Academic International Journal of Engineering Sciences
ISSN: 2984-7745
Aca. Intl. J. E. Sci. 2024; 2(1) 15-21
Journal homepage: www.aipublishers.org/aijes

Document heading.

Intrusion Detection Systems: A Revisit of Performance Evaluation Parameters

[1]Tamara Saad Mohamed, [2]Saad Mohammed Khalifah

[1]kut University College, kut, Iraq

[2]kut University College, Kut, Iraq.
tamara.mohhh@gmail.com

Article Info.
Article history:
Received 20 February 2024
Revised 26 April 2024
Accepted 14 May 2024
Published 20 June 2024
Keywords:
Computer systems security, Intrusion detection system, Network Security, Performance evaluation metrics, Cyber-Physical Systems.
How to cite:
Tamara Saad Mohamed, Saad Mohammed Khalifah Intrusion Detection Systems: A Revisit of Performance Evaluation Parameters. Aca. Intl. J. E. Sci. 2024; 2(1) 15-21
DOI:
https://doi.org/10.59675/E212

Abstract

The effectiveness efficiency of various intrusion detection systems (IDSs) have always been evaluated on various datasets using different techniques. This evaluation normally covers a range of metrics, ranging from those that evaluated their accuracy to those that focus on their usability and performance. The evaluation of several attributes of IDS depends on the use of several metrics which have been proposed and used in various studies. Regrettably, there are no existing benchmark metric for the network intrusion detection systems as studies are still ongoing in this regard. Quantitative evaluation of the performance of IDS is mainly based on the use of several performance but in this article, most of the existing performance metrics were reviewed and used for the evaluation of the performance of various IDSs using different benchmark datasets. Furthermore, this study will aid in better understanding of the various performance metrics used for the evaluation of the performance of novel IDSs.; it will also further the insights to those that have found interest in the uses and the development of IDSs and related areas.

Introduction

Cyber-Physical Systems (CPSs) refers to the group of highly interconnected systems that are put together to perform various functions in different fields, such as defense, energy, manufacturing, and healthcare. The role of CPSs in these fields has been revolutionized by the efflorescence of different techniques, such as the artificial intelligence (AI), Internet of Things (IoT)and “5G”, thereby leading to the emergence of more paradigms like smart grids, smart companies, and smart cities. The

emergence of these smart technologies has contributed to the amendment of the general goodness of life in a cost-effective and efficient manner[1].

Despite being an emerging field of technology, most of the early applications of smart technologies have proven their potential application benefits; for instance, the emergence of smart grids has enabled better management and maintenance of public electricity grid through the incorporation of smart elements and devices throughout the power supply chain. In this development, the role of "CPS" and the evolution of smart cities relies on IoT owing to its importance in the realization of the next generation of infrastructures that will offer sustainable and affordable technological solutions. Hence, the capability of the smart elements of smart technologies to establish internet-based communication is important in enabling IoT to achieve connected infrastructures[2].

However, the IoT used in smart infrastructures has being the target of numerous attacks recently due to its open network structure [3]. Many studies have recently focused on the identification of these malicious attacks, especially those targeted against IDSs [4]. An IDS is one of the primary components of a representative security architecture, which supplies visibility into the activities of a system, enabling timely detection and response to any undesired cases. An IDS can be classified into the following main approaches: anomaly-based detection, signature-based detection, hybrid intrusion detection, and specification-based intrusion detection[5]. Hence, the robustness and performance of new IDSs are being evaluated on different datasets using different techniques [6]. The evaluation of IDS can be done from two basic perspectives which are[7]:

- ***Efficiency***: Here, the evaluation is focused on the resources that must be allocated to the system, such as the main memory and CPU cycles.
- ***Effectiveness***: This evaluation which is also known as the classification accuracy focus on the system's ability to classify network activities into intrusive and non-intrusive classes.

Intrusion detection systems can be evaluated based on different features, such as the performance, usability, and accuracy but in most studies, attention is normally given to the effectiveness and accuracy of novel IDS by measuring only the rate of false alarm and the extent of successfully detected attacks. These studies gave little attention to the efficiency of the proposed systems. New IDSs are mostly evaluated for performance based on the measurement of the benchmark metrics on certain benchmark datasets; then, the observed performance of the analyzed system is compared to that of other existing techniques in the same category. Regrettably, there are no existing benchmark metric for the network intrusion detection systems as studies are still ongoing in this regard[8]. Quantitative evaluation of the performance of IDS in the existing studies is mainly based on the use of several performance measures[9]. For instance, the review by within the period of 2000-2008 noted that there are about 276 articles that have been published on intrusion detection but 42% of these articles reported performance assessment using the False Positive Rate (FPR), Detection Rate (DR), and area under the ROC (AUC) as the performance metrics. Furthermore, the study by[6] observed that reliance on only DR, FPR and AUC cannot help in complete evaluation of the performance of new IDSs certain situations and to address this issue, they study suggested Intrusion Detection Capability(CID) as a single objective metric that measures performance in terms of the Positive Predictive Value (PPV), Base Rate, and Negative Predictive Value (NPV).

Classifications of performance metrics

Researchers have developed numerous performance metrics for IDS evaluation but these metrics have been categorized into three types which are probability, ranking, and threshold metrics[4]. Regarding threshold metrics, they include metrics like classification rate (CR), Cost per example (CPE), F-measure (FM), etc. The closeness of a prediction to a threshold is not so important as the consideration is mainly on if the threshold was reached or exceeded. In most studies, the value of threshold metrics ranges from "0" into "1". For the ranking metrics, they include metrics like FPR, Precision (PR), CID, AUC, and Detection Rate (DR). The ranking metrics have values in the range of "0" to "1". The performance evaluation metrics are based on the ordering of the cases and not on the actual values that were predicted. Once the ordering is preserved, no variance is expected as these metrics measure the level of ordering of the attack instances before the normal instances; hence, they are considered a summary of the performance of a model across all possible thresholds. Among the probability metrics is root mean square error (RMSE) and its value is in the range of "0" to "1". RMSE for each attack class is minimized when the predicted value for that class coincides with the true conditional probability of the considered class being a normal class. But, "CID" is a metric that is based on information theory as it offers a preferable comparison of various IDSs compared to the rest of the performance metric, such as AUC[10]. "CID" has values in the range of "0" and "1", with good performing IDS having higher "CID" values. Being that these evaluation metrics are calculated from a confusion matrix (CM), A CM remains a proper way of expressing the classification results of IDSs[11].

Confusion matrix (CM)

This is a well-known metric used in solving classification problems; it is applicable to multiclass and binary classification problems owing to its ability to achieve true and false classification results. Figure 1 provides a detailed explanation of CM. The possible classes of events are as follows:

- True positive (TP): These are the successfully detected intrusions by the IDS.
- False positive (FP): These are the normal activities that were wrongly detected as attacks by the IDS.
- True Negative (TN): These are normal activities that were successfully detected as normal by the IDS.
- False Negative (FN): These are abnormal activities that were wrongly classified as normal by the system[11].

Actual	Predicted	Predicted
	Attack	Normal
Attack	TP	FN
Normal	FP	TN

Fig. 1. General knowledge of the CM[12]

B. Metrics from confusion matrix

Despite the power of a CM in classification representation, it is not considered a powerful metric for comparison tasks. Hence, various performance metrics have been presented and clarified in terms of the CM variables; these metrics offer some numeric values that can be compared and explained with ease in sections.

Aca. Intl. J. E. Sci. 2024; 2(1) 15-21 DOI: https://doi.org/10.59675/E212

- Classification rate (CR): This is the ratio of instances correctly classified to the overall number of instances.

$$\text{CR} = \frac{Correctly\ classified\ instances}{Total\ number\ of\ instances} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Detection rate (DR): This is the ratio of the number of attacks correctly detected by the system to the overall number of attacks.

$$\text{DR} = \frac{Number\ of\ correctly\ detected\ attacks}{Total\ number\ of\ attacks} = \frac{TP}{TP+FN} \quad (2)$$

False-positive-rate (FPR): This is the ratio of the number of ordinary activities wrongly detected as an attack to the overall number of common instances.

$$\text{FPR} = \frac{Number\ of\ normal\ \ instances\ detected\ as\ attacks}{Total\ number\ of\ normal\ instances} = \frac{FP}{FP+TN} \quad (3)$$

- Precision (PR): This is the ratio of the correctly predicted positive instances.

$$\text{PR} = \frac{TP}{TP+Fp} \quad (4)$$

- Recall: This metric serves as a measure of the missing aspect of the Precision; it is the percentage of the real attacks detected by the system. Classifiers are expected to have a rise recall value as this evaluation metric is equivalent to the "DR" of the system.
- F-measure (FM): The FM of any given threshold represents the harmonic mean of the recall and precision at such threshold.

$$\text{FM} = \frac{2}{\frac{1}{PR}+\frac{1}{Recall}} \quad (5)$$

This metric is only preferred when using only one accuracy metric for the evaluation. Area under ROC curve (ROC): When the sensitivity is plotted against (1 - specificity) for all the possible thresholds, the resulting plot is called the ROC curve. Sensitivity is defined as P (Pred = positive TP) and is approximated by the ratio of the correctly predicted TPs to the overall number of considered cases (also the same as a recall). Specificity is defined as P (Pred = TN = negative) and can be approximated by the ratio of the truly predicted negatives to the overall number of considered cases. Generally, the field under the "ROC" curve is mainly considered a summary statistic. ROC curves are based on the Signal detection theory[13]; hence, they can be used to express the relationship between "DR" and "FPR" of a classifier during the tuning process; it can also be used to compare the level of accuracy of different classifiers. Despite the high effectiveness of this metric, it is associated with certain limitations, such as its dependence on the ratio of normal-to-attack traffic.

Furthermore, "ROC"-based comparison of classifiers works best when using the same dataset as comparison based on different datasets is not acceptable except if the involved datasets have the same ratio of normal-to-attack instances. Another problem of the ROC curves is that sometimes, they may be misleading or insufficient to capture the strengths and weaknesses of the evaluated system[13]. Therefore, there is a need to determine the ability of the ROC curve correctly classifying events into normal and attack groups when evaluating the effectiveness of an IDS; this approach should also be backed with other performance measures, such as the efficiency of resource utilization, robustness to attacks and resilience to stress. The evaluation of the performance of new IDSs is of industrial and research importance, especially in tuning and comparing different IDSs in

Aca. Intl. J. E. Sci. 2024; 2(1) 15-21 DOI: https://doi.org/10.59675/E212

a better way. As earlier stated, there are numerous metrics for the evaluation of different aspects of IDS, but until now, there is no single metric that can sufficiently and objectively capture the capability of the IDSs[13].

The study by[7] noted that most of the reported studies on IDs have mainly used metrics such as TPR, FPR, and ROC for the evaluation of the performance of newly developed IDSs. Alternatively, some studied report the use of FNR which is equal to 1 – TPR, as well as TNR which is equal to 1 - FPR. The determination of a better performing IDS from a list of IDSs is usually difficult when relying on the values of these two metrics alone; this is normally the case when there is a need for a tradeoff, such as in a case where one IDS is reporting TPR and FPR values of 0.8 and 0.1 respectively, while another IDS is reporting TPR and FPR values of 0.9 and 0.2, respectively. In this case, reliance on these two values may not be appropriate in determining the better-performing IDS. In this case, the ROC curve provides the needed tradeoff even though it cannot tell the better IDS in most cases.The study by[6] addressed this problem by proposing a new evaluation metric called Intrusion Detection Capability (CID) which is based on Information theory where three metrics (TPR, base rate, and FPR) were considered collectively. This new metric exhibits many important features as follows:

- Naturally considers all the relevant aspects of detection capability, including FPR, FNR, base rate, positive predictive value (PPV), and negative predictive value (NPV).
- Supports the objective evaluation of the capability of IDSs.
- High sensitivity to the operation parameters of the IDS, such as the base rate, FPR, and FNR.

Capability metrics of intrusion detection system (CID)

To evaluate intrusion detection systems (IDS) must use the intrusion detection capability (CID) which is metric of an information theoretic approach. It measures the mutual information between IDS input and output relative to the entropy of the input, which quantifies the reduction in uncertainty about an intrusion event. While CID offers a unified metric for tuning IDS, it does not account for the associated operational costs or provide a direct measure for practical concerns such as false alarm and detection rates. This mentions the complexity of designing an IDS that is both effective and efficient in real world scenarios to occur[14].

$$CID\,\frac{I(X;Y)}{H(x)} = \frac{H(X) - H\left(\frac{x}{y}\right)}{H(x)} \quad (6)$$

Where $I(X;Y)$ = the mutual information of X & Y.

$H(X)$ = entropy of X.

$H\left(\frac{x}{y}\right)$ = the conditional entropy of X when Y is known.

$$H(X)\sum p(x)\log\big(p(x)\big) = -\beta\log(\beta) - (1-\beta)(\log(1-\beta)) \quad (7)$$

$$CID = -B(1-\beta)\log(PPV) - B(1-\beta)\log(1-NPV) - (1-B)(1-\alpha)\log(NPV) - (1-B)\alpha\log(1-PPV) \quad (8)$$

Aca. Intl. J. E. Sci. 2024; 2(1) 15-21 DOI: https://doi.org/10.59675/E212

Readers are referred to the studies by[14,15] and[16] for more detail of the CID.

Conclusion

In this article, a review of the existing performance metrics used in evaluating the performance of IDSs on different benchmark datasets, their strengths and their limitations, was presented. Different performance metrics were compared and validated for the evaluation of IDS performance. Mostly, the performance of IDSs is validated based on the several performance metrics on different benchmark data sets; then, this is followed by the comparison of the evaluated systems with some of the existing systems in the field in terms of performance. Even though various researchers have reported the use of different performance metrics for quantitative performance evaluation of IDSs, literature evidence suggests that most of these studies rely on the DR, FPR, and AUC for performance evaluation. This approach has been noted to fail in certain cases of determining the best-performing IDS. However, efforts have been made to address this issue by proposing the use of CID as a single objective metric but in this study, the use of FPR, DR and CID along with DR of each attack type is being proposed when comparing the performance of new IDSs with some of the existing ones in the field.

References

1. Abushark YB, Khan AI, Alsolami F, Almalawi A, Alam MM, Agrawal A, et al. Cyber Security Analysis and Evaluation for Intrusion Detection Systems. Comput Mater Contin. 2022;72(1):1765–83.
2. Ozkan-Okay M, Samet R, Aslan O, Gupta D. A Comprehensive Systematic Literature Review on Intrusion Detection Systems. IEEE Access. 2021;9:157727–60.
3. Eskandari M, Janjua ZH, Vecchio M, Antonelli F. Passban IDS: An Intelligent Anomaly-Based Intrusion Detection System for IoT Edge Devices. IEEE Internet Things J. 2020;7(8):6882–97.
4. Lansky J, Ali S, Mohammadi M, Majeed MK, Karim SHT, Rashidi S, et al. Deep Learning-Based Intrusion Detection Systems: A Systematic Review. IEEE Access. 2021;9:101574–99.
5. Golrang A, Golrang AM, Yayilgan SY, Elezaj O. A novel hybrid ids based on modified NSGAII-ANN and random forest. Electron. 2020;9(4):1–19.
6. Gu G, Fogla P, Dagon D, Lee W, Skorić B. Measuring intrusion detection capability. 2006;90–101.
7. Huang S, Lei K. IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks. Ad Hoc Networks. 2020;105.
8. Liu L, Ma Z, Meng W. Detection of multiple-mix-attack malicious nodes using perceptron-based trust in IoT networks. Futur Gener Comput Syst [Internet]. 2019;101:865–79. Available from: https://doi.org/10.1016/j.future.2019.07.021
9. Kumar S, Gupta S, Arora S. Research Trends in Network-Based Intrusion Detection Systems: A Review. IEEE Access. 2021;9:157761–79.
10. Gendreau AA, Moorman M. Survey of intrusion detection systems towards an end to end secure internet of things. Proc - 2016 IEEE 4th Int Conf Futur Internet Things Cloud, FiCloud 2016. 2016;84–90.
11. Nam K, Kim K. A Study on SDN security enhancement using open source IDS/IPS Suricata. 9th Int Conf Inf Commun Technol Converg ICT Converg Powered by Smart Intell ICTC 2018. 2018;1124–6.
12. Mohamed TS, Aydin S. IoT-Based Intrusion Detection Systems: A Review. Smart Sci [Internet]. 2022;10(4):265–82. Available from: https://doi.org/10.1080/23080477.2021.1972914
13. Yaacoub JPA, Salman O, Noura HN, Kaaniche N, Chehab A, Malli M. Cyber-physical systems security: Limitations, issues and future trends. Microprocess Microsyst [Internet]. 2020;77:103201.

Available from: https://doi.org/10.1016/j.micpro.2020.103201
14. Imoize AL, Oyedare T, Otuokere ME, Shetty S. Software Intrusion Detection Evaluation System: A Cost-Based Evaluation of Intrusion Detection Capability. Commun Netw. 2018;10(04):211–29.
15. Guruprasad S, Rio D'Souza GL. Chaos multiobjective evolutionary based technique to obtain accurate solutions in intrusion detection systems. Indian J Comput Sci Eng. 2020;11(2):188–94.
16. Klangjorhor J, Phanphaisarn A, Teeyakasem P, Chaiyawat P, Phinyo P, Settakorn J, et al. In vitro drug sensitivity (IDS) of patient-derived primary osteosarcoma cells as an early predictor of the clinical outcomes of osteosarcoma patients. Cancer Chemother Pharmacol [Internet]. 2020;85(6):1165–76. Available from: https://doi.org/10.1007/s00280-020-04081-5