Network Security: Exploring the Use of Autoencoders and Variational Autoencoders for Anomaly Detection in Cyber Security

Project Leader: George Deskas

Project Helper: Stephen O'Sullivan, Dimitrios Ntakoulas

Background: Anomaly detection is the process of identifying data points, events, or observations that do not conform to an expected pattern or behaviour [1]. In the context of **Banking Cyber Security,** anomaly detection is a crucial task as it allows for the identification of unusual and potentially malicious activity. Within Lloyds Banking Group, **Security Operations Centres (SOCs)** undertake the correlation and analysis of security events to protect the organisation from external cyber-attacks. Predominantly, these data are activity logs sourced from many heterogenous systems using multiple and evolving communication protocols (e.g., TCP, DNS, HTTP, properties of transferred files, etc.).

Traditional anomaly detection techniques such as rule-based systems and statistical methods have been applied to cyber security data with varying degrees of success. However, these methods can be limited in their ability to adapt to changing patterns of behaviour and may produce a high number of false positives [2]. Deep Learning models, such as **Autoencoders (AEs)** [3] and **Variational Autoencoders (VAEs)** [4], have recently gained attention as a potential solution to the anomaly detection problem. These models are able to learn the underlying distribution of the data and can reconstruct the original inputs that conform to this distribution. When presented with an anomalous sample, the model will be less likely to reconstruct it, making it possible to detect anomalies.

The project will focus on the implementation of AEs / VAEs on real data, their performance in comparison to other anomaly detection approaches, and the interpretability of these models.

Methodology: The project will involve the use of the **UNSW-NB15 Network Intrusion** Dataset [5] [6] for training the AEs / VAEs models and evaluate their performance using standard metrics such as precision, recall, and F1-score. The project will also investigate the interpretability of the models, such as the ability to understand which features are important for anomaly detection. We recommend using one of the following explainability frameworks, **LIME - Local Interpretable Model-Agnostic Explanations** [7] or **SHAP – Shapley Additive exPlanations** [8] for the interpretation of the results.

Research questions:

- How do AEs / VAEs compare to other anomaly detection approaches in terms of performance and interpretability when applied to real-world cyber security data?
- How can these models be implemented for anomaly detection in the context of cyber security?
- What are the exploitability concerns of using AEs / VAEs for anomaly detection in cyber security?

References

[1] [Online]. Available: https://cumulocity.com/guides/machine-learning/anomaly-detection/.

[2] R. H. J. B. S. R.-N. Giulia Moschini, “Anomaly and Fraud Detection in Credit Card Transactions Using the ARIMA Model”, [Online]. Available: https://arxiv.org/abs/2009.07578.

[3] I. G. a. Y. B. a. A. Courville, “Deep Learning”, [Online]. Available: http://www.deeplearningbook.org.

[4] M. W. Diederik P Kingma, “Auto-Encoding Variational Bayes”, [Online]. Available: https://arxiv.org/abs/1312.6114.

[5] [Online]. Available: https://research.unsw.edu.au/projects/unsw-nb15-dataset.

[6] Z. Z. a. G. Serpen, “UNSW-NB15 Computer Security Dataset: Analysis through”, [Online]. Available: https://arxiv.org/ftp/arxiv/papers/2101/2101.05067.pdf.

[7] [Online]. Available: https://github.com/marcotcr/lime.

[8] [Online]. Available: https://github.com/slundberg/shap.