INTRUSION DETECTION SYSTEM FOR CYBERSECURITY

[1]SYED FURQAN JAMAL, [2]SURAJ SINGH LALOTRA, [3]ANKITA SHARMA

[1,2]Chandigarh University Punjab, India
[3]Assistant Professor Chandigarh University Punjab, India
E-mail: [1]imfurqanjamal@gmail.com, [2]suraj.lalotra@gmail.com, [3]Ankita.e11389@cumail.com

Abstract - As the chance of cyber-attacks continues to raise, the Ask for flooding and adaptable Impedances Divulgence Frameworks (IDS) to bolster cybersecurity measures is squeezing. In this paper, we propose a comprehensive approach to organizing an IDS that saddles the control of machine learning strategies, emphasizing the integration of Python, machine learning calculations, and coordinate frameworks. Our imaginative framework cements cutting-edge machine learning calculations to analyze organize movement plans and recognize unordinary behaviors which is able appear up up potential unsettling impacts. We utilize the adaptability and capability of the Python programming lingo for data control, preprocessing, and unfaltering integration with machine learning libraries. By synergizing these prompts, we have made a boundless and adaptable impedances divulgence component that can ceaselessly learn and progression to keep pace with the ever-evolving hazard scene. Unmistakable datasets are utilized to organize and test the system, and the execution estimations are carefully overviewed to degree the ampleness of the system in recognizing specific cyber dangers. The Consider through and through burrows into how coordination organize structure capabilities can update the exactness and ampleness of the IDS. These cements joining live data supports, grouping from the standard relationship, and organize topology examination, making a comprehensive approach to recognizing and settling potential security breaches. This framework not since it was shows up the control of machine learning in cybersecurity, but in extension highlights the centrality of a comprehensive approach, utilizing Python and organize systems integration in organize to fittingly ensure against progressing cyber perils. This asks around joins to the progression of cybersecurity sharpens, laying the establishment for future headways and headways internal parts the field of impedances divulgence.

Keyword - RFE, Cyberattack, Interruption Discovery Framework, Decision Tree Algorithm, SFEM

I. INTRODUCTION

As cyberattacks gotten to be progressively advanced and organize activity proceeds to blow, up Interruption Discovery Frameworks (IDS) confront a pivotal hurdle: recognizing malevolent behavior within the middle of tremendous sums of genuine information. Whereas this information gives important experiences, it can moreover make wasteful aspects and potential errors.[2] Analyzing unimportant or excess highlights places strain on computational assets and can cloud genuine dangers. In this manner, the cautious determination of related highlights is fundamental in making strides the precision and viability of cutting edge IDS.

The center of this consider is the creation of a Subset Incorporate Conclusion Instrument (SFEM) especially laid out for recognizing interferences. Our suggestion sets out to realize two principal destinations:

Apportion with immaterial highlights: Firstly, through the area and transfer of unessential highlights that have little or no influence on recognizing between ordinary and intrusive behavior, SFEM focuses to unravel the examination procedure and lessen the require for over the best computation.

Recognize significantly discriminative highlights: On the other hand, SFEM is committed to recognizing characteristics that appear a tall level of affiliation with harmful behavior. These highlights, much acknowledged to their capacity to isolated, offer basic understanding for correct interference detection.[4] At SFEM, our approach to intrusion revelation utilizes recursive highlight transfer (RFE), a significantly compelling strategy for selecting highlights. By mixing RFE with a choice tree-based classifier, prepared to intentionally evaluate the influence of each include and choose the driving set for our purposes. Through this key combination, we are going ceaselessly refine and optimize our incorporate assurance for most extraordinary reasonability.

The SFEM, if implemented as planned, would boast a broad range of potential key central facilities.

1. Better precision in detection: These results are obtained by the classifier making more accurate classifications of non-interferometer and normal behaviors.
2. More efficient computational processes: In this regard, SFEM reduces the number of features that therefore minimizes scheduling overheads and enables faster analysis and improved system responsiveness.
3. Lessened storage requirements: The smaller feature set means that the need for storing everything is reduced, thus data management becomes more effective.

Our consider makes a important commitment to the field of interruption discovery by presenting a cutting-edge SFEM that utilizes RFE and choice tree learning.[5] Through experimentation with a well-known interruption location dataset, we clearly

grandstand the capabilities of our approach in highlight determination, location precision improvement, and in general framework execution advancement.

II. RELATED WORK

The interface concept becomes relevant at this stage to serve as a blueprint to observe and remove flaws from the overall product. IDS processes within the creation of Obstacles to be transferred to the internet for application.

Indeed, the spatiality involved in the inner the area of data the manifestation of the dreadfulness that bluntly show us the reality specifically that data hold. its grave control. In this area of research, the study of renowned techniques for cement is considered. certification on Integration of Distributed Systems (IDS), thus, acknowledging their robustness, efficiency, and reliability in a context of the present Subset Front-End Module (SFP).

1. Filter-based strategies: This approach even installs careful considerations in each interface. are worthy of the attention of the target course if they consist of the cautious aspects or the main referrals to the subject matter of the course. (e.g., impedances). The Information theory and Chi-square strategies are certainly the ones that we are well-familiar with. up, and ReliefF.

Strengths:
Valuable and computationally cheap.

Limitations:
This supports creation of highly essential components and facilities as it.: individual evaluation, rejecting interfacial standard. Throughout our long history, art has had a profound impact on human civilization. From the cave paintings of ancient times to the colorful murals and statues adorning modern-day cities, art has played a critical role in our ability to communicate, preserve traditions, and connect with others in the most intimate ways.

2. Wrapper-based techniques: These strategies assess just small subsets from a given set. depends on the fact that they figuratively present a simple method of getting a mental image of the course of the calculation. The iteration of two concentrate groups trying to evacuate the spotlights is known as Recursive Set Trade. such a low-level implementation is simply base-coding.

Strengths:
Think through that which makes that sentence generally applicable and restate the confirmation as is better fit for that category.

Limitations:
A methodology is computationally powerful and little replicable to err on the side of customizing particular models.

3. Embedded-based methods: They enabled weighted underrepresentation of complex models compared to other features of learning process continually using regularization techniques to frame the punishment for complex models. Utilizes L1/L2 descent such as one can explain, built-in decisions as well as feature importance with it.

Strengths:
Additionally, convolutional neural networks can be repeatedly structured to be compact and aid in avoiding overfitting by penalizing complex models.

Limitations:
However, may no away can recognize the constituting and vital elements in isolation.

4. Hybrid methods: These methods consist of two major categories, that is, filter-based approach being like bagging and boosting to improve efficiency is accomplished using wrapper-based technique and the classification.

Strengths:
Can implement both techniques alike to have two birds with one stone, so to say, after all.

Limitations:
The growing difficulty in understanding their words and the complexity of their messages serve as eye-openers, demonstrating certain issues.

Comparison with SFEM: The proposed SFEM exploits the properties of RFE in by using it to iteratively select highlights. In any case, this goes beyond consolidating a classifier based on a selection tree to consider the value of the appearance and select not the highlights with the smallest impact but also the distinguishing points most important to reveal intrusion behavior. This hybrid approach focuses on achieving improvements in accuracy and common sense compared to filter or wrapper-based strategies.

III. PROPOSED SUBSET FEATURE ELIMINATION MECHANISM

The proposed Subset Feature Elimination Mechanism (SFEM) points to viably selecting a subset of highlights that improve the execution of Interruption Discovery Frameworks (IDS). It leverages the control of Recursive Include Disposal (RFE) coupled with a choice tree-based classifier to attain two key targets:

Eliminate irrelevant features: Highlights with negligible or no discriminative control in recognizing

ordinary and meddlesome behavior are recognized and expelled, streamlining the investigation handle, and decreasing computational overhead.[11]

Distinguish exceedingly discriminative highlights: Highlights showing a solid relationship with noxious action are pinpointed and held, giving significant bits of knowledge for precise interruption discovery.

Key Components:

- **Decision Tree Classifier:** Its characteristics include the significance scoring component plays a pivotal part in assessing highlights and directing the RFE handle.
- **RFE:** Empowers iterative highlight ends based on their effect on the classifier's execution.
- **Stopping Criterion:** Guarantees satisfactory include choice without overfitting or relinquishing precision.

Benefits of SFEM:

- **Improved Detection Accuracy**: Centering on pertinent features enhances the classifier's capacity to distinguish between ordinary and meddlesome behavior, driving to more exact locations and less untrue positives.
- **Enhanced Computational Efficiency:** Diminished highlight dimensionality deciphers to quicker investigation and progressed framework system framework Synonym responsiveness.
- **Reduced Storage Demand:** Putting away and overseeing a smaller highlight set interprets to lower capacity needs and encourages productive data handling.
- **Interpretability:** Choice tree-based classifiers offer inalienable interpretability, permitting a simpler understanding of how chosen highlights contribute to interruption location.
- **Compared to existing strategies:** SFEM offers a few points of interest over filter-based strategies by considering highlighting intelligence and fitting determination to the chosen classifier.[1]

It overcomes the impediments of unadulterated wrapper-based strategies by joining extra direction by including significance scores.

Compared to hybrid methods, SFEM offers possibly way better productivity whereas holding interpretability through the decision tree classifier.

IV. METHODOLOGY

The test strategy utilized in this paper is outlined in Fig. 1 and portrayed as taken after:

Data Gathering: Essentially in this step, the dataset needs to go through a cleaning preparation to remove copy records, as the NSL KDD dataset was utilized which has as of now been cleaned, this step is not any longer required. Another Pre-processing operation needs to be taken in put since the dataset contains numerical and non-numerical occurrences. For the most part, the estimator (classifier) defined within the scikit-learn works well with numerical inputs, so a one-of-K or one-hot encoding strategy is utilized to form that change.[16] This method will transform each categorical highlight with m conceivable inputs to n parallel highlights, with one dynamic at a time only.

Feature Scaling: Highlight scaling could be a common prerequisite of machine learning strategies, and to dodge that highlights with expansive values may weigh as well on the ultimate comes about.[9] For each include, calculate the normal, subtract the cruel esteem from the highlight esteem, and partition the result by their standard deviation. After scaling, each highlight will have a zero normal, with a standard deviation of one.

Feature Selection: Highlight selection is utilized to kill the excess and unessential information. It could be a strategy of selecting a subset of important highlights that completely speaks to the given issue near a least weakening of introduction [26], two conceivable reasons were analyzed why it would be prescribed to limit the number of features:

Firstly, it is conceivable that unessential highlights might propose relationships between highlights and target classes that arise just by chance and do not accurately demonstrate the issue. This perspective is additionally related to over-fitting, as a rule in a choice tree classifier.[19] Besides, an expansive number of highlights might enormously increment the computation time without a comparing classifier advancement.

The highlight determination process starts with a univariate highlight choice with ANOVA F-test for include scoring, univariate include choice analyzes each highlight independently to determine the strength of the relationship of the highlight with names. The Select Percentile strategy within the sklearn_feature_selection module was utilized, this strategy selects highlights based on a percentile of the most elevated scores.

Once, the leading subset of highlights was found, a recursive highlight end was applied which over and over built a model, setting the feature aside and after that rehashing the method with the remained highlights until all features within the dataset were depleted. As such, it could be a great optimization for finding the most excellent performing subset of features. The thought is to utilize the weights of a classifier to produce a feature ranking.

Model: Here, a decision tree show was built to parcel the information utilizing data pick-up until occurrences in each leaf hub have uniform course

labels.

Usually, an awfully straightforward but effective hierarchical method for supervised learning (classification or regression) whereby the nearby space (region) is recognized in an arrangement of dreary parts in fewer steps (little). At each test, a single highlight is utilized to part the hub concurring to the feature's values.[2] On the off chance that after the part, for each branch, all the occasions chosen have a place to the comparable course, the part is considered total or immaculate.

A Decision Tree is made up internal decision nodes and terminal leaves. A test work is actualized by each choice hub with a discrete comes about naming the branches. Giving an input, at each hub, a test is built and based on the outcome, one of the branches will be considered. [13]

Here the learning calculation begins at the root and until a leaf node is reached, the prepare will be done recursively at which minute the esteem spoken to within the leaf hub is the yield. Each leaf hub has a results name, which is the course target in case of classification and numeric esteem for regression.
A leaf hub can depict a localized space or locale where occasions finding in this input space (region) possess the same labels for classification and similar numeric esteem for relapse.[18]

Prediction And Evaluation: The test information was used to make expectations of our demonstration and for evaluation, different settings were considered such as the exactness score, accuracy, review, f-measure, and a perplexity framework. A 10-fold cross-validation was performed throughout all the method.

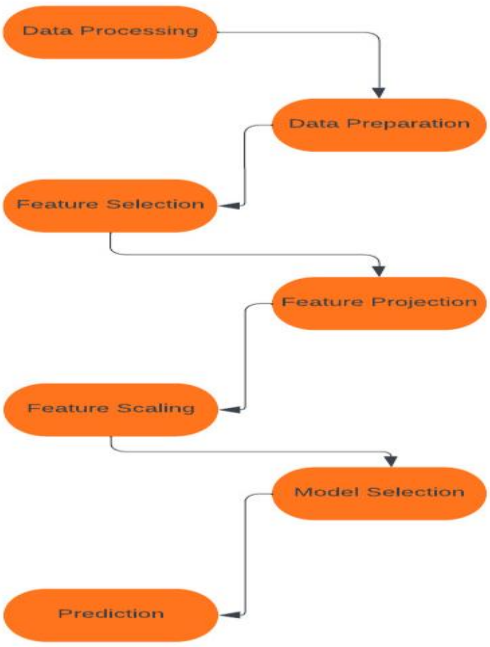

Fig I Flowchart

V. EXPERIMENT AND RESULT

The Decision Tree learning calculation was utilized within the
explore. Decision Tree tends in some cases to overfit, so to discover the finest parameters to fit the demonstration, a thorough grid search parameters tuning was computed and data pick up was utilized to choose highlights.[8] Subsequently, building from the prepared information, a tree was gotten with its clears out being lesson names. When building a decision tree, one highlight is utilized to part the hub and parcel the information. Subsequently, highlights are utilized in a univariate way.

After getting the satisfactory number of highlights amid the univariate choice preparation, a recursive feature elimination (RFE) was worked with the number of highlights passed as a parameter to identify the highlights chosen. Amid the RFE handle, to begin with, the classifier is prepared on the initial set of highlights, and weights are ascribed to each highlight. At that point, highlights whose outright weights are the smallest are pruned from the current set of highlights. That preparation is recursively rehashed on the pruned set until the required number of highlights to select is finally come to.[15]

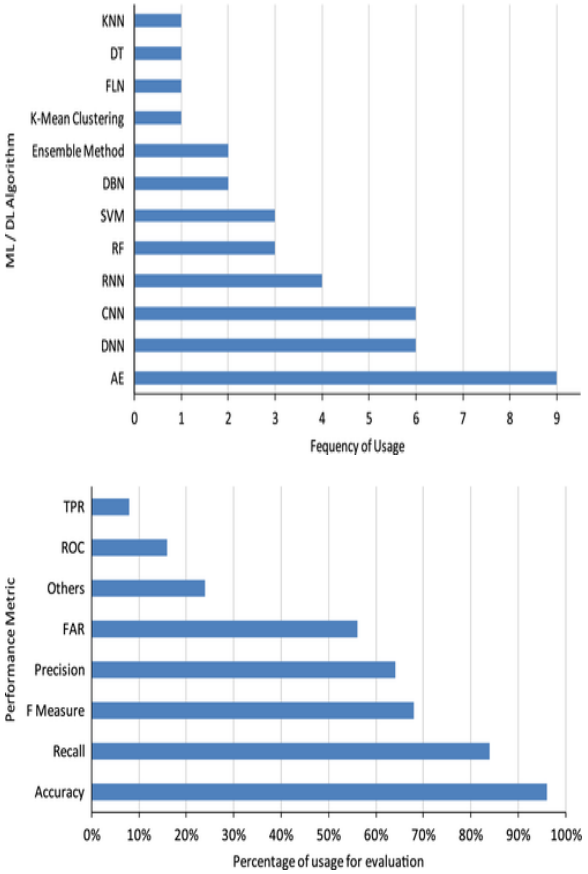

Fig II and III: Illustrating the Accuracy and Precision and Other Factors

Firstly, the classifier is prepared at the

beginning to gather properties, and weights are relegated to each quality. At that point, the absolute weights of a few traits that are the littlest are pruned from the current sets of traits. That method is recursively rehashed on the pruned set until the specified number of properties to choose is in the long run come to. It could be a great optimization for finding the finest-performing subset of highlights.[11] It ought to be noted that RFE has no impact on relationship strategies since the positioning basis is computed with data about a single highlight.

An investigation was performed to determine the accuracy of our estimator after selecting important highlights as outlined in Fig. 3, 4, and 5 and the detail is summarized within Table 4. When comparing the result near the execution assessment with all highlights depict within the Table 5, a noteworthy change of the by and large execution of the proposed show has been observed.

Accuracy	Precision	Recall	F-measure	N of Features	Class
99.90	99.69	99.79	99.74	12	Dos
99.80	99.37	99.37	99.37	15	Probe
99.88	97.40	97.41	97.40	13	R2L
99.95	99.70	99.69	99.70	11	U2R

Accuracy	Precision	Recall	F-measure	N of Features	Class
99.66	99.505	99.71	99.61	41	Dos
99.57	99.04	98.84	98.94	41	Probe
97.03	95.83	95.59	95.71	41	R2L
99.64	99.66	99.61	99.65	41	U2R

Table iv And v. Performance Evaluation With Selected Features

Table 6 shows a 2x2 confusion matrix after features selection on the dataset for a combination of two target classes (normal class and an attack class).

Confusion Matrix		**Predicted Label**			
		Normal	**DoS**		
True Label	**Normal**	9676	25	**Positive predictive value**	99.74 %
	DoS	15	7445	**Negative predictive value**	99.79 %
Confusion Matrix		**Predicted Label**			
		Normal	**Probe**		
True Label	**Normal**	9652	59	**Positive predictive value**	99.39 %
	Probe	30	2391	**Negative predictive value**	98.76 %

Confusion Matrix		Predicted Label			
		Normal	R2L		
True Label	Normal	9594	117	Positive predictive value	98.79 %
	R2L	87	2798	Negative predictive Value	96.98 %
Confusion Matrix		Predicted Label			
		Normal	U2R		
True Label	Normal	9683	28	Positive predictive Value	99.71 %
	U2R	7	60	Negative predictive Value	89.53 %

Table vi. Details Confusion Matrix after Features Selection

99.84	2.368e-03	1.544e-03	1.153e-02	4.119e-04
3.485e-03	99.691	9.38337802e-04	6.702e-04	0
8.674e-03	2.891e-03	99.45	2.891e-03	4.130e-04
4.540e-02	1.039e-03	2.079e-03	96.10	6.239e-03
5.970e-02	0	0	2.089e-01	87.5

Table vii. Confusion Matrix Details

The completed confusion matrix is outlined within the Figure 7 and Table 7 appear the number of redress and erroneous forecasts made by the classification demonstrate compared to the genuine results within the dataset.[17] The investigation for include determination has been worn out terms of the class that accomplished great levels of entropy or Gini index from others within the preparing set and the examination of include significance within the preparing set. The significance of a highlight is computed as the (normalized) total lessening of the basis brought by that include. Figures 8, 9, 10, 11, and Table 7 appears that the foremost pertinent highlights for Dos, Tests, R2L, and U2R are respectively "same_srv_rate", "src_bytes", "dst_host_srv_count" and "root_shell"[10].

With the enhancement the exactness, the proposed demonstration illustrated that it performs well after selecting significant highlights. From Figure 12, it is obvious that the time taken to construct a classifier is diminished through highlight determination, particularly the proposed approach. Building a Decision Tree classifier on the dataset with highlights chosen by our approach takes as it were 0.956 seconds for the DoS attack class, which is quicker than building on the dataset with all the 41 highlights by 14.541s as appeared in Table 9. This result gives a modern understanding employing a classification learning algorithm and reduction technique to choose important and critical include in arrange to make strides in the precision discovery rate of the framework and to recognize conceivable highlights that may contribute to this advancement.[11] As our objective was to decide whether a highlights choice prepare will progress the exactness location of a show on distinctive set of attacks lesson found within the dataset utilized for our test, each set of assaults lesson were treated independently as they show distinctive characteristics and are diverse by nature.

This choice was too made in arrange to recognize all pertinent highlights for each distinctive assaults class and to compare the exactness advancement from the initial set of highlights. As significant highlights which are appropriated to classify those diverse assaults course have been found, the result examination has appeared that the execution of the show has truly been progressed. As a case, the R2L attack accuracy detection has been advancement from 97.03% to 99.88% as well as its execution time as appeared on the Table 9. The Table VIII appears a comparison between our strategy and a few past one.

Author	Method used	Classifier used	Accuracy for Attack Classes (N^{br} of selected features)			
			DOS	Probe	R2L	U2R
(Dhanabal & Shantharajah 2015) [27]	Correlation based Feature Selection method	J48	99.1 % (6)	98.9 % (6)	97.9 % (6)	98.7 %(6)

(Senthilnayaki et al. 2015) [28]	Optimal Genetic Algorithm	SVM	99.15 % (10)	99.08 % (10)	96.50 % (10)	97.03 % (10)
(Zhang & Wang 2013) [29]	Sequential search	Naïve Bayes	99.3 % (11)	97.4 % (11)	95.0 % (11)	59.6 % (11)
(Alazab et al. 2012) [30]	Information gain	J48	99.7 % (12)	97.8 % (12)	91.3 % (12)	97.2 % (12)
(Mukherjee & Sharma 2012) [31]	Feature vitality based Method	Naïve Bayes	98.7% (24)	98.8 % (24)	96.1 % (24)	64% (24)
(Parsazad et al. 2012) [32]	Correlation Coefficient	K-nearest neighbor	98.34 % (30)	98.38 % (30)	97.03 % (30)	83.33 % (30)
(Parsazad et al. 2012) [32]	Fast feature Reduction	K-nearest neighbor	98.28 % (10)	98.50 % (10)	97.79 % (20)	82.00 % (10)
(Parsazad et al. 2012) [32]	Least Square Regression Error	K-nearest neighbor	98.34 % (30)	98.98 % (20)	97.62 % (20)	82.61 % (20)

Table VIII. Comparison With Other Features Selection Techniques

VI. CONCLUSION

In this paper, the importance of employing a set of pertinent highlights with a satisfactory classification learning calculation for modeling an IDS has been illustrated.[9] Benchmark datasets are an imperative fix utilized to test the execution of the proposed strategy. The examination of the utilization of the open datasets appears in Figure 4. It is outlined that 60% times NSL-KDD and KDD Cup'99 were utilized for testing and approving purposes.

Both are very ancient datasets but are still exceptionally prevalent among analysts due to the availability of broad comes about within the writing. Cutting-edge organized engineering is very distinctive from the one 20 a long time ago. It is very self-evident that a demonstration prepared and confirmed utilizing the most recent dataset will perform comparatively better than the show prepared and confirmed utilizing an ancient dataset within the genuine world.[6] fier to distinguish imperative highlights have been done. This preparation over and over builds a show placing the feature aside and after that rehashing the method with the remaining highlights until all highlights shown within the dataset are depleted.

The highlight determination strategy proposed in this paper had accomplished a tall result in terms of exactness and highlights were distinguished based on data pick-up and positioning method.

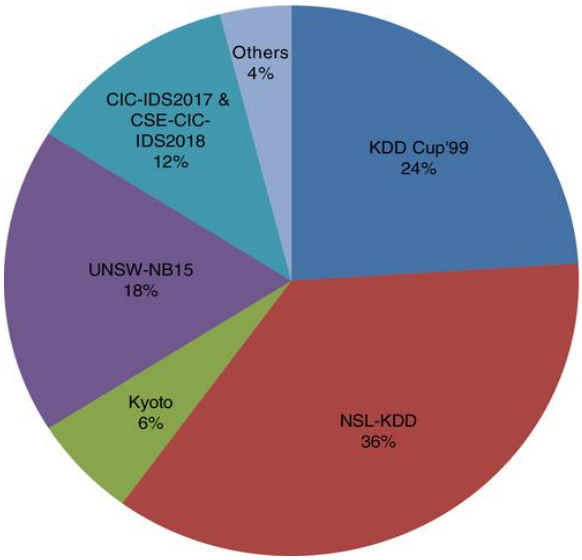

Fig IV Evaluation Metrics for intrusion Detection Systems

REFERENCES

[1] M. P. K. Shelke, M. S. Sontakke, and A. D. Gawande, “Intrusion Detection System for Cloud Computing,” Int. J. Sci. Technol. Res., vol. 1, no. 4, pp. 67–71, 2012.

[2] L. Han, "Using a Dynamic K-means Algorithm to Detect Anomaly Activities," 2011, pp. 1049-1052.

[3] Levin, "KDD-99 Classifier Learning Contest: LLSoft s Results Overview," SIGKDD explorations, vol. 1, pp. 67-75, 2000.

[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] J. McHugh, “Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory,” ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262–294, 2000.

[7] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” Submitted to Second IEEE Symposis

[8] An introduction and recommendation of a feature selection stra-

[9] tegy that comprises a univariate highlights determination related

[10] NSL KDD dataset, Accessed December 2015, https://github.com/defcom17/NSL_KDD

[11] P. Ghosh, C. Debnath, and D. Metia, “An Efficient Hybrid Multilevel Intrusion Detection System in Cloud Environment,” IOSR J. Comput. Eng., vol. 16, no. 4, pp. 16–26, 2014.

[12] William D. How AI can help improve intrusion detection systems [Internet]. GCN. Available from: https://gcn.com/cybersecurity/2020/04/how-ai-can-help-improve intrusion-detection-systems/291266/

[13] Dhanabal, L., Dr. S.P. Shantharajah, "A Study on NSL_KDD Daaset for Intrusion Detection System Based on Classification Algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, issue 6, pp. 446-452, June 2015…

[14] C. F. Tsai, et al., "Intrusion detection by machine learning: A review," Expert Systems with Applications, vol. 36, pp. 11994-12000, 2009.

[15] C. Cowan et al., "Stackguard: automatic adaptive detection and prevention of buffer-overflow attacks," in USENIX security symposium, 1998, vol. 98, pp. 63–78: San Antonio, TX

[16] Creech G, Hu J (2014a) A semantic approach to host-based intrusion detection systems using Contiguousand Discontiguous system call patterns. IEEE Trans Comput 63(4):807–819

[17] Agrawal S, Agrawal J (2015) Survey on anomaly detection using data mining techniques. Procedia Computer Science 60:708–713

[18] Valdovinos I., Perez-Diaz J., Choo K.K., Botero J. Emerging DDoS attack detection and mitigation strategies in software-defined networks: Taxonomy, challenges and future directions. Journal of Network and Computer Applications [Internet]. 2021 Aug 1 [cited 2021 Sep 23];187:103093. Available from: https://www.sciencedirect.com/science/article/pii/S1084804521001156

[19] Niksefat S., Kaghazgaran P., Sadeghiyan B. Privacy issues in intrusion detection systems: A taxonomy, survey and future directions. Computer Science Review. 2017 Aug;25:69–78.

[20] Cybersecurity Spotlight – Signature-Based vs Anomaly-Based Detection [Internet]. CIS.Available from: https://www.cisecurity.org/insights/spotlight/cybersecurity spotlight-signature-based-vs-anomaly-based-detection

[21] Australian. (2017, November). Australian cyber security center threat report2017. Available: https://www.acsc.gov.au/publications/ACSC_Threat_Report_2017.pdf

[22] Bhuyan MH, Bhattacharyya DK, Kalita JK (2014) Network anomaly detection: methods, systems and tools. IEEE Communications Surveys & Tutorials 16(1):303–336

★★★