. 177, no. 18, pp. 3799–3821, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025507001648

[28] B. Senthilnayaki, K. Venkatalakshmi, and A. Kannan, "Intrusion detection using optimal genetic feature selection and SVM based classifier," in *Proc. 3rd Int. Conf. Signal Process., Commun. Netw. (ICSCN)*, Mar. 2015, pp. 1–4.

[29] Z. Yu, J. J. P. Tsai, and T. Weigert, "An automatically tuning intrusion detection system," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 373–384, Apr. 2007.

[30] J. Esmaily, R. Moradinezhad, and J. Ghasemi, "Intrusion detection system based on multi-layer perceptron neural networks and decision tree," in *Proc. 7th Conf. Inf. Knowl. Technol. (IKT)*, May 2015, pp. 1–5.

[31] C.-H. Tsang, S. Kwong, and H. Wang, "Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection," *Pattern Recognit.*, vol. 40, no. 9, pp. 2373–2391, Sep. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320306005218

[32] Z. Banković, D. Stepanović, S. Bojanić, and O. Nieto-Taladriz, "Improving network security using genetic algorithm approach," *Comput. Elect. Eng.*, vol. 33, nos. 5–6, pp. 438–451, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0045790607000584

[33] M. S. Rani and S. B. Xavier, "A hybrid intrusion detection system based on C5.0 decision tree and one-class SVM," *Int. J. Current Eng. Technol.*, vol. 5, no. 3, pp. 2001–2007, 2015.

[34] A. Abraham, R. Jain, J. Thomas, and S. Y. Han, "D-SCIDS: Distributed soft computing intrusion detection system," *J. Netw. Comput. Appl.*, vol. 30, no. 1, pp. 81–98, Jan. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804505000421

[35] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowl.-Based Syst.*, vol. 78, pp. 13–21, Apr. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705115000167

[36] S. Peddabachigari, A. Abraham, C. Grosan, and J. Thomas, "Modeling intrusion detection system using hybrid intelligent systems," *J. Netw. Comput. Appl.*, vol. 30, no. 1, pp. 114–132, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804505000445

[37] Y. Danane and T. Parvat, "Intrusion detection system using fuzzy genetic algorithm," in *Proc. Int. Conf. Pervas. Comput. (ICPC)*, Jan. 2015, pp. 1–5.

[38] 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.

[39] S. A. Jebur and H. H. O. Nasereddin, "Enhanced solutions for misuse network intrusion detection system using SGA and SSGA," *Int. J. Comput. Sci. Netw. Secur.*, vol. 15, no. 5, p. 12, 2015.

[40] L. P. Rajeswari and K. Arputharaj, "An active rule approach for network intrusion detection with enhanced C4.5 algorithm," *Int. J. Commun., Netw. Syst. Sci.*, vol. 1, no. 4, pp. 314–321, 2008.

[41] A. Tesfahun and D. L. Bhaskari, "Effective hybrid intrusion detection system: A layered approach," *Int. J. Comput. Netw. Inf. Secur.*, vol. 7, no. 3, pp. 35–41, 2015.

[42] Y.-X. Wei and M.-Q. Wu, "KFDA and clustering based multiclass SVM for intrusion detection," *J. China Univ. Posts Telecommun.*, vol. 15, no. 1, pp. 123–128, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1005888508600746

[43] M. Abdulrazaq and A. Salih, "Combination of multi classification algorithms for intrusion detection system," *Int. J. Sci. Eng. Res.*, vol. 6, no. 1, pp. 1364–1371, 2015.

[44] K. Shafi and H. A. Abbass, "An adaptive genetic-based signature learning system for intrusion detection," *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12036–12043, Dec. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417409002589

[45] R. Singh, H. Kumar, and R. K. Singla, "An intrusion detection system using network traffic profiling and online sequential extreme learning machine," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8609–8624, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417415004753

[46] S.-Y. Wu and E. Yen, "Data mining-based intrusion detectors," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5605–5612, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417408004089

[47] M. P. Arthur and K. Kannan, "Cross-layer based multiclass intrusion detection system for secure multicast communication of MANET in military networks," *Wireless Netw.*, vol. 22, no. 3, pp. 1035–1059, Apr. 2016, doi: 10.1007/s11276-015-1065-2.

[48] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, "Intrusion detection using fuzzy association rules," *Appl. Soft Comput.*, vol. 9, no. 2, pp. 462–469, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494608000975

[49] S. Elhag, A. Fernández, A. Bawakid, S. Alshomrani, and F. Herrera, "On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 193–202, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414004783

[50] H. Tang and Z. Cao, "Machine learning-based intrusion detection algorithms," *J. Comput. Inf. Syst.*, vol. 5, no. 6, pp. 1825–1831, 2009.

[51] C. Anita S and S. Gupta, "An effective model for anomaly IDS to improve the efficiency," in *Proc. Int. Conf. Green Comput. Internet Things (ICGCIoT)*, Oct. 2015, pp. 190–194.

[52] G. Wang, J. Hao, J. Ma, and L. Huang, "A new approach to intrusion detection using artificial neural networks and fuzzy clustering," *Expert Syst. Appl.*, vol. 37, no. 9, pp. 6225–6232, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417410001417

[53] H.-H. Chou and S.-D. Wang, "An adaptive network intrusion detection approach for the cloud environment," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, Sep. 2015, pp. 1–6.

[54] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320309002155

[55] K. S. Desale, C. N. Kumathekar, and A. P. Chavan, "Efficient intrusion detection system using stream data mining classification technique," in *Proc. Int. Conf. Comput. Commun. Control Autom.*, Feb. 2015, pp. 469–473.

[56] S. Rajasegarar, C. Leckie, J. C. Bezdek, and M. Palaniswami, "Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 518–533, Sep. 2010.

[57] P. Sornsuwit and S. Jaiyen, "Intrusion detection model based on ensemble learning for U2R and R2L attacks," in *Proc. 7th Int. Conf. Inf. Technol. Elect. Eng. (ICITEE)*, Oct. 2015, pp. 354–359.

[58] D. M. Farid and M. Z. Rahman, "Anomaly network intrusion detection based on improved self adaptive Bayesian algorithm," *J. Comput.*, vol. 5, no. 1, pp. 23–31, 2010.

[59] B. Vrat, N. Aggarwal, and S. Venkatesan, "Anomaly detection in IPv4 and IPv6 networks using machine learning," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2015, pp. 1–6.

[60] M. S. Abadeh and J. Habibi, "A hybridization of evolutionary fuzzy systems and ant colony optimization for intrusion detection," *ISC Int. J. Inf. Secur.*, vol. 2, no. 1, pp. 33–46, 2010.

[61] F. A. A. Alseiari and Z. Aung, "Real-time anomaly-based distributed intrusion detection systems for advanced Metering Infrastructure utilizing stream data mining," in *Proc. Int. Conf. Smart Grid Clean Energy Technol. (ICSGCE)*, Oct. 2015, pp. 148–153.

[62] M. Ektefa, S. Memar, F. Sidi, and L. S. Affendey, "Intrusion detection using data mining techniques," in *Proc. Int. Conf. Inf. Retr. Knowl. Manage. (CAMP)*, Mar. 2010, pp. 200–203.

[63] N. F. Haq, A. R. Onik, and F. M. Shah, "An ensemble framework of anomaly detection using hybridized feature selection approach (HFSA)," in *Proc. SAI Intell. Syst. Conf. (IntelliSys)*, Nov. 2015, pp. 989–995.

[64] K. Jaswal, P. Kumar, and S. Rawat, "Design and development of a prototype application for intrusion detection using data mining," in *Proc. 4th Int. Conf. Rel., Infocom Technol. Optim. (ICRITO)*, Sep. 2015, pp. 1–6.

[65] Y. Yi, J. Wu, and W. Xu, "Incremental SVM based on reserved set for network intrusion detection," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7698–7707, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417410015046

[66] P. Singh and A. Tiwari, "An efficient approach for intrusion detection in reduced features of KDD99 using ID3 and classification with KNNGA," in *Proc. 2nd Int. Conf. Adv. Comput. Commun. Eng.*, May 2015, pp. 445–452.

[67] Y. Li, W. Li, and G. Wu, "An intrusion detection approach using SVM and multiple kernel method," *Int. J. Adv. Comput. Technol.*, vol. 4, no. 1, pp. 463–469, 2012.

[68] P. Amudha, S. Karthik, and S. Sivakumari, "Intrusion detection based on core vector machine and ensemble classification methods," in *Proc. Int. Conf. Soft-Comput. Netw. Secur. (ICSNS)*, Feb. 2015, pp. 1–5.

[69] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 424–430, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417411009948

[70] S. K. Gautam and H. Om, "Anomaly detection system using entropy based technique," in *Proc. 1st Int. Conf. Next Gener. Comput. Technol. (NGCT)*, Sep. 2015, pp. 738–743.

[71] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417411009080

[72] M. V. Kotpalliwar and R. Wajgi, "Classification of attacks using support vector machine (SVM) on KDDCUP'99 IDS database," in *Proc. 5th Int. Conf. Commun. Syst. Netw. Technol.*, Apr. 2015, pp. 987–990.

[73] N. Wattanapongsakorn *et al.*, "A practical network-based intrusion detection and prevention system," in *Proc. IEEE 11th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Jun. 2012, pp. 209–214.

[74] S. Sahu and B. M. Mehtre, "Network intrusion detection system using J48 decision tree," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2015, pp. 2023–2026.

[75] S. Thaseen and C. A. Kumar, "An analysis of supervised tree based classifiers for intrusion detection system," in *Proc. Int. Conf. Pattern Recognit., Informat. Mobile Eng.*, Feb. 2013, pp. 294–299.

[76] T. H. Hadi and M. R. Joshi, "Handling ambiguous packets in intrusion detection," in *Proc. 3rd Int. Conf. Signal Process., Commun. Netw. (ICSCN)*, Mar. 2015, pp. 1–7.

[77] S. M. H. Bamakan, H. Wang, T. Yingjie, and Y. Shi, "An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization," *Neurocomputing*, vol. 199, pp. 90–102, Jul. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231216300510

[78] Y. J. Zhao, M. J. Wei, and J. Wang, "Realization of intrusion detection system based on the improved data mining technology," in *Proc. 8th Int. Conf. Comput. Sci. Educ.*, Apr. 2013, pp. 982–987.

[79] M. M. Rathore, A. Ahmad, and A. Paul, "Real time intrusion detection system for ultra-high-speed big data environments," *J. Supercomput.*, vol. 72, no. 9, pp. 3489–3510, Sep. 2016, doi: 10.1007/s11227-015-1615-5.

[80] E. Wonghirunsombat, T. Asawaniwed, V. Hanchana, N. Wattanapongsakorn, S. Srakaew, and C. Charnsripinyo, "A centralized management framework of network-based intrusion detection and prevention system," in *Proc. 10th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, May 2013, pp. 183–188.

[81] S. Masarat, S. Sharifian, and H. Taheri, "Modified parallel random forest for intrusion detection systems," *J. Supercomput.*, vol. 72, no. 6, pp. 2235–2258, Jun. 2016, doi: 10.1007/s11227-016-1727-6.

[82] B. Senthilnayaki, K. Venkatalakshmi, and A. Kannan, "An intelligent intrusion detection system using genetic based feature selection and modified J48 decision tree classifier," in *Proc. 5th Int. Conf. Adv. Comput. (ICoAC)*, Dec. 2013, pp. 1–7.

[83] S.-H. Kang and K. J. Kim, "A feature selection approach to find optimal feature subsets for the network intrusion detection system," *Cluster Comput.*, vol. 19, no. 1, pp. 325–333, Mar. 2016, doi: 10.1007/s10586-015-0527-8.

[84] M. Kakavand, N. Mustapha, A. Mustapha, and M. T. Abdullah, "A text mining-based anomaly detection model in network security," *Global J. Comput. Sci. Technol.*, vol. 14, no. 5, pp. 22–31, 2014. [Online]. Available: https://computerresearch.org/index.php/computer/article/view/1110

[85] M. A. Jabbar, R. Aluvalu, and S. S. S. Reddy, "Cluster based ensemble classification for intrusion detection system," in *Proc. 9th Int. Conf. Mach. Learn. Comput. (ICMLC)*. New York, NY, USA: ACM, 2017, pp. 253–257. [Online]. Available: http://doi.acm.org/10.1145/3055635.3056595

[86] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1690–1700, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413006878

[87] M. R. G. Raman, K. Kirthivasan, and V. S. S. Sriram, "Development of rough set—Hypergraph technique for key feature identification in intrusion detection systems," *Comput. Electr. Eng.*, vol. 59, pp. 189–200, Apr. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0045790617300460

[88] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Appl. Soft Comput.*, vol. 18, pp. 178–184, May 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494614000477

[89] R. A. R. Ashfaq, X.-Z. Wang, J. Z. Huang, H. Abbas, and Y.-L. He, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Inf. Sci.*, vol. 378, pp. 484–497, Feb. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025516302547

[90] B. Luo and J. Xia, "A novel intrusion detection system based on feature generation with visualization strategy," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4139–4147, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414000074

[91] H. Bostani and M. Sheikhan, "Modification of supervised OPF-based intrusion detection systems using unsupervised learning and social network concept," *Pattern Recognit.*, vol. 62, pp. 56–72, Feb. 2017.

[92] A. Chaudhary, V. N. Tiwari, and A. Kumar, "Design an anomaly based fuzzy intrusion detection system for packet dropping attack in mobile ad hoc networks," in *Proc. IEEE Int. Adv. Comput. Conf. (IACC)*, Feb. 2014, pp. 256–261.

[93] W. Chen, F. Kong, F. Mei, G. Yuan, and B. Li, "A novel unsupervised anomaly detection approach for intrusion detection system," in *Proc. IEEE 3rd Int. Conf. Big Data Secur. Cloud (Bigdatasecurity), IEEE Int. Conf. High Perform. Smart Comput. (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS)*, May 2017, pp. 69–73.

[94] O. Al-Jarrah and A. Arafat, "Network intrusion detection system using neural network classification of attack behavior," *J. Adv. Inf. Technol.*, vol. 6, no. 1, pp. 291–295, 2015.

[95] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in *Proc. 3th Int. Conf. Web Res. (ICWR)*, Apr. 2017, pp. 178–184.

[96] S. E. Benaicha, L. Saoudi, S. E. B. Guermeche, and O. Lounis, "Intrusion detection system using genetic algorithm," in *Proc. Sci. Inf. Conf.*, Aug. 2014, pp. 564–568.

[97] L. Yang, J. Li, G. Fehringer, P. Barraclough, G. Sexton, and Y. Cao, "Intrusion detection system by fuzzy interpolation," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2017, pp. 1–6.

[98] S. Kumar and A. Yadav, "Increasing performance Of intrusion detection system using neural network," in *Proc. IEEE Int. Conf. Adv. Commun., Control Comput. Technol.*, May 2014, pp. 546–550.

[99] E. A. Shams and A. Rizaner, "A novel support vector machine based intrusion detection system for mobile ad hoc networks," *Wireless Netw.*, vol. 24, no. 5, pp. 1821–1829, Jan. 2017, doi: 10.1007/s11276-016-1439-0.

[100] D. Pal and A. Parashar, "Improved genetic algorithm for intrusion detection system," in *Proc. Int. Conf. Comput. Intell. Commun. Netw.*, Nov. 2014, pp. 835–839.

[101] S. Jamali and P. Jafarzadeh, "An intelligent intrusion detection system by using hierarchically structured learning automata," *Neural Comput. Appl.*, vol. 28, no. 5, pp. 1001–1008, May 2017, doi: 10.1007/s00521-015-2116-4.

[102] S. K. Wagh and S. R. Kolhe, "Effective intrusion detection system using semi-supervised learning," in *Proc. Int. Conf. Data Mining Intell. Comput. (ICDMIC)*, Sep. 2014, pp. 1–5.

[103] N. Acharya and S. Singh, "An IWD-based feature selection method for intrusion detection system," *Soft Comput.*, vol. 22, no. 13, pp. 4407–4416, May 2017, doi: 10.1007/s00500-017-2635-2.

[104] S. Ramakrishnan and S. Devaraju, "Attack's feature selection-based network intrusion detection system using fuzzy control language," *Int. J. Fuzzy Syst.*, vol. 19, no. 2, pp. 316–328, Apr. 2017, doi: 10.1007/s40815-016-0160-6.

[105] S. Potluri and C. Diedrich, "Accelerated deep neural networks for enhanced intrusion detection system," in *Proc. IEEE 21st Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2016, pp. 1–8.

[106] S. R. Kumari and P. Kumari, "Adaptive anomaly intrusion detection system using optimized hoeffding tree," *ARPN J. Eng. Appl. Sci.*, vol. 33, no. 1, pp. 102–108, 2014.

[107] B. M. Aslahi-Shahri *et al.*, "A hybrid method consisting of GA and SVM for intrusion detection system," *Neural Comput. Appl.*, vol. 27, no. 6, pp. 1669–1676, 2016, doi: 10.1007/s00521-015-1964-2.

[108] J. Hussain, S. Lalmuanawma, and L. Chhakchhuak, "A novel network intrusion detection system using two-stage hybrid classification technique," *IJCCER*, vol. 3, no. 2, pp. 16–27, 2015.

[109] A. Sultana and M. A. Jabbar, "Intelligent network intrusion detection system using data mining techniques," in *Proc. 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. (iCATccT)*, Jul. 2016, pp. 329–333.

[110] A. N. Devi and K. P. M. Kumar, "Intrusion detection system based on genetic–SVM for DoS attacks," *Int. J. Eng. Res. Gen. Sci.*, vol. 3, no. 2, pp. 107–113, Mar./Apr. 2015.

[111] A. Hadri, K. Chougdali, and R. Touahni, "Intrusion detection system using PCA and Fuzzy PCA techniques," in *Proc. Int. Conf. Adv. Commun. Syst. Inf. Secur. (ACOSIS)*, Oct. 2016, pp. 1–7.

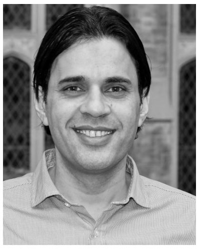
[112] N. G. Relan and D. R. Patil, "Implementation of network intrusion detection system using variant of decision tree algorithm," in *Proc. Int. Conf. Nascent Technol. Eng. Field (ICNTE)*, Jan. 2015, pp. 1–5.

[113] M. A. Manzoor and Y. Morgan, "Real-time support vector machine based network intrusion detection system using Apache storm," in *Proc. IEEE 7th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2016, pp. 1–5.

[114] D. P. Gaikwad and R. C. Thool, "Intrusion detection system using bagging ensemble method of machine learning," in *Proc. Int. Conf. Comput. Commun. Control Autom.*, Feb. 2015, pp. 291–295.

[115] H. M. Tahir, A. M. Said, N. H. Osman, N. H. Zakaria, P. N. M. Sabri, and N. Katuk, "Oving K-means clustering using discretization technique in network intrusion detection system," in *Proc. 3rd Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Aug. 2016, pp. 248–252.

[116] A. Kannan, G. Q. Maguire, Jr., A. Sharma, and P. Schoo, "Genetic algorithm based feature selection algorithm for effective intrusion detection in cloud networks," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 416–423.

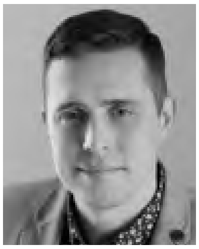


FADI SALO received the B.Sc. degree in computer science from Al-Ahliyya Amman University, Jordan, in 1999, the M.Sc. degree in computer science from University Putra Malaysia, Malaysia, in 2005, and the M.E. degree in electrical and computer engineering from Western University, Canada, in 2015, where he is currently pursuing the Ph.D. degree in software engineering with the Department of Electrical and Computer Engineering. His research interests include data mining, machine learning, social network analysis, data analytics, cloud computing, network security, and intrusion detection.

MOHAMMADNOOR INJADAT received the B.Sc. degree in computer science from Al al-Bayt University, Jordan, in 2000, the M.Sc. degree in computer science from University Putra Malaysia, Malaysia, in 2002, and the M.E. degree in electrical and computer engineering from Western University, Canada, in 2015, where he is currently pursuing the Ph.D. degree in software engineering with the Department of Electrical and Computer Engineering. His research interests include data mining, machine learning, social network analysis, data analytics, and cloud computing.

ALI BOU NASSIF received the master's degree in computer science and the Ph.D. degree in electrical and computer engineering from Western University, Canada, in 2009 and 2012, respectively. He is currently an Assistant Professor and an Assistant Dean of the Graduate Studies, University of Sharjah, United Arab Emirates, and an Adjunct Research Professor with Western University. He His research interests include the applications of statistical and artificial intelligence models in different areas such as software engineering, electrical engineering, e-learning, security, and social media. He is a member of the IEEE Computer Society and a Registered Professional Engineer (P.Eng.) in Ontario.

ABDALLAH SHAMI received the B.E. degree in electrical and computer engineering from Lebanese University in 1997 and the Ph.D. degree in electrical engineering from the Graduate School and University Center, City University of New York, in 2002. In 2002, he joined the Department of Electrical Engineering, Lakehead University, Thunder Bay, ON, Canada, as an Assistant Professor. Since 2004, he has been with Western University, where he is currently a Professor with the Department of Electrical and Computer Engineering. His current research interests are in the areas of network optimization, cloud computing, and wireless networks.



ALEKSANDER ESSEX (M'18) received the Ph.D. degree in computer science from the University of Waterloo in 2012. He is currently an Assistant Professor of software engineering with the Department of ECE, Western University, Canada. Specializing in cybersecurity, his research focuses on cyber threats to electronic and online voting, and on secure multi-party cryptographic techniques for private health informatics. Part of his research focuses on the applications of cryptography to the sharing of health information, such as private records linkage and genomic privacy. His work in applied cryptography includes the recent discovery of an RSA-based public-key encryption scheme for homomorphically computing one-sided threshold functions. He is a member of the ACM, EVN, and is a Licensed Professional Engineer in Ontario.