Intrusion Detection Systems: Evaluating Machine Learning Techniques

JAYDON KING

CYS 485 – CYBERSECURITY ANALYSIS II

DR. AWAD MUSSA

SPRING 2024

1. Introduction:

With the increase in malware and threat actors in recent years, there is a need for software to be able to help alert users of attacks. Threat actors are using more and more sophisticated malware to attack users and companies around the world. These attacks have evolved to the point of being able to bypass firewalls and infect computers. With the change in threat actors' approach to infiltration, there is a need for a new form of detection. With the help of the evolution of Machine Learning and Artificial Intelligence in recent years, the use of Intrusion Detection Systems is increasing in value.

With IDSs being an older innovation in cybersecurity, different techniques can be used using different types of machine learning. These techniques have strengths and weaknesses, but their usage relies on the type of company that is going to be using them. While some of these techniques are not new, some of them have been shown to have major strengths when used for detection. With the evolution of artificial intelligence in the past five years, it can be inferred that this can help IDS increase their detection ratings. Another thing to think about is ensuring the detection rates are real threats and not false positives.

With Machine Learning evolving more and more, it is reasonable to think about implementing one of these systems in an IDS. It is, however, important to look at the various techniques that are available and compare their strengths and weaknesses through different tests conducted. The main point of this research is to analyze the different types of IDS techniques available, compare them, and determine what scenarios they would work in. The key techniques that will be compared will be the older Rule-based techniques and Machine Learning techniques. Other techniques will be discussed and compared as well, but the main objective of this research is to compare the two most common techniques. This will also take into account the amount of human intervention needed for them and how it impacts their effectiveness.

The literature used in this paper all came from Google Scholar, which was recommended by the professor. The only thing the papers had to meet was that they were some form of survey paper, conference paper, or journal. While researching the papers, many papers did not meet the criteria or did not pertain to the main topic of this paper. Some of the papers were unfortunately only accessible through a paid account, which was not used in the research. Others used keywords about the topic, but upon reading were about a different topic as a whole. The sources included were for the reasons of either discussing the pros and cons of different IDS techniques, including

tests of different IDS techniques, or discussing how the techniques work and how to implement them.

This research paper will start by describing what an IDS is, how it works, and its purpose in network security. It will then move to discussing the different techniques of IDSs, how they work, and how they are implemented. The next topic will be a comparison of the different techniques using tests from previous research. Lastly, the paper will discuss which technique is better before giving an overall review of IDS and how they may change soon.

2. Literature Review:

This section of the paper starts with explaining what an Intrusion Detection System is and its purpose in network security. Next, the paper explains the different techniques that are used in IDSs. Next, the section compares the different techniques to each other and how they perform in tests against different scenarios. Lastly, the review will determine which technique is most effective using its accuracy rating and false-positive rate.

A key part of this literature review is the works and papers I gathered that pertain to my topic. I used a summary table, Table 1, that displays all the works used in my research, what they discuss, and what they bring that is different from the rest of the papers. Some of the papers do have some overlap in information as they are about similar topics, but some have different approaches to displaying their idea. The summary table helped to display the information from the works and saved time by looking at the table rather than having to go to multiple different papers to find a detail. It also allowed me to draw better ideas from the papers and understand how most of them differ from one another.

Table 1: Summary Table

Authors/Year/Type/URL	Problem/Product	Intervention/Method/Improvement	Outcomes	Limitations/Gaps/Future Work
Abdallah, E. E., & Otoom, A. F. 2022 Survey https://www.sciencedirect.com/science/article/pii/S1877050922004422?ref=pdf_download&fr=RR-2&rr=870f3e7e596813cd	Accurately Detecting New attacks IDS systems have high false alarm rates	Implementing supervised machine learning in IDSs. Use of different learning models of machine learning to help IDSs predict new attacks. Utilize supervised learning algorithms to improve detection rate for IDSs as well as lower false positive rate.	Accuracy and False-positive rate (81% - 99% accuracy and .001%-.13% FPR)	Limitations/Gaps: Does not give solutions to issues it brings up in paper. Future Work: Giving a possible solution on all problems mentioned would be beneficial to the paper.

Ahmim, A., Maglaras, L., Ferrag, M. A., Derdour, M., & Janicke, H. 2019 Journal https://arxiv.org/pdf/1812.09059.pdf	An intrusion detection system that combines different approaches to better detect attacks.	Using different models of Decision Tree and Rules-based Models to create a hierarchical IDS Conducts an experiment to compare their model to other models on how they detect attacks and provides their accuracy	Accuracy (96% accurate)	Limitations/Gaps: Complex equations are not explained very well and lead to confusion. Overall topics discussed are also somewhat difficult to understand and convey. Future Work: Explaining complex topics in a manner that is easier for the reader to understand would help the research be able to convey their main point about their IDS system.
Da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. 2019 Survey https://d1wqtxts1xzle7.cloudfront.net/96594360/Internet-of-Things-A-survey-on-machine-learning-based-intrusion-detection-approaches-libre.pdf	Implementing IDSs onto IoT devices to detect and prevent attacks.	Implementing IDSs onto IoT devices as more attacks have been focusing them more and more. Goes over how to implement IDSs onto IoT systems and how they use datasets from other papers to test the security issues. Reduces false positive detection at the cost of increased training time.	Detection Rates Machine Learning Techniques	Limitations/Gaps: Does not answer issues that are brought up in the paper. Fails to explain the cost of increased training time to have low FPR. Future Work: Going into detail over the estimated cost of training time for the IDS would help provide a reason as to why increasing training time would be worth it.
Hamid, Y., Sugumaran, M., & Journaux, L. 2016 Journal https://hal.science/hal-01392098/document	Compare different Machine Learning techniques for IDSs.	Goes over various techniques in Weka that can be implemented in IDSs to improve detection. Techniques include Rule Based, Bayes Rule, Functions, Lazy Learners, Tree, and Misc. Conducts a test on all techniques and displays the Accuracy and False Positive rates for them.	Accuracy and False Positive Rates.	Limitations/Gaps: Paper starts discussing using IDSs for anomaly detection, which should not be the focus of IDSs. Future Work: Focusing on IDSs for detecting intrusions rather than anomalies will be beneficial to the topic of the whole paper.
Haq, N. F., Onik, A. R., Hridoy, M. A. K., Rafni, M., Shah, F. M., & Farid, D. M. 2015 Survey	Increase in internet users has lead to an increase in vulnerabilities in network security.	Gives an overview of IDSs and their use in network security Lists other research papers that discuss different techniques of machine learning used in IDSs.	Number of different datasets used Distribution of single	Limitations/Gaps: Paper mainly goes over what paper discuss the various techniques rather than

https://thesai.org/Downloads/IJARAI/Volume4No3/Paper_2-Application_of_Machine_Learning_Approaches_in_Intrusion_Detection_System.pdf	Implementing Machine Learning in IDSs to better detect attacks		classifiers in IDSs in various years.	explaining their implementation. Future Work: Discussing the implementation would help support their point for IDSs.
Liu, Q., Hagenmeyer, V., & Keller, H. B. 2021 Journal https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9395457	Implementation of Rule Learning-Based IDSs	Provides an overview of various Rule Learning-based IDSs found in other research Provides performance of different rule learning based IDSs found in their research	Accuracy and Error Rate	Limitations/Gaps: Paper requires a good memory to remember what is being discussed in the overall paper. Future Work: Focusing on the most used rule-based IDSs would help shorten the amount of information needed to remember.
Sajja, G. S., Mustafa, M., Ponnusamy, R., & Abdufattokhov, S 2021 Conference https://annalsofrscb.ro/index.php/journal/article/view/7837	Threats and attacks are becoming more complex and harder to detect.	Displays how the most common Machine Learning techniques in IDSs perform in experimentation.	Accuracy and Error Rate	Limitations/Gaps: Does not display a variety of different techniques. Only uses 3 techniques rather than some of the other techniques in other papers. Future Work: Utilizing other techniques in research would provide a more broad spectrum of techniques that could be beneficial to IDSs.
Sharma, R. K., Kalita, H. K., & Borah, P. 2016 Conference https://www.researchgate.net/profile/Rupam-Sharma-2/publication/277328598_Analysis_of_Machine_Learning_Techniques_Based_Intrusion_Detection_Systems/links/57c7a8fe08ae28c01d4f8d7d/Analysis-of-Machine-Learning-Techniques-Based-Intrusion-Detection-Systems.pdf	Utilizing Machine Learning in IDSs to help improve detection rate and lower false positive rates.	Explains supervised and unsupervised learning Explains commonly used machine learning techniques Displays Average detection rate or accuracy of various techniques.	Accuracy and detection rate	Limitations/Gaps: Very limited on the amount of techniques it describes compared to other research. Future Work: Providing more techniques used in IDSs would help give a better idea of what is available in IDSs when it comes to Machine Learning.

2.1 Intrusion Detection Systems:

Intrusion Detection System is a security tool or set of tools that is used to detect intrusions that occur on either the network or the host itself. Network Intrusion Detection Systems, or NIDS, detect intrusions through analyzing network traffic. Host-based Intrusion Detection Systems are split into four categories of analysis, File System monitor, Log file monitor, Connection analyzers, and Kernel-based IDS (Sharma et al., 2016). The most common form of Intrusion Detection is usually NIDS as they pertain to more devices rather than focusing on one device only. A scenario to use a Host-based would be in the case of a server that holds sensitive data that would need to be protected better than others (Abdallah, E. E., & Otoom, A. F, 2022).

Both types of Intrusion Systems have two key methodologies for how they detect intrusion: anomaly, and misuse-based detection. Anomaly detection uses previous usage data as well as monitoring current usage to determine what is considered normal for the system or network it is monitoring. Once it has the normal behavior determined, it then compares traffic that is coming through the IDS and determines whether or not it is an anomaly. If it does detect an anomaly, it will raise an alert of the anomaly and alert the proper security staff. This method is very useful for zero-day attacks that have little or no information about the attack. Misuse detection utilizes monitored events and uses signatures to check for an attack if it is already known. It uses already-known attacks from different databases and compares them to the traffic it is observing. If any of them match an attack, it throws an alert of the attack (Liu, Q., et al, 2021).

While both methodologies can detect attacks, they both have major disadvantages in their detection. Anomaly detection may be able to detect a zero-day attack, but it suffers from a high false-positive rate as well as difficulty in changing the "normal" usage. Misuse detection can detect already known attacks and has a lower false-positive, but it is unable to detect new attacks (Abdallah, E. E., & Otoom, A. F, 2022). With these disadvantages, there is the added risk of an attack getting through into the network. Even though it will never be completely risk-free, there are still better alternatives that use Machine Learning and Artificial Intelligence to better detect attacks.

2.2 Machine Learning IDS

With the advancements in Artificial Intelligence and Machine Learning in the past two decades, it has come very far since its beginning. Machine learning was first mentioned in 1959 by Arthur Samuel and defined as " a study that allows computers to learn knowledge without being programmed" (Haq, N. F et al, 2015). In recent years, AI has grown to be more powerful than

before with the introduction of GPT-4 and other AI bots used now. Machine learning is a subfield of AI that teaches a machine how to train itself for tasks (Sharma et al., 2016).

Machine learning is split into two different types: supervised learning and unsupervised learning. Supervised learning is when there is training data used in teaching the machine and includes the output of the vector. Unsupervised learning is where the machine learns from the environment more than the training data. The system still does learn from training data, but this data does not include the vector in the data (Sharma et al., 2016).

With the increase of cyberattacks getting through network firewalls and older IDSs, security professionals have started to implement machine learning algorithms into IDSs. The main objective of implementing Machine Learning is to allow for the prediction of attacks to occur from datasets of different known attacks. It also utilizes anomaly detection as well to compare normal traffic flow to the ones being observed as well as predict anomalies as well (Haq, N. F et al, 2015).

As stated by Sharma et al., several different techniques can be used when implementing Machine Learning into intrusion systems. Each of the techniques has different approaches and algorithms from one another. The most used techniques are artificial neural networks (ANN), decision trees, support vector machines (SVM), Bayesian classification, and self-organizing maps.

ANN machine learning uses layers that are meant to replicate neurons found in the human brain. This method allows for the IDS to gather network patterns in the input layer and process them in the deeper, hidden layers. The neurons that are on one layer are in some way connected to the neurons of the output layer. Then once the data is done processing, the detection results of the connection are output and display whether there was an attack detected.

The Decision Tree technique is noted as being one of the more simple techniques out there. It is essentially a bunch of if-then rules that the algorithm will go through until it finds the matching statement for the connection. The system starts at the beginning of the “tree and”, then it goes through the branches that have the scenarios until it finds a match of the scenario. Then, once the branch is found, it will use the leaf, or the solution, of the issue. This solution can be to alert you of an attack or to determine that it is a false positive.

The SVM technique uses the input vector provided by the network to map it into a higher-dimensional feature space. Then the machine creates two classes, one on each side of the space. The machine then processes the input vectors separates them into the two classes and puts them

on their corresponding side. It is essentially used to determine the normal traffic in a network and then anything that is not normal is considered an anomaly.

Bayesian Classification uses probabilities of hypothesis to learn and predict attacks. The machine assigns a probability to every hypothesis it uses from the network based on prior knowledge. Then once it has the probabilities linked to the corresponding hypothesis, it will classify them by using predictions of several hypotheses. These are then weighted out by their probabilities to determine their classification. Since the Bayesian method requires prior knowledge of probabilities, this technique is widely considered the most difficult to implement into an IDS.

Lastly, a Self-organizing Map, or SOM, is a form of unsupervised learning ANN that has units on a low-dimensional grid. Each unit is first assigned an initial weight vector. The initial vector is then compared to the known weight vectors found in the SOM. The weights are then updated after every iteration of the training. After the training is done, the input vector will then have an output vector and a distance between the input and every unit. The Best Matching Unit is then determined by finding the smallest distance between the input and output units.

2.3: Machine Learning IDS Effectiveness and Comparison

This section of the paper is going to use prior experiments and tests completed in my research to determine the effectiveness of each technique in Intrusion Detection. This section will also compare the techniques to each other and determine which one provides the best results. The outcomes that are going to be used that the tests have in common are their accuracy, sometimes called precision, and the false-positive rate sometimes referred to as error.

The first test results discussed are from Hamid et al. and provide a comparison between different techniques that are found in Weka, which is a collection of machine learning techniques. The techniques in Weka include more than the ones discussed earlier in other works, but they are like some of the other ones. The techniques are separated into different groups based on how they classify the inputs provided. The experiment KDD99 intrusion data set, which is also used in a test done by Saija et al.. The results of the test are found in Table 2 below.

Sno	Group	Technique	CC	TP Rate	FP Rate	Precision	Recall	F-Measure	RA	KS	MAE
1.	Rule Based	DT	99.757	0.998	0.001	0.998	0.998	0.997	1	0.9959	0.0014
		CR	78.538	0.785	0.061	0.677	0.785	0.713	0.937	0.6318	0.0203
		ZeroR	56.838	0.568	0.568	0.323	0.568	0.412	0.5	0	0.0514
		OneR	98.122	0.981	0.005	0.979	0.981	0.979	0.988	0.9682	0.0016
		PART	99.970	1	0	1	1	1	1	0.9995	0
2.	Bayes Rule	BayesNet	99.670	0.997	0	0.998	0.997	0.997	1	0.9944	0.0003
		NaiveBayes	92.748	0.927	0	0.989	0.927	0.949	1	0.8802	0.0063
		NBUpdatable	92.748	0.927	0	0.989	0.927	0.949	1	0.8802	0.0063
3.	Functions	MLP	98.75	0.988	0.004	0.977	0.988	0.982	0.998	0.979	0.0011
		SMO	99.552	0.996	0.001	0.995	0.996	0.995	0.999	0.9924	0.0793
		Simple Logistic	99.941	0.999	0	0.999	0.999	0.999	1	0.999	0.0001
4.	Lazy Learners	IB1	99.941	0.999	0	0.999	0.999	0.999	1	0.9999	0.0001
		IBk	99.941	0.999	0	0.999	0.999	0.999	1	0.9999	0.0001
		Kstar	99.904	0.999	0	0.999	0.99	0.999	1	0.9984	0.0001
		LWL	98.261	0.983	0.005	0.966	0.983	0.974	0.999	0.9702	0.0058
5.	Tree	DecisionStump	78.537	0.785	0.061	0.677	0.785	0.713	0.937	0.6318	0.0203
		J48	99.960	1	0	1	1	1	1	0.9993	0
6.	Meta-Algorithm	AdaboostM1	97.859	0979	0.005	0.962	0.979	0.97	0.993	0.9636	0.0478
7.	Misc	InputMappedClassifier	56.838	0.568	0.568	0.323	0.568	0.412	0.5	0	0.0514

Table 2: Results using the full dataset (Hamid et al., 2016)

With the above results, it is shown that the technique that performed the best is the PART technique with a 99.970 accuracy and false-positive rate of 0. The PART technique is defined as building a decision tree that makes the best leaf into a rule. This is similar to the decision tree technique mentioned earlier by Sharma et al. and Haq et al..

The next test results that are discussed are the ones from Ahmim et al. who also provided a comparison between different techniques. The thing that makes this test unique, however, is that it includes the authors' hierarchical system. Their system is based on a combination of both Decision Tree and Rules-base models, which are both noted as being pretty strong from other sources. The results from Ahmim et al. are presented in two tables that display similar data. The first one goes in-depth on the specific data in the dataset and the accuracy the techniques got on it. The second table displays the overall performance of the techniques as well as the training time and test time for the models. The results from both of the mentioned tables are listed below in Tables 3 and 4.

	Our Model	WISARD [18]	Forest PA [19]	J48 Consolidated [20]	LIBSVM [21]	FURIA [22]	Random Forest	REP Tree	MLP	Naive Bayes	Jrip	J48
TNR (BENIGN)	98.855%	97.135%	96.450%	93.355%	94.870%	96.835%	98.120%	95.165%	92.650%	66.545%	95.530%	94.960%
DR DDoS	99.879%	54.697%	99.818%	93.212%	55.970%	99.758%	99.818%	99.788%	91.212%	93.879%	99.667%	99.788%
DR DoS slowloris	97.758%	78.909%	92.848%	95.030%	78.182%	93.758%	93.758%	92.727%	78.485%	82.667%	93.333%	93.879%
DR DoS Slowhttptest	93.841%	23.353%	86.826%	83.832%	76.561%	78.358%	81.352%	75.364%	88.537%	70.060%	85.543%	80.325%
DR DoS Hulk	96.782%	67.600%	93.945%	95.891%	73.709%	98.655%	95.164%	92.218%	86.891%	73.782%	97.364%	93.600%
DR DoS GoldenEye	67.571%	48.714%	67.571%	67.143%	57.571%	65.143%	67.571%	66.429%	65.429%	62.143%	63.857%	67.286%
DR Heartbleed	100%	80.000%	100%	80.000%	0.000%	40.000%	100%	100%	0.000%	80.000%	80.000%	100%
DR PortScan	99.881%	51.407%	99.594%	99.046%	48.521%	87.118%	99.881%	99.881%	48.521%	99.499%	99.881%	98.569%
DR Bot	46.474%	1.442%	48.718%	52.083%	0.000%	48.077%	49.679%	47.756%	51.282%	29.968%	46.474%	47.756%
DR FTP-Patator	99.636%	0.000%	99.727%	100%	0.000%	99.636%	99.727%	99.182%	99.000%	99.455%	99.545%	99.545%
DR SSH-Patator	99.909%	0.000%	100%	99.727%	0.000%	100%	99.818%	100%	99.727%	99.182%	100%	100%
DR Web Attack - Brute Force	73.265%	4.694%	73.469%	55.102%	80.816%	49.796%	70.408%	70.816%	90.408%	5.102%	61.837%	60.408%
DR Web Attack - XSS	30.625%	1.250%	34.375%	48.750%	0.000%	38.750%	37.500%	32.500%	1.875%	91.875%	38.125%	41.250%
DR Web Attack - Sql Injection	50.000%	0.000%	50.000%	100%	0.000%	50.000%	100%	50.000%	50.000%	100%	75.000%	50.000%
DR Infiltration	100%	50.000%	83.333%	100%	0.000%	83.333%	83.333%	83.333%	16.667%	83.333%	100%	66.667%

Table 3: Performance by Attack Type and Benign; (Ahmim et al., 2019)

	Our Model	WISARD [18]	Forest PA [19]	J48 Consolidated [20]	LIBSVM [21]	FURIA [22]	Random Forest	REP Tree	MLP	Naive Bayes	Jrip	J48
FAR	1.145%	2.865%	3.550%	6.645%	5.130%	3.165%	1.880%	4.835%	7.350%	33.455%	4.470%	5.040%
DR (Overall)	94.475%	48.175%	92.920%	92.020%	54.595%	90.500%	93.050%	91.640%	77.830%	82.510%	93.400%	91.990%
Accuracy	96.665%	72.655%	94.685%	92.688%	74.733%	93.668%	95.585%	93.403%	85.240%	74.528%	94.465%	93.475%
Training Time	159.5 s	13.47 s	110.64 s	105.46 s	318.6 s	234.98 s	20.03 s	2.73 s	942.98 s	0.45 s	76.65 s	8.34 s
Test Time	2.27 s	243.28 s	0.99 s	0.61 s	343.96 s	0.96 s	1.7 s	0.52 s	1.61 s	12.39 s	0.51 s	1.12 s

Table 4: Overall Performance of Techniques; (Ahmim et al., 2019)

Looking at the results of Table 3 show that the model proposed by the writers performs well against other techniques and classifiers. Their model had the highest percentage in about half of the different attack types tested in this experiment. Some types, however, did not perform as well as others. That is not to say that it did not beat others that ended up being worse, but it did not outperform in all attack types.

The results from Table 4 show the overall false-positive rate, the detection rate, accuracy, training time, and test time for each of the models. As seen in the results, their model did the best on average in terms of low false positive rate, high detection rate, and high accuracy. Their model did not have the best training time or test time than some of the other models. The model with the best training time is the Naïve Bayes, a type of Bayesian classification, with a time of 0.45 seconds. That short time, however, led to a major increase in false positives and a decrease in detection rate and accuracy. The one with the shortest time is the Jrip method with a test time of .51 seconds. While it did perform better than most of the other systems, it did not perform as well as the authors'

hierarchical model. The main takeaway from this test is that even though the training time and test time are a little more than most, the performance makes their model worth implementing and using.

With these results in mind, we can have an idea as to how the different techniques compare to each other. We are then able to observe the performance of machine learning IDS, which outperforms normal, statically set IDS as it can adapt over time to changes in the system and behaviors. Machine Learning also allows for the IDS to be able to predict a new attack through these techniques, which is something not found in static rule-based IDS. Machine Learning also removes the need for a lot of human involvement as it would be able to come up with the rules and signatures and link them to solutions when they occur. Whereas with human interaction, there is the risk of human error, which can reduce the effectiveness of an IDS. The only important human involvement would be to monitor the machine to ensure it is training correctly and to provide data to it if going down that approach.

Even though Machine Learning would decrease the amount of human interaction involved in IDSs, some limitations are worth noting. One limitation would be the chance of bias within the algorithm. Most Machine Learning algorithms are not created on their own, there is someone who designs and programs them. It is worth noting that this could potentially lead to the IDS being biased on certain attacks, especially if it also decides on the possible solution for it. Another limitation would be the need for constant updates to the algorithms. As different attacks are occurring every day, the IDS would need to be updated to detect the latest attacks. This would require either automation through programming or manually updating them. Regardless of the way, there will still be the need to have some form of human interaction. Even with these limitations, though, the integration of Machine Learning would help improve detection rates as well as decrease false positives, which is a major improvement compared to older IDS systems.

3. Conclusion

In conclusion, this research paper has provided a comprehensive comparison of Intrusion Detection Systems and their integration with Machine Learning. Through this literature review, we discussed the need for IDS in networks, the limitations of rule-based systems, and the potential for ML to improve threat detection capabilities. With the increase in complexity of attacks, older IDSs that rely on human interaction are not enough to detect new attacks. Intrusion Detection Systems have been around for decades and have further evolved over the years thanks to newer

technology. The idea of Machine Learning has been around for over 70 years and has evolved into a field that shows a lot of potential within IDSs. With newer advancements coming out with AI and Machine Learning, there is potential for new features and better models to come out to be used.

Several techniques can be considered that offer different detection ratings and accuracies. The results found in the research have shown that Machine Learning is effective in Intrusion Detection, with some techniques performing better than others in certain categories. Something to consider, though, is the cost of implementing this kind of technology in a network. It would require a decent amount of time to train the machine, as well as to test it to ensure it is performing at an optimal rate. There is also the need for security analysts to understand Machine Learning, which is complex and difficult to understand for some. Outside of these issues, there is a lot of growth in performance to be offered from Machine Learning integration such as higher detection rates and fewer false positives. These are crucial for security analysts as they ensure that no attacks are coming through without them knowing. It also lowers the risk of several false positives appearing and covers the actual malicious attacks.

As mentioned in the last section, there are both pros and cons when it comes to implementing Machine Learning in IDSs. It is still valuable for an organization to look into this technology as it will grow with time and improve in the future. Proper research needs to be done to ensure they pick the correct technique for their business, as well as ensuring they have the right resources to implement it. It is also important to ensure that an organization evaluates its IDS and ensures it is performing as needed and no leakage of attacks is getting into the network. With this in mind, it can prove to be beneficial to utilize Machine Learning in IDS as it comes with many benefits and its potential to grow.

The amount of research focusing on recent advancements in Machine Learning is still somewhat hard to find, but there is most likely to be more as this technology grows. Newer research should look into the different techniques and figure out newer systems that take the best techniques into one. This idea is similar to the hierarchal method mentioned in the report done by Ahmim et al.. With the research provided in this paper, there is room for further research to be done and to find newer studies that can help improve this field of security. The most important aspect of the research is to find any newer techniques that come up and to determine their effectiveness and their false-positive rate. With this, organizations can determine if it is worth

implementing. Finally, any new ideas or methods that come up with Machine Learning should be investigated and evaluated to see if it would work in an IDS. With the findings from this research in mind, cybersecurity professionals need to consider implementing ML in IDS and explore ongoing research. With the use of ML, we can see that there are potential improvements in both detection of threats as well as lower false positives.

References

Abdallah, E. E., & Otoom, A. F. (2022). Intrusion detection systems using supervised machine learning techniques: a survey. Procedia Computer Science, 201, 205-212.0

Ahmim, A., Maglaras, L., Ferrag, M. A., Derdour, M., & Janicke, H. (2019, May). A novel hierarchical intrusion detection system based on decision tree and rules-based models. In 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS) (pp. 228-233). IEEE.

Da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. (2019). Internet of Things: A survey on machine learning-based intrusion detection approaches. Computer Networks, 151, 147-157.

Hamid, Y., Sugumaran, M., & Journaux, L. (2016, August). Machine learning techniques for intrusion detection: a comparative analysis. In Proceedings of the International Conference on Informatics and Analytics (pp. 1-6).

Haq, N. F., Onik, A. R., Hridoy, M. A. K., Rafni, M., Shah, F. M., & Farid, D. M. (2015). Application of machine learning approaches in intrusion detection system: a survey. IJARAI-International Journal of Advanced Research in Artificial Intelligence, 4(3), 9-18.

Liu, Q., Hagenmeyer, V., & Keller, H. B. (2021). A review of rule learning-based intrusion detection systems and their prospects in smart grids. IEEE Access, 9, 57542-57564.

Sajja, G. S., Mustafa, M., Ponnusamy, R., & Abdufattokhov, S. (2021). Machine learning algorithms in intrusion detection and classification. Annals of the Romanian Society for Cell Biology, 25(6), 12211-12219.

Sharma, R. K., Kalita, H. K., & Borah, P. (2016). Analysis of machine learning techniques based intrusion detection systems. In Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics: ICACNI 2015, Volume 2 (pp. 485-493). Springer India.