CSCM38: Advanced Topics: AI and Cyber Security – Report

Deep learning applications for network intrusion detection

Luke Hengstenberg

878876

10/04/2020

Table of Contents

1. Introduction

Sophisticated intrusion detection is integral to maintaining security within organisations and protecting sensitive information from malicious parties. Intrusion Detection Systems (IDS) are software applications or devices that monitor networks for malicious activity or policy violations [1]. Network-based IDSs (NIDS) are distributed devices that inspect network traffic traversing the devices they are positioned on, passively analysing and reporting suspicious behaviour [2]. NIDSs are typically divided into two classes: signature-based NIDS (SNIDS) and anomaly detection-based NIDS (ADNIDS). SNIDSs operate by looking at the signature inside data packets for a pattern matching that of a known attack pattern [3]. This allows for attack detection with high accuracy if an attack is known to the SNIDS. However, performance is limited to the pre-installed signatures thus rendering SNIDS ineffective in the detection of new or unknown attacks. ADNIDS operate by estimating the normal state of a system and raising an alarm if a certain amount of abnormal behaviour is detected [4]. This addresses the weakness of SNIDSs as unknown attacks should create an anomaly and will therefore be detected. However, ADNIDS based on certain techniques are not always accurate and can create issues due to raising false alarms.

The way in which the behaviour of a network can be accurately monitored by an NIDS, to establish a "normal" state and minimise false positives has posed a challenge in the scientific community. The demand and dynamicity of modern networks has increased due to receiving and handling larger amounts of traffic, with more focus placed on real-time human interaction. This increase in demand creates concerns that an attack could be missed by an NIDS as an overwhelming amount of data must be analysed at a lower level of specificity [5]. An attack that occurs undetected could have catastrophic consequences for an organisation and its users. It is therefore essential that an intrusion detection system maintains a high level of accuracy at all times.

Deep learning is a subclass of machine learning that uses "hierarchical levels of artificial neural networks" built in a similar structure to the human brain to processes data unconventionally and create patterns utilized for decision making [6]. Deep learning enables the learning and analysis of vast amounts of data, making it an attractive tool for use in an intrusion detection system. This report investigates how deep learning techniques can be used as a basis for an intrusion detection system that operates with a high level of accuracy. This report reviews the research of several academic publications, contrasting the similarities and differences in application of deep learning to network intrusion detection. This report aims to provide an overview of the research and how each contribution improves intrusion detection accuracy. Furthermore, this report evaluates the significance of contribution of each publication, identifying the strengths and weaknesses of each methodology. Overall, this report will attempt to establish the most accurate and best performing deep learning technique from the chosen publications, suggesting improvements where possible.

2. Review of Literature

This section conducts an in-depth literature review of several academic publications detailing and contrasting the proposed deep learning approaches to intrusion detection. Subsection 2.1 defines the problem the research addresses. Subsection 2.2 summarises and compares the methodologies. Subsection 2.3 lists the key findings and compiles experimental results to establish similarities and differences. Subsection 2.4 measures the comparative significance of each contribution, identifying strengths and weaknesses, and suggesting possible improvements.

2.1. Overview of Problem Definition

When investigating academic publications proposing deep learning approaches to intrusion detection a common problem definition can be deduced. All researchers feel the current state of intrusion detection is inadequate and therefore must be improved to address a variety of increasingly serious threats. Shone, et al. [9] express concern that most NIDSs continue to operate using signature-based techniques (SNIDS) instead of anomaly detection techniques (ADNIDS). [8] and [10] convey similar concerns, describing SNIDS as being "less effective" with performance suffering due to a "growing number of new cyber attacks". Yin, et al. [7] describe the method in which various unknown network attacks are identified as being an "unavoidable key technical issue" and how intrusion detection is typically "equivalent to a classification

problem" where network traffic is either a binary i.e. normal or anomalous, or five-category classification problem. [8], [9], and [10] share this view and have chosen to approach intrusion detection as a classification problem, but with differing methodologies and varying classifications. All researchers maintain the view that the current state of intrusion detection at their time of research can be improved, whether that's proposing an effective ADNIDS to replace obsolete SNIDSs [8, 9, 10], building on limitations of shallow machine learning techniques [7, 9], or creating a deep learning based NIDS with superior accuracy when compared to existing deep learning techniques [7, 8, 9, 10].

Moving forward this literature review will continuously evaluate each publication against its goal of improving the accuracy and efficiency of anomaly-based intrusion detection. Each publication will be cross-examined to identify similarities, novelties, and limitations. It is expected that the most modern piece of research, a technique proposed by Shone, et al. [9], will report the best performance and highest accuracy due to advancements in technology and the researcher's ability to relate to more existing work. This report will attempt to confirm or disprove this prediction in the succeeding sections.

2.2. Overview of Methodology

To solve the issue of accuracy in anomaly based IDSs each research piece proposed a unique methodology for realising their deep learning approach. This section is written on the assumption that the reader holds a level of preliminary knowledge of deep learning techniques to ensure each methodology can be covered succinctly and more focus can be placed on making comparisons. It is encouraged that each of the referenced academic publications is visited to gain a deeper understanding of the summarised content (*see references [7, 8, 9, 10] in section 4*).

Yin, et al. [7] proposed a supervised learning approach using recurrent neural networks (RNN-IDS) with data from an intrusion detection benchmark dataset called NSL-KDD. In this approach the data is first pre-processed by converting all non-numeric features to numerical values and normalizing them. The RNN-IDS model is then trained with the labelled pre-processed training set using forward propagation and backpropagation algorithms. [7] makes use of a directional loop within the hidden layer of the model that stores previous information about the feature data and alters the output passed to the next network layer accordingly. The forward propagation algorithm is used to calculate and pass on the output values and within the algorithm a soft-max regression function is called. The backpropagation algorithm is used to pass accumulated residuals to update weights. The RNN-IDS model is then tested with the testing dataset to measure its performance in binary classification (normal and anomaly) and 5-category classification (normal and four attack).

Javaid, et al. [8] proposed an unsupervised learning approach using self-taught learning (STL) with sparse autoencoder based feature learning and a soft-max regression classifier. Much like [7], the NSL-KDD dataset was used to benchmark performance and pre-processed by conversion of non-numeric values and normalization. Training similarities to [7] are seen in the first stage of [8] as a neural network applying a backpropagation algorithm is used for feature learning. However, the neural network differs as a sparse autoencoder is not recurrent and therefore does not use a directional loop. Another major difference is the learning is unsupervised, meaning the model is trained with an unlabelled training dataset. Once unsupervised feature learning is completed by the sparse autoencoder the second stage consists of applying the newly learned feature representation to the labelled training dataset using soft-max regression for classifier training. Soft-max regression is used in both [7] and [8] because it allows for multiclass classification, meaning performance in [8] is also measured for 5-category classification (normal and four attack) with the labelled testing dataset. However, the STL model differs because soft-max is applied directly instead of within the training algorithm as in [7]'s forward propagation algorithm. This may negatively impact the performance of [8] when compared to [7] and will be revisited in later sections. As well as 5-category, [8] measures performance for binary classification (normal and anomaly) and 23-category classification (normal and 22 attack).

Shone, et al. [9] proposed an unsupervised learning approach using stacked non-symmetric deep autoencoder (NDAE) based feature learning and random forest classification. [9] used the NSL-KDD dataset (much like [7] and [8]) and its predecessor the KDD Cup '99 dataset to benchmark performance and pre-processed the

datasets by conversion of non-numeric values and normalization. [9] takes a novel approach to intrusion detection by combining the power of deep learning through stacked non-symmetric deep autoencoders (NDAEs) and the speed of shallow learning with a Random Forest algorithm. The feature learning proposed by [9] is most similar to the STL approach proposed by [8] in its unsupervised learning approach based on an autoencoder. However, the extent of this similarity is minimal as the NDAE is designed to be non-symmetric meaning only the encoder part is utilised instead of both the encoder and decoder (as in [8]). The model in [9] also stacks NDAE's to create a deep learning hierarchy that allows complex relationships between features to be learned. When compared, the unsupervised learning approach proposed by [9] appears to be an expansion of [8], streamlining and improving feature extraction with their methodology. For the classification problem a shallow learning Random Forest classifier is trained using the encoded feature representations from the stacked NDAEs. The incorporation of a shallow learning technique makes [9] stand out from the other chosen publications and its impact on accuracy will be reviewed in section 2.3. [9] measures performance for 5-category classification (normal and four attack) for the NSL-KDD dataset (as with [7] and [8]) and the KDD Cup '99 dataset, as well as for 13-category classification (normal and 12 attack) with the NSL-KDD dataset.

K. Alrawashdeh and C. Purdy [10] proposed an unsupervised learning approach using Restricted Boltzmann Machine (RBM) based feature learning structured as a deep belief network and a logistic regression classifier. [10] uses the KDD Cup '99 seen in [9] which is known to have several defects when compared to its successor NSL-KDD. However, the methodology is reported as minimizing the effects of this in the models learning approach. Similarly, [10] pre-processes the dataset by conversion of non-numeric values and normalization. Much like [8] and [9], the model is trained in an unsupervised fashion using unlabelled training data. [10] proposes and trains two models, one comprised of a single RBM and one comprised of several RBM's forming a Deep Belief Network. One of its novelties is that the structure of an RBM is made of only input and hidden layers whereas the RNN [7], STL [8], and S-NDAE [9] IDS models all consist of input, output and hidden layers. This means alternate sampling and feature reduction was required using Gibbs sampling and contrastive divergence in what [10] deemed as a "pre-training" stage. Once the RBM has learned the features they are passed to the next network layer and further reduced until the final "fine-tuning" stage. In this stage a variation of the wake-sleep algorithm is used, and the resultant data is classified with soft-max regression. The proposed model by [10] is most like that of [9] due to its stacked hierarchical design and 5-category (normal and four attack) classification of the KDD Cup '99 dataset. However, these similarities are trivial and [10] presents an approach that is widely different from the other reviewed literature.

Studying each methodology showed that although the proposed deep learning approaches to intrusion detection share clear similarities overall each research piece is vastly different and novel in its own right. In the next section the key findings and experimental results will be discussed, relating the model design to performance.

2.3. Key Findings

Testing and evaluation of the previously explored anomaly-based IDS models varied in methodology with certain researchers taking a more thorough, real-world approach, and others taking a less realistic but performance focused approach. Each researcher carried out several experiments with similar and differing evaluation metrics. For the RNN-IDS model, Yin, et al. [7] chose to evaluate based on the following 7 metrics: Accuracy (AC), True Positive (TP), True Positive Rate (TPR), False Positive (FP), False Positive Rate (FPR), True Negative (TN), and False Negative (FN). Accuracy is the percentage of correctly classified records and was calculated using the total correctly classified records (TP and TN) against the total records. TPR is the percentage of correctly detected records and was calculated using the correctly detected records (TP) over the total anomalies (TP and FN). FPR is the exact opposite of TPR showing the percentage of incorrectly rejected records using the incorrectly rejected records (FP) over the total normal (FP and TN). These metrics are a standard within the mathematic and machine learning community for evaluating accuracy and classification. It therefore came as no surprise that these metrics and formulas were selected to evaluate the other IDS models.

In [8] and [9], an additional 2 metrics were used: Precision and F-Measure. The metrics also differed in the fact that TPR was termed "Recall" and FPR was termed "False Alarm" ([9] only). The precision metric was calculated as the percentage ratio of TP against TP and FP, measuring the amount of correct classifications (TP) impacted by incorrect classifications (FP). The F-Measure metric was calculated to find the harmonic mean between the precision and recall, discovering the balance between them. The inclusion of more metrics is a strength of [8] and [9] in comparison to [7] and [10] as the additional information can be used for a richer and more thorough evaluation. On the other hand, the focus of an intrusion detection system is accuracy so it can be argued that all publications provide sufficient evaluation metrics.

As discussed in the previous section the IDS models were trained and tested with data from the NSL-KDD dataset ([7], [8], [9]) and/or its predecessor the KDD Cup '99 dataset ([9], [10]). The experiments can be summarised as follows:

- Yin, et al. [7] used a personal laptop with average specs to perform binary classification (normal and anomaly) and five-category classification (normal and four attack) on $KDDTrain^{+}$ (training set), $KDDTest^{+}$ (testing set), and $KDDTest^{-22}$ (more difficult testing subset) included in NSL-KDD.
- Javaid, et al. [8] used an unspecified device to perform binary classification, five-category classification and 23-category classification (normal and 22 attack) on a 10-fold cross-validation of the NSL-KDD training set. Moreover, [8] performed binary classification and five-category classification on the NSL-KDD testing set.
- Shone, et al. [9] used a GPU-enabled high spec computer to perform five-category classification on a 10% sample of the KDD Cup '99 dataset (both training and testing). Furthermore, [9] performed 5-category classification and 13-category classification on a 10-fold cross-validation of the entire NSL-KDD dataset.
- K. Alrawashdeh and C. Purdy [10] used a personal laptop with low specs (lower than [7] and [9]) to perform five-category classification on a 10% sample of the testing data of the KDD Cup '99 dataset. [10] also experimented with continuous submission of 1000 record batches to simulate an active system.

Yin, et al. [7] trained the RNN-IDS model with varied hidden nodes and learning rates to discover when the model's accuracy is at its peak. For the binary classification task, the optimal number of hidden nodes was found to be 80 with a learning rate of 0.1. The model obtained an accuracy of 99.81% for the KDD training set, 83.28% for the KDD testing set, and 68.55% for the challenging KDD testing subset in an elapsed time of 5516s. For the five-category classification task, the optimal number of hidden nodes was also 80 with a learning rate of 0.5. The model obtained an accuracy of 99.53% for the KDD training set, 81.29% for the KDD testing set, and 64.67% for the challenging KDD testing subset in an elapsed time of 11444s. For the tested attack types (DoS, R2L, U2R, and Probe) the model reported TPRs of 83.49%, 24.69%, 11.50%, 83.40%, and FPRs of 2.06%, 0.80%, 0.07%, and 2.16% retrospectively.

Javaid, et al. [8] found the STL model obtained a similar accuracy for binary, five-category and 23-category classification tasks with the KDD training set, with accuracy ranging between 98-99%. Precision, Recall and F-Measure values were only recorded for binary classification with the KDD training set, with values ranging between 97.5%-99%. For the binary and five-category classification tasks with the KDD testing set the model reported accuracies of 88.39% and 79.10% retrospectively. For binary classification the precision, recall and f-measure results were 85.44%, 95.95% and 90.4%. In comparison to [7], the model has less accuracy using the training dataset, higher accuracy for binary classification with the test dataset, and lower accuracy for five-category classification with the test dataset. A limitation of [8] is the precision, recall and f-measure results are not presented individually for each of the attacks meaning this aspect cannot be compared to [7]. Furthermore, the presentation of results is done with bar charts meaning results not specifically referenced in text must be judged by eye.

Shone, et al. [9] found the S-NDAE model obtained an overall accuracy of 85.42% for the five-category classification and 89.22% for the 13-category classification of the NSL-KDD dataset (10-fold cross-validated sample). In comparison to the other models S-NDAE appears to produce the highest accuracy for five-category classification on the KDD dataset. However, the differences in the dataset sample used makes

this point questionable. In [7] and [8] the experiment is performed using the KDD testing set whereas [9] uses a 10-fold cross-validated sample of the entire KDD dataset. To accurately compare these models an identical dataset must be used. In comparison to [7], [9] reports higher recall or TPR values for DoS (94.58%/83.49%), Probe (94.67%/83.40%) and lower values for U2R (2.70%/11.50%) and R2L (3.82%/24.69%). Furthermore, [9] reports lower false alarm or FPR values for DoS (1.07%/2.06%) and higher values for Probe (16.84%/2.16%), U2R (50%/0.07%), R2L (3.45%/0.80%). This shows that for their datasets [7] was able to detect DoS and Probe attacks the best but raised more false alarms for Probe, U2R and R2L attacks. Again, the comparability of these statistics is questionable, and an identical dataset would need to be used as there was likely a differing occurrence of attacks.

[9] found the S-NDAE model obtained an overall accuracy of 97.85% for five-category classification of the KDD Cup '99 dataset (10% train/test sample). K. Alrawashdeh and C. Purdy [10] also performed experiments with variations of their model for five-category classification of a 10% KDD Cup '99 sample. [10] reported overall accuracies of 92% with a single RBM, 95% with a two hidden layer deep belief network (DBN), and a 97.9% accuracy with a four hidden layer DBN with soft-max logistic regression for fine tuning. Although [10] reports a higher accuracy than [9] with the KDD Cup '99 dataset the difference is marginal and at a greater cost to performance, with [9] showing on average training time for the S-NDAE model was 97.72% less. Moreover, [9] conducted a test of [10] within their research using the NSL-KDD dataset and found their method to be more accurate. Overall, [9] appears to have the highest accuracy for five-category classification using the KDD dataset with 85.42%, followed by [7] with 81.29%, [10] with 80.58%, and finally [8] with 79.10%. However, as previously mentioned [9] and [10] have been tested on a sample of the whole NSL-KDD dataset (both training and test), whereas [7] and [8] were tested with the testing set of NSL-KDD. This means the accuracies of [9] and [7] must be tested using the same sample to determine the most accurate. [9] also reported significantly less training time than [7], but again consistency in the device used is necessary to reach a conclusion as [9] used a much higher spec machine.

2.4. Significance of contribution

The anomaly-based intrusion detection models proposed by each academic publication were all significant to some degree. Yin, et al. [7] presented an RNN approach to intrusion detection that showed high accuracy for 2 out of 3 of the datasets and good performance with an average spec laptop. The experimentation was the most thorough in terms of testing accuracy for training, testing and a difficult testing dataset, allowing the models performance to be clearly shown for challenging data samples. In its own comparisons to related work [7] achieved better results than all selected traditional classification methods, therefore achieving its own aims for the model. The primary limitation of [7] is the model is trained for supervised classification using labelled training data. This means RNN-IDS has shown no application for unknown or new attacks making it infeasible for an intrusion detection system in the real-world. Furthermore, the model performed with an unacceptable accuracy (68.55%/64.67%) with the difficult testing set giving the impression that it would not adequately protect a real system. The contribution of [7] proves the use of recurrent neural network for intrusion detection from a scientific standpoint. However, from an industry standpoint the contribution is less significant, and the model has little to no application as an IDS. Improvements could be made by focusing on designing the model to work unsupervised (with unlabelled data), new intrusion types, and increasing performance using a GPU.

Javaid, et al. [8] presented an STL approach to intrusion detection that demonstrated sophisticated feature learning and extraction. The focus on an unsupervised self-teaching model demonstrates STLs application as a real-world IDS because it can be used for unknown/new attacks. [8] also performed the widest variety of classification tasks on multiple datasets and provides many evaluation metrics. The primary limitation of [8] is the research is purely focused on accuracy rather than both performance and accuracy. No information is given on the time taken or device used, therefore it is difficult to gauge how applicable [8] is as a real-world IDS. On top of showing capabilities for identifying unknown attacks [8] must also prove that it operates with high performance. Furthermore, the presentation of results was found to be the least useful when compared to the other publications. Improvements could be made by recording the model's performance, looking at improving accuracy with design variations such as stacked autoencoders, and considering how the technique could be incorporated into a real-time system.

Shone, et al. [9] presented a unique approach to intrusion detection that combined deep learning and shallow learning using stacked-NDAEs and a random forest classifier. [9] produced the highest accuracy and was the most performance focused, recording a multitude of time data and utilising a GPU to maximise efficiency. The model's high performance with an unsupervised learning process shows [9] has real-world application for the accurate detection of unknown or new attacks with a high level of efficiency. The reporter feels the contribution of [9] is superior to that of [7], [8] and [10] because it performed the best, has the most real-world application and was evaluated with a large amount of data for multiple classification types. The primary limitation is that the dataset samples taken through 10-fold cross-validation may have provided a dataset allowing the model to obtain a higher accuracy than it would've if the dataset from [7] or [8] was used. Despite this, it is felt that [9]'s contribution is no less significant, and the model would be likely to perform highly with different samples. Improvements could be made by measuring the model's performance with real network traffic and real-time NIDS simulation. Furthermore, the obtained accuracy of 85.42% is not sufficient for fully securing a network and shows room for improvement before [9]'s use in industry.

K. Alrawashdeh and C. Purdy [10] presented an RBM based deep belief network with soft-max logistic regression that incorporated unsupervised learning and facilitated human interaction. [10] displayed higher accuracy than its chosen related work and demonstrated how variations of the model using layers of RBM's could increase the accuracy. The approach detailed a user interface for modifying the number of hidden layers and learning rate dynamically. This feature set [10] apart from the others and showed the contributions real-world application. [10] demonstrated using the model as an online system for intrusion detection by continuously testing batches of 1000 samples, maintaining both an industry and scientific viewpoint of the model's potential. The primary limitation of [10] is the choice in dataset and classification task. Only the KDD Cup '99 was used which is known to suffer from many defects and only five-category classification was performed. The lack of dynamicity in the testing creates concern that the model's performance with other datasets and real data would be poor. Furthermore, [9] proved it was superior in both accuracy and efficiency with its own testing of both models. Improvements could be made by testing the model with more realistic data, increasing the accuracy with design modifications, and utilising hardware to increase efficiency as training time was too high.

3. Conclusion

In conclusion this report reviewed four academic publications detailing deep learning approaches to intrusion detection. The purpose was to understand how deep learning can be utilised to detect anomalies in network behaviour to warn of a possible attack. It was discovered that recurrent neural networks [7] can provide a high accuracy supervised intrusion detection approach with efficient performance. However, it was also concluded that a real-world system would be required to operate in an unsupervised fashion with unlabelled data as seen in [8], [9] and [10]. The ways in which a self-taught learning technique can be used for superior feature learning and extraction was shown by [8], stressing the importance of strong self-teaching for detecting unknown types of intrusion. However, it was also clear that complex feature learning has an impact on performance as seen in [10]'s RBM deep belief network, and that the efficiency of an NIDS is as important as the accuracy. The reporter's prediction that the most modern publication [9] would show the highest accuracy and best efficiency was proven by the superior performance of the stacked non-symmetric deep auto-encoder combined with random forest. Finally, all reviewed literature showed significance within the scientific community and respective field of research, but in their current state none of the proposed models could be used as an intrusion detection system in industry. The defects in both the KDD Cup '99 and NSL-KDD datasets damage the credibility of the model's performance evaluations as neither dataset is representative of real network data. The actual accuracy of all techniques is unknown and will remain so until testing of a real network can take place. Nevertheless, the literature has demonstrated deep learning techniques can be used for anomaly-based intrusion detection so it is only a matter of time before researchers have accessibility to the right resources and can take full advantage of deep learnings capabilities.

4. References

[1] barracuda, "Intrusion Detection System," barracuda.com, 2020, [Online], Available: https://www.barracuda.com/glossary/intrusion-detection-system, [Accessed Mar. 31, 2020].

[2] J. Burton, I. Dubrawsky, V. Osipov, C. T. Baumrucker, and M. Sweeney, "Chapter 1 - Introduction to Intrusion Detection Systems," in *Cisco Security Professional's Guide to Secure Intrusion Detection Systems*, Syngress, 2003, pp. 1-38, doi: https://doi.org/10.1016/B978-193226669-6/50021-5.

[3] S. N. Shah, and P. Singh, "Signature-Based Network Intrusion Detection System Using SNORT And WINPCAP," in *International Journal of Engineering Research & Technology (IJERT) 1*, no. 10, 2012, pp. 1-7.

[4] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," in *computers & security 28*, no. 1-2, 2009, pp. 18-28.

[5] T. Wilhelm, and J. Andress, "Chapter 8 - Use of Timing to Enter an Area," in *Ninja Hacking*, Syngress, 2011, pp. 119-134, doi: *https://doi.org/10.1016/B978-1-59749-588-2.00008-1*.

[6] M. Hargrave, "Deep Learning," Investopedia.com, Apr. 30, 2019, [Online], Available: https://www.investopedia.com/terms/d/deep-learning.asp, [Accessed Apr. 1, 2020].

[7] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," in *Ieee Access 5*, 2017, pp. 21954-21961.

[8] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, 2016, pp. 21-26.

[9] N. Shone, T. N. Ngoc, and V. D. Phai, "A deep learning approach to network intrusion detection," in *IEEE Transactions on Emerging Topics in Computational Intelligence 2*, no. 1, 2018, pp. 41-50.

[10] K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 195-200.