proceedings

Extended Abstract

Comparative Results with Unsupervised Techniques in Cyber Attack Novelty Detection †

Jorge Meira

Computing Department, University of Coruña, Coruña 15071, Spain; j.a.meira@udc.es; Tel.: +34-981-167-000

† Presented at the XoveTIC Congress, A Coruña, Spain, 27–28 September 2018.

Published: 17 September 2018

Abstract: Intrusion detection is a major necessity in current times. Computer systems are constantly being victims of malicious attacks. These attacks keep on exploring new technics that are undetected by current Intrusion Detection Systems (IDS), because most IDS focus on detecting signatures of previously known attacks. This work explores some unsupervised learning algorithms that have the potential of identifying previously unknown attacks, by performing outlier detection. The algorithms explored are one class based: the Autoencoder Neural Network, K-Means, Nearest Neighbor and Isolation Forest. There algorithms were used to analyze two publicly available datasets, the NSL-KDD and ISCX, and compare the results obtained from each algorithm to perceive their performance in novelty detection.

Keywords: unsupervised learning; anomaly detection; outlier detection; novelty detection

1. Introduction

Cyber-Security is a field that is constantly evolving, the rate by which new threats and attacks appear is enormous and this requires a constant research for vulnerabilities and ways of solving them by the people responsible for the Security Systems [1].

Intrusion Detection Systems (IDS) are tools based on attack detection techniques to finding out new vulnerabilities. IDS tend to follow one of two different approaches: signature based, or anomaly based. Signature based detection requires previous knowledge of an attack before being able to identify it, on the other hand anomaly base detection only requires knowledge of regular data, and any potential deviation from that norm can correspond to an attack, even if the attack has not been discovered yet [2]. This is an arduous task, and classification algorithms can be used to aid in this scenario. Some algorithms, called supervised learning algorithms are well suited for problems where exiting classified examples can be used as training data for the algorithm. However, with new vulnerabilities there are no classified examples for supervised algorithms to learn from. One possibility to help with this problem is the usage of unsupervised learning algorithms. Unsupervised learning algorithms can learn what is normal data and find deviations from that, which in this case would indicate a possible attack previously unknown (anomaly based).

2. Experimental Work

In this work, was studied the behavior of some unsupervised algorithms based in one class classification, to verify if these techniques are a viable solution to discover and detect unknown attacks. In this section is presented and described the network anomaly detection workflow as shown in Figure 1.

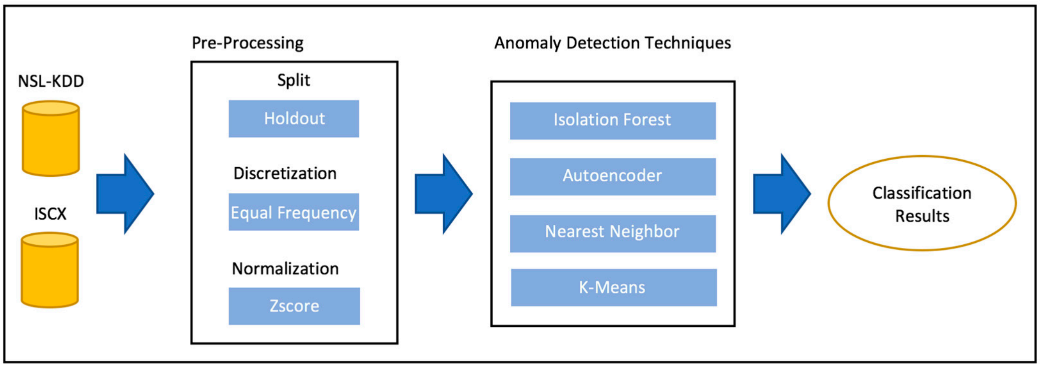


Figure 1. Anomaly detection workflow.

Datasets and Preprocessing

In our exploration, we analyzed the NSL-KDD [3,4] and the ISCX datasets [5]. These datasets contain samples from normal activity and from simulated attacks in computer systems and are commonly used in the literature. Before using the learning algorithms, we've applied some pre-processing methods to prepare the data. As shown in Figure 1, it was first applied the holdout method to both datasets, where 2/3 of the data (corresponding to normal activity in network) were used to train the algorithms and 1/3 of the data where 10% of this portion corresponds to anomalies, were used to test the algorithms. The next step was discretization, where all continuous features of both datasets were converted to categorical features trough the equal frequency technique. The last pre-processing method applied was the data normalization, to have all the features within the same scale. This way, prevents some classification algorithms to give more importance at features with large numeric values. Z-Score was the normalization technique applied to the data. This technique transforms the input, so the mean is zero and the standard deviation is one. After the data cleaning and transformation, we applied four one-class algorithms, namely Autoencoder, Nearest Neighbor, K-Means and Isolation Forest and evaluate is performance in the NSL-KDD and ISCX datasets.

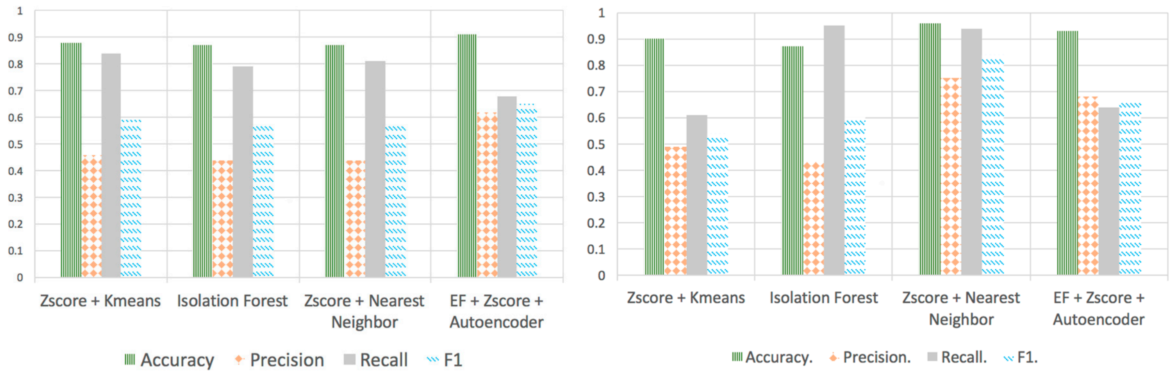
3. Comparative Evaluation and Conclusions

We tested all combinations of pre-processing techniques with the unsupervised learning algorithms and graphically presented the results of the best techniques applied to each algorithm for NSL-KDD and ISCX datasets.

In both datasets all the algorithms had a high accuracy. That was expected as most of the samples are from normal activity. To achieve better conclusions about the algorithms efficacy the F1 metric will also be analyzed. Starting by the NSL-KDD dataset shown in Figure 2a, we can see that all of the algorithms had a similar result close to 60% of F1. It is not a high-performance score, however the recall was much higher than precision in K-Means, Nearest Neighbor and Isolation forest algorithms, around 80%, which means that the false negatives were much less than the false positives. In cybersecurity it is important to have a low false negative rate, since it represents data predicted as normal, while in fact it represents malicious or abnormal activity.

In the ISCX dataset (Figure 2b) all algorithms showed a slightly better performance, except the Nearest Neighbor algorithm which had a much higher score compared to the NSL-KDD. These results were expected in the ISCX, whereas this dataset only has 4 different types of attacks compared to the 38 different types in the NSL-KDD dataset.

The results showed that these unsupervised techniques combined with the best preprocessing techniques could detect most of the anomaly instances but also generate a lot of false positives. These occur due to the similarity between normal and abnormal instances.



(**a**) (**b**)

Figure 2. Anomaly detection results in (**a**) NSL-KDD and (**b**) ISCX datasets.

Author Contributions: His work was conceived and performed by J.M. and guided by Amparo Alonso Betanzos, Veronica Bolón Canedo and Goreti Marreiros, the colleagues from the research group GECAD and LIDIA helped in the reviewing process of this work.

Acknowledgments: This is an extension of the work made in the SASSI project (ANI|P2020 17775). This work is supported by the research project ED431G/01, and has received funding from the Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2016-2019) and the European Union (European Regional Development Fund-ERDF).

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Zanero, S.; Serazzi, G. Unsupervised learning algorithms for intrusion detection. In Proceedings of the 2008 IEEE Network Operations and Management Symposium, Salvador, Bahia, Brazil, 7–11 April 2008; pp. 1043–1048.
2. Casas, P.; Mazel, J.; Owezarski, P. Unsupervised Network Intrusion Detection Systems: Detecting the Unknown without Knowledge. *Comput. Commun.* **2012**, *35*, 772–783.
3. Noto, K.; Brodley, C.; Slonim, D. FRaC: A feature-modeling approach for semi-supervised and unsupervised anomaly detection. *Data Min. Knowl. Discov.* **2012**, *25*, 109–133.
4. *KDD Cup 1999 Data*; UCI Machine Learning Repository: Irvine, CA, USA, 1999. Available online: http://archive.ics.uci.edu/ml/datasets/kdd+cup+1999+data (accessed on 1 January 2018).
5. Shiravi, A.; Shiravi, H.; Tavallaee, M.; Ghorbani, A.A. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Comput. Secur.* **2012**, *31*, 357–374.