Comparison of entropy-based and machine learning approaches in intrusion detection

Slavko Gajin* and Valentina Timčenko **
* School of Electrical Engineering, Belgrade, Serbia
** School of Electrical Engineering, Mihailo Pupin Institute, University of Belgrade, Belgrade, Serbia
slavko.gajin@rcub.bg.ac.rs, valentina.timcenko@pupin.rs

***Abstract*— This paper provides an analysis and comparison of the most important characteristics of entropy-based techniques and different categories of machine learning approaches. The main goal is to better understand the existing techniques and results, which can be found in a wide range of scientific studies. Although both classes of approaches rely on the network traffic structure analysis, the entropy-based techniques are inherently much simpler for the application, and in some cases can easily deal with the detection of the zero-day attacks. On the other side, due to the lack or presence of the labels, the machine learning algorithms can be applied for specific attack cases, and under such conditions providing more accuracy when compared with some entropy-based methods.**

I. INTRODUCTION

The aim of network intrusion detection, as a part of a wider area of network behaviour analysis, is to inspect network activities in real-time and detect the cybersecurity threats in the early phase before the main security breach occurs [1], [2]. Both entropy-based and machine learning approaches rely on network traffic structure rather than a deep packet inspection, which is based on the attack signature in the packet payload. This characteristic makes them very applicable for detecting security threats on encrypted traffic and occupy a large interest in the research community in networking and cybersecurity. However, the differences between entropy-based and machine learning techniques are significant in many aspects and their results can be hardly directly compared. Entropy-based detection techniques are attractive due to their simplicity and applicability in real-time network traffic, with no need to train the system with labelled data, which allows detection of the zero-day attacks and suspicious behaviour even for the traffic unknown in advance. On the other hand, both supervised and unsupervised machine learning approaches cover many different techniques and algorithms, and their applicability and detection efficiency depend on the use cases and available features that present the network structure.

In this paper, we analyse the most important characteristics of entropy-based techniques and both supervised and unsupervised machine learning approaches to better understand the results that are widely presented in the scientific literature.

The rest of this paper is organized as follows: Section 2 discusses the scope of this research, a wide range of referent literature and already applied solutions, as well the problems that arise from the necessity of applying the cutting-edge techniques and algorithms in the context of the network intrusion detection systems (IDS). Section 3 tackles the motivation, objectives and methodology used in this research. In section 4, we analyse and evaluate the entropy-based technique, while Section 5 addresses the machine learning techniques. Section 6 concludes the paper and summarise the main differences between the entropy-based and machine learning approaches.

II. RESEARCH PROBLEM AND RELATED WORK

The steady increase of the malicious network activities has implicated the bustling research and development activities in network anomaly and intrusion detection systems (NADS/NIDS), both in academia and in industry. The existing approaches clearly indicate the need to examine the attacker activity from different angles. In [3], anomaly detection systems are categorized based on a normal behaviour specification, systems relying on standard and trained mechanisms. The standard mechanisms use normal behaviour characteristics, taking as a base set of rules or protocols that specify this behaviour (e.g. TCP and its three-way handshake procedure for the connection setup can be used to detect half-open connections and delete them from the queue). The trained specification mechanisms use trained characteristics related to normal behaviour collected by the means of network traffic and system behaviour monitoring, whereas rely on defining threshold values depending on a range of traffic parameters. In the case when traffic exceeds defined values, it is regarded as anomalous. If considering the time-series data that represent network activities, anomaly detection approaches can be seen through a specific macro-classification into the methods relying on the statistical, probabilistic, proximity-based, clustering-based and prediction-based methods [4].

Some recent studies support the wider use of flow-based techniques, where the use of the Cisco NetFlow protocol is most frequently used for network traffic monitoring and data collection [5], [6]. Due to its simplicity to collect statistical data about network communication from network devices, the use of the flow data instead of the full network packets has proven to be highly beneficial [7]. Entropy-based techniques can be applied in the context of the NetFlow feature distributions [9], as it is showed that there is a strong correlation between address and port features,

especially in the case of using bidirectional data flows. A group of authors has used the parameterized Rényi and Tsallis entropy-based anomaly detection along with the supervised machine learning methodology, obtaining, as a result, a low false-positive rate [10]. The results indicated the dominance of the parametrized entropies over the generally applied Shannon entropy. In all cases, the entropy-based approaches produce a number of time time-series values, which are analysed for anomalous behaviour [8]. However, some studies have reported the weaknesses of the entropy-based approaches, mostly related to deceiving the flow-based detection solutions in the context of the additionally injected spoofed DDoS traffic [11].

The introduction of the machine learning techniques to the NADS and NIDS systems has brought new air into this area. The continuous appearance of different "zero-day" attacks, that correspond to the previously unidentified events, which are therefore unable to be analysed and, in many cases, can skip the detection system procedures, leads to the necessity for the introduction of different, sometimes highly adaptive, data-driven machine learning algorithms for the needs of the intrusion detection.

However, the use of machine learning techniques has brought a bag of different challenges to deal with, such as techniques for data handling, cleaning, and preparing, proper selection of features, choice of an adequate model and settings for hyper-parameters, techniques to deal with the underfitting and overfitting (the bias/variance trade-off), as well as the proper data dimensionality reduction. Nevertheless, the main issue arises from having available the data labels for supervised algorithms and if the data is realistic enough to be used in the analysis for better detection algorithms research [12].

The practise has proved that the complex networking area, different communication behaviour, and specific service needs cannot rely on the solutions that are built on the anomaly detection methods with a single algorithm approach, neither with only one machine learning technique (artificial neural-networks, pattern matching, etc.). The researchers should follow the recent advances in machine learning that suggest hybrid approaches, as well as the cascading of multiple methods.

III. MOTIVATION, OBJECTIVES AND METHODOLOGY

When presenting the complex scientific contributions in papers with very limited pages, it can often lead to selecting only a few, isolated results, that demonstrate just the best characteristics of the presented research contribution. Accordingly, many research papers in network behaviour analysis claim a high efficiency of the proposed approaches when using entropy and machine learning techniques. However, it is not rare that the presented results demonstrate just some isolated cases, forcing the efficiency as high as possible by fine-tuning the calculating parameters, and choosing the most appropriate datasets and features which sometimes can be unintentionally biased. In some cases, the researchers do not thoroughly explain the research environment information, thus leaving some important details hidden. Only deeper research and experiments in the field can revile other aspects that are not well addressed. Although there are numerous published studies, there is still a lot of misunderstanding left, where some results that hold for the isolated setup are generalised, leading to wrong conclusions.

Therefore, the motivation behind our work is to explain some of those details by comparing the two most frequent approaches in network behaviour analysis – entropy-based and machine learning techniques. The main objectives of this research are directed towards the comprehensive analysis of both categories of techniques, to clearly explain their advantages and disadvantages, especially in terms of practical usage. Starting from that premise, the characteristics of entropy-based and machine learning techniques in network behaviour analysis can be analysed in a much wider context, especially when considering the applicability in the real-world environment.

IV. ENTROPY-BASED APPROACH

The entropy-based network traffic anomaly detection in many aspects completely differs from the machine learning methods, which makes it difficult and even impossible to directly compare their performances. For that reason, we rather discuss their general characteristics and leave a decision which is better for specific use-cases.

Information about network communications, the so-called *flows*, can be easily exported from the network routers using standardized IPFIX protocols or similar industry standards, such as NetFlow, JFlow, NetStream, etc. The original flows are unidirectional, and further pre-processing is required to convert them to bidirectional flows which are shown to provide more information for accurate anomaly detection.

In general, the flow-based entropy approach relies on the use of two groups of features, the identification attributes (5-tuple), which correspond to the source and destination IP addresses, protocol and port numbers, and the traffic volumetric features that include the total packet and byte number. The volumetric features are useful for traffic-intensive anomalies, as in the case of Distributed Denial of Service (DDoS).

Entropy is calculated over the distribution as a measure of the data evenness. Data distributions are generated during certain time intervals, the so-called *epochs*, by the process of aggregation, where some of the features are chosen as the aggregation keys, while others are further counted (flow records) or summarized (bytes and packets). For instance, the aggregation can be realized by source IP, destination IP, source port, destination port, and their combinations. To better recognized hidden structure in network communications and detect low-intensive anomalies, the third type of features are calculated, the so-called *behaviour features*. They present the total number of distinct occurrences of other features, calculated by the second-degree aggregation, such as the number of distinct destination IP addresses in the communication with a certain source IP address, which is illustrated in Figure 1.

Figure 1. The procedure of the aggregation and feature calculation.

Entropy is calculated per epoch for each data distribution generated by the aggregation process, resulting in many time series entropy data (one feature value per epoch). A significant change in the data distribution due to changes in network communication activities usually occurs as a spike in the distribution, which results in sudden entropy changes. These changes can be detected by using a windowing mechanism or EMA (Exponential Moving Average) [14], calculating the margin of accepted values. When the calculated entropy value crosses the margins, the system raises an alarm as such an event is recognizes as an anomaly. These anomalies are an indication of security threats, and therefore further root cause analysis from raw flow records is needed to extract the information about the attackers, victims and services used.

A peak in data distribution, that causes the entropy change, can be suppressed by adding additional data to extend the distribution tail to make the distribution more even. Consequently, the entropy change can be annulled, and the anomaly is deceived. This method can be used by some attackers, which are generating spoofed traffic in parallel to the attack, to camouflage the attack and deceive the entropy-based anomaly detection systems [11].

A. Evaluation

The entropy-based methods can be categorized as non-parameterized, as the well known Shannon entropy, or parameterized, such as Renyi or Tsallis entropies.

The Shannon entropy is defined as:

$$H_S(X) = \sum_{i=1}^{N} p(x_i)\ \log_a \frac{1}{p(x_i)} \qquad (1)$$

where N is a total number of elements in the distribution, while $p(x_i)$ is an empirical probability, calculated by the relative contribution of element x_i with value m_i in the total sum of all values, M:

$$p(x_i) = \frac{m_i}{M},\ M = \sum_{i=1}^{N} m_i \qquad (2)$$

The Shannon entropy relies on a trade-off between achieved influences from the main mass of the distribution and the tail part of the distribution, which can be controlled using parameterized entropy generalizations, Renyi (eq. 3) and Tsallis (eq. 4) [15], [16]. Both rely on setting the specified parameter α. When having a positive value, this parameter uncovers the main mass in the distribution, which represents the concentration of the events that arise often. In the case when this value is negative, it refers to the tail of the distribution, indicating the dispersion that is produced by the seldom events.

$$H_R(X) = \frac{1}{1-\alpha}\log_b\left(\sum_{i=1}^{N} p(x_i)^{\alpha}\right) \qquad (3)$$

$$H_T(X) = \frac{1}{1-\alpha}\left(\sum_{i=1}^{N} p(x_i)^{\alpha} - 1\right) \qquad (4)$$

Additionally, it is useful to apply a specific scaling factor to normalize the entropy to a value of 1 for fully randomized distribution. According to this procedure, the scaling factor that is applied for the Shannon and Rényi is presented as $1/\log_b N$, while for Tsallis it is $(1-\alpha)/(N^{1-\alpha}-1)$. Shannon entropy is always giving values in the range between 0 and 1, which is also a case with Rényi and Tsallis entropies when α parameter has a positive value. The negative α generates entropies with values above 1.

Certain entropy research studies highlight the dominance of the Tsallis and Rényi entropies over the Shannon entropy, mostly pointing to the results that indicate better detection of peaks or tails in the obtained feature distributions [10], [17]. Based on our experiments and analysis, it seems that their deductions are strongly dependent on the used detection techniques, choice of the data and selection of features that are further used for the experiments, thus, we believe that their conclusion cannot be generalised. Accordingly, we evaluate these entropies considering the feature sensitivity to detect an anomaly.

The experience in using Rényi and Tsallis entropy shows that the optimal performances can be obtained for the fixed value of α (+2 and -2). To estimate the behaviour of the evaluated entropies more thoroughly, we consider a reference distribution of 100 elements, defined by the reciprocal function 1/x, shown in Figure 2, taking values of 100, 50, 33, 25, and terminated with 'a long tail' of elements with value 1. The test result indicates that obtained distribution nearly approximates the deviation of the flow feature values distribution obtainable in a real network traffic case.

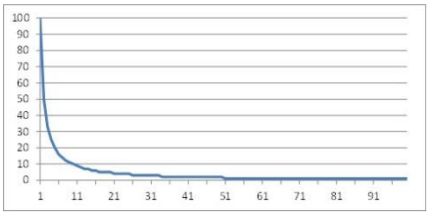

Figure 2. Generic data distribution: the long tail of elements with value 1.

There are two scenarios for entropy changes, when increasing the peak and when increasing the tail of the distribution. When progressively increasing the peak of the distribution, from the value 100 to 1000, the entropy of all discussed types is changed as presented in Figure 3. The Shannon, as well as Rényi and Tsallis entropies with the positive value of the parameter α, result in decreased values. In the case of the negative α, the entropy increases. In the case of Tsallis entropy, the increase is much higher and it is separately presented in Figure 5.

Conversely, when increasing the tail of the distribution taking the value from range 1 to 1000, the introduction of the new elements with a value of 1, involves more similarities in the data, while as a result, the entropies approximate the value 1, as presented in Figure 4.

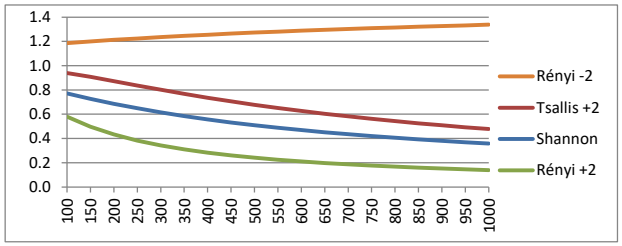

Figure 3. The entropy change: increase of the distribution peak.

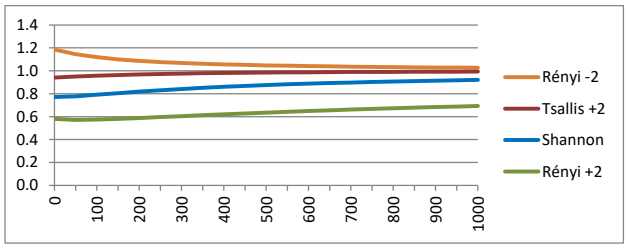

Figure 4. The entropy change: increase of the distribution tail.

In the case of the positive parameter, Rényi entropy provides the lowest values, while Tsallis contributes with much higher values (in a range from 1.7 to 106, which is not shown as these values are out of the scale). The obtained results indicate that when having entropy with lower values it is more difficult to detect a drop, particularly in the context of higher standard deviation. It could be seen in the case of Rényi entropy when having a positive parameter, that manifests more sensitivity to the regular variation of data (the highest slope presented in Figure 3) while providing the lowest values.

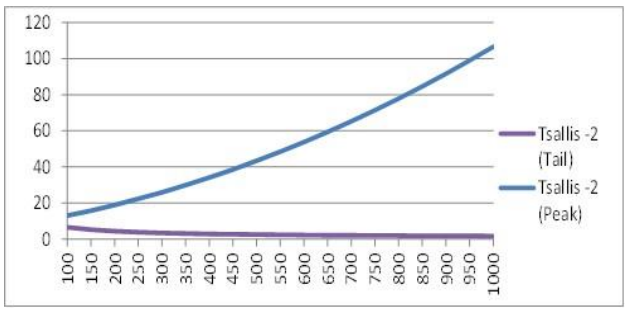

Figure 5. Tsallis entropy, negative α, and values from 100 to 1000.

To evaluate anomaly detection performances, we used the CTU-13 dataset with the regular background traffic taken from the Czech Technical University network. Then, we have modelled the DDoS communication pattern using synthetically generated anomalies, gradually increasing the anomalous flows in different epochs, taking the values of 10, 25, 50, 100, 200, 500 and 5000 flows.

Since the targeted IP address receives a huge amount of traffic during the DDoS attack, a spike appears in the data distribution of the traditionally used byte count feature aggregated by the destination IP address. As the result, it makes corresponding drops of Shannon entropy, which are presented in Figure 6.

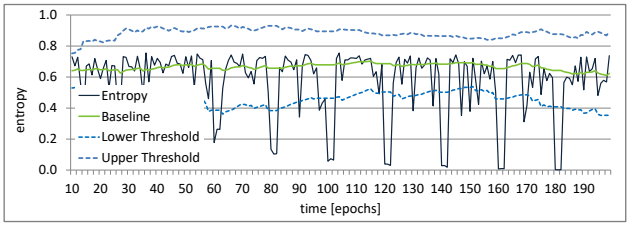

Figure 6. Shannon entropy of the source byte aggregated by destination IP address.

At the same time, the amount of regular traffic variates significantly, which also makes a lot of additional drops below the threshold, generating false positive alarms.

Since Rényi entropy with a positive parameter (α=2) generates the lowest values according to Figure 3, a high standard deviation of the byte count feature results in the negative lower threshold, which is shown in Figure 7. Obviously, this negative threshold can never be crossed, which makes the Rényi entropy of this feature completely useless.

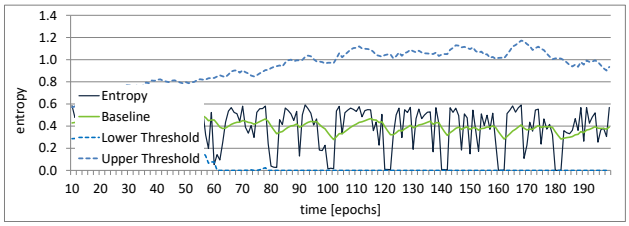

Figure 7. Rényi entropy (α=2) of the source byte aggregated by destination IP address.

On the other hand, the behaviour feature that counts distinct destination ports aggregated by the destination IP address, presented in Figure 8, accurately detects all the anomalies except the two with the lowest intensity, with no false-positive alarms.

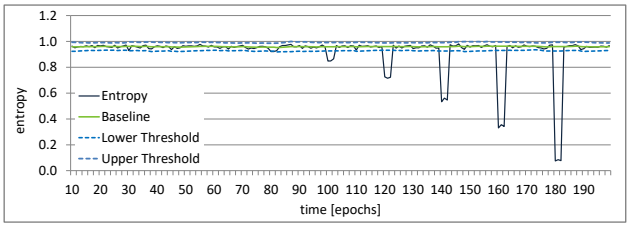

Figure 8. Rényi entropy (α=2) of the destination port feature aggregated by destination IP address.

Similar results are achieved with other entropy types, which demonstrates a higher quality of the behaviour feature in comparison to the volumetric features. Table 1 presents the number of detected anomalies in the aforementioned dataset with seven series of synthetically generated DDoS attack. The source and destination IP address and port, and flow count are labelled with the letters ‘S’, ‘D’, ‘s’, ‘d’, and ‘f’ respectively, while the labels in the square brackets relate to the aggregation key.

It is obvious that there is no single entropy type that outperforms the others in all cases.

For that reason, we believe that the right selection of the entropy type is not a straightforward task and should consider many aspects, such as specific network traffic and its variety and deviations, a technique used for entropy change detection, as well as previously analysed resilient to deception.

TABLE I.
NUMBER OF DETECTED ANOMALIES BY DIFFERENT ENTROPY TYPES

Features	Shannon	Renyi +2	Renyi -2	Tsallis +2	Tsallis -2
f[S]	2	1	3	2	2
S[D]	2	1	3	1	3
d[D]	5	6	5	5	5
S[s]	5	6	4	4	6
d[s]	4	3	4	3	5
S[d]	4	2	3	2	3
f[S.d]	2	1	2	2	2
S[D.s]	6	6	4	4	6

V. MACHINE LEARNING APPROACH

In contrast to the entropy-based approach, which is conducted on the time interval level (per epoch), machine learning methods provide anomaly detection granularity on the data level and classify every data point as either normal or anomalous. For that reason, the identification flow attributes defined in the flow 5-tuple cannot be used since these could lead to the overfitting of the model, which would then learn about the attacker and victims instead of the attack characteristics. Therefore, there is a need to use additional features, such as packet size, round trip time, TCP window, TCP flags, their minimum, maximum, and average values, as well as standard deviations. Since most of the routers cannot export these features, they must be extracted and calculated from the raw network traffic, which involves additional processing and complexity.

Besides, the supervised machine learning techniques require data labelling. If the goal of the detection is to classify the attack, then the attack type must be also labelled. As an advantage, this individual data labelling can allow calculation of various performance metrics, such as the number of True/False Positive/Negative instances, Accuracy, Precision, Recall, F-Measure, and ROC curve characteristics etc. These metrics are inconvenient for application in the entropy-based approach. Still, data labelling is highly impractical in a real-world environment and running a corporate network.

Even in the research area, the lack of reliable datasets that reflect real, up to date regular and anomalous traffic profiles and encompasses a variety of network anomalies, attacks and their features is one of the main issues. Some of such datasets that are available for scientific experiments are CTU-13, UNSW-NB15, CIC-IDS-2017, CIC-DDOS-2019 and others. For the application in a real-life environment, there is also a need to encompass all the diversity of the network types (ISP, corporate, small office, academic), usage and user activities, with a focus on the services, communication pattern, traffic volumes.

Among the machine learning algorithms, the unsupervised category encompasses different algorithms which provide specific complexity, speed, and applicability. In general, the unsupervised algorithms are found as easier to implement in a real-life network, but harder to evaluate than supervised.

A. Evaluation

In our experiments, we have used one of the latest and most complete publicly available datasets, namely the CICIDS2017 [13]. It is flow-based labelled datasets, which includes the most common attack scenarios, covering the profiles of Web-based, Brute force, DoS, DDoS, Infiltration, Heartbleed, Botnet and Scan attacks, each in a different file named according to the weekday when the dataset was created. More than 80 flow-based features related to network communications were generated by processing real traffic with simulated attacks, which can be used for both entropy-based and machine learning techniques.

The authors of the CICIDS2017 achieved very high detection performances of the web attack using supervised machine learning algorithms [13]. We confirmed this finding by applying Random Forest (RF), Multilayer Perceptron (MLP), and Naive-Bayes (NB) algorithms to the same set of features: initial TCP window size in both directions and the total bytes transferred from the source to destination.

As the size of the TCP window can be defined to have an arbitrary value, even in the case of the attack communication, our idea was to evaluate the anomaly detection ability of the evaluated machine learning algorithms in the case of different TCP window size values. Thus, we have slightly modified data and changed TCP window size by 3%, 10% and 30%, from the size defined in the comparative study. When repeating the experiment with this way changed window size, very poor results are obtained (Table II). It is especially interesting in the case of the MLP which even for the slightest change of the value, 3%, was not able to detect any attack at all, while for RF the detection capability is severely lost. The NB showed to be more resilient to the TCP window size changes but having the lowest performances.

TABLE II.
SUPERVISED MACHINE LEARNING PERFORMANCE EVALUATION

Alg.	Dataset	Precision	Recall	F1
RF	**Original**	**0.850**	**0.981**	**0.911**
	Modified, 3%	0.176	0.037	0.061
	Modified, 10%	0.176	0.037	0.061
	Modified, 30%	0.176	0.037	0.061
MLP	**Original**	**0.771**	**0.840**	**0.804**
	Modified, 3%	0.000	0.000	N/A
	Modified, 10%	0.000	0.000	N/A
	Modified, 30%	0.000	0.000	N/A
NB	**Original**	**0.132**	**0.909**	**0.230**
	Modified, 3%	0.132	0.908	0.230
	Modified, 10%	0.123	0.842	0.215
	Modified, 30%	0.123	0.842	0.215

The obtained results point out that in the case of some machine learning algorithms that rely on such specific feature values can be deceived without too many efforts with a slight change in the attack setup. Consequently, it is obvious that some straightforward measurement over specific features can mislead in dataset evaluation, thus it is recommendable to proceed with some additional analysis of raw data and to explore the meaning of the features more profoundly in the context of the used machine learning algorithms.

VI. CONCLUSION

In this paper, we have evaluated the entropy-based and machine learning approaches from the intrusion detection capabilities perspective and compared the features of these two approaches from several characteristic aspects.

We have thoroughly compared the three most popular entropies, Shannon, Tsallis and Renyi, and estimated their response to changes in the feature distributions. The proper choice of the used entropy depends on the network traffic characteristics, variations and specific deviations, as well as of selected features and parameters, which can have a strong impact on the resiliency to the entropy deception.

From the application of the machine learning algorithms overview, it can be noticed that the supervised algorithms contribute with significant limitations for the potential use in the real-case networks. It is also indicated that the dependence on some specific feature value, as it is demonstrated with the TCP window size value, can be misleading in conclusions, thus there is a need of building solutions that would be more immune to this kind of tendencies. We believe that the use of unsupervised machine learning algorithms could contribute more positively to this issue.

Our main conclusions about differences between the entropy-based and machine learning approach in anomaly detection are summarised in Table III.

TABLE III.
DIFFERENCES BETWEEN THE ENTROPY-BASED AND MACHINE LEARNING ALGORITHMS APPROACH IN ANOMALY DETECTION

Characteristics	Entropy-based	Machine Learning
Features	Basic flow data exported by NetFlow like protocols	Additional features calculated by preprocessing network traffic
Processing	Aggregation per epoch	Based on each data instance
	Intensive, but straightforward	Complex, depends on the algorithm
Results	Simplified	Detailed
	One value per feature for all instances in epoch	Per instance (classification, clustering)
	Time series values per feature	Rich performance metrics
Analysis	Data correlation	Straightforward, detailed
	Root cause analysis	Performance metrics
Implementation	Relatively easy	Problematic
	NetFlow data collection	Supervised ML requires labelled dataset
	Straightforward calculation	Unsupervised ML is complex for large number of instances
	Demanding analysis to reduce false positive alarms	Overfitting

A promising solution that can combine the advantages of both approaches is to use unsupervised machine learning algorithms to perform multivariate analysis of entropy results of a large set of features. Our further research is conducted in that direction.

ACKNOWLEDGEMENT

This work was partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia under the EUREKA project "Network Traffic Anomaly Detection system based on NetFlow data analysis – TRADE" [grant number E! 13304].

REFERENCES

[1] N. Moustafa, J. Hu, J. Slay, "A holistic review of network anomaly detection systems: a comprehensive survey," Journal of Network and Computer Applications 128, 2019, pp. 33-55. doi: 10.1016/j.jnca.2018.12.006.

[2] M. Ahmed, A.N. Mahmood, J. Hu, "A survey of network anomaly detection techniques," Journal of Network and Computer Applications 60, 2016, pp. 19–31. doi: 10.1016/j.jnca.2015.11.016.

[3] J. Mirkovic, and P. Reiher, "A taxonomy of DDoS attack and DDoS defense mechanisms," ACM SIGCOMM Computer Communication Review, 34(2), 2014, pp. 39-53.

[4] F. Giannoni, M. Mancini, and F. Marinelli, "Anomaly detection models for IoT time series data," 2018. arXiv preprint arXiv:1812.00890.

[5] B. Li, J. Springer, G. Bebis, and M. Hadi Gunes, "A survey of network flow applications," J. Netw. Comput.Appl., vol. 36, no. 2, 2013, pp. 567–581.

[6] B. Claise, "Cisco systems netflow services export version 9," No. RFC 3954. 2004.

[7] V. Carela-Español, P. Barlet-Ros, A. Cabellos-Aparicio, and J. Solé-Pareta, "Analysis of the impact of sampling on NetFlow traffic classification," Comput. Netw., vol. 55, no. 5, 2011, pp. 1083–1099.

[8] K. Flanagan, E. Fallon, P. Connolly, and A. Awad, "Network anomaly detection in time series using distance based outlier detection with cluster density analysis," In 2017 Internet Technologies and Applications (ITA), 2017, pp. 116-121. IEEE.

[9] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, H. Zhang, "An Empirical Evaluation of Entropy-based Traffic Anomaly Detection," In Proc. of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC'08), 2008, pp. 151–156.

[10] P. Berezinski, B. Jasiul, M. Szpyrka, "An entropy-based network anomaly detection method," Entropy 17 (4), 2015, pp. 2367–2408. doi: doi.org/10.3390/e17042367.

[11] I. Özçelik, R. R. Brooks, "Deceiving entropy based DoS detection," Computers & Security 48, 2015, pp. 234-245, doi: 10.1016/j.cose.2014.10.013.

[12] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*, 2019, O'Reilly Media.

[13] I. Sharafaldin, A.H. Lashkari, A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in: Proceedings of the 4th International Conference on Information Systems Security and Privacy - Volume 1 2018, pp. 108-116. doi: 10.5220/0006639801080116

[14] A. J. Lawrance, and P. A. W. Lewis. "An exponential moving average sequence and point process (EMA1)," Journal of Applied Probability 14, no. 1, 1977, pp. 98-113.

[15] C. Tsallis, "Possible generalization of Boltzmann–Gibbs statistics, Journal of Statistical Physics 52 (1-2), 1988, pp. 479–487, doi: 10.1007/BF01016429.

[16] A. Rényi, "On measures of entropy and information, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, (1961) 547–561.

[17] B. Tellenbach, M. Burkhart, D. Schatzmann, D. Gugelmanna, D. Sornette, "Accurate network anomaly classification with generalized entropy metrics, Computer Networks 55 (11), 2011 pp. 3485-3502, doi: 10.1016/j.comnet.2011.07.00.