2. Nusriddinovich, I. N., & Ilhomkhojayevna, A. N. (2022). CYBER THREATS,VULNERABILITIES AND RISKS IN ECONOMIC SECTORS. *Galaxy International Interdisciplinary Research Journal*, *10*(9), 139-140.

3. Маллабоев Н., Шокиров Д. СПОСОБЫ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ //Теория и практика современной науки. – 2016. – №. 6-1. – С. 826-830

4. MU Xujaevich. Pedagogical conditions for the formation of communicative competencies of students of vocational schools using interactive teaching methods // International Journal on Integrated Education 3 (12), 345-347.

DATASETS FOR INTRUSION DETECTION SYSTEMS: ENHANCING NETWORK SECURITY

Bozorov Suhrobjon
Toshkent axborot texnologiyalari universiteti
bek.muminovich.95@mail.ru

Abstract. Intrusion Detection Systems (IDS) play a pivotal role in safeguarding networks against cyber threats. To effectively develop and evaluate IDS solutions, access to diverse and comprehensive datasets is crucial. This article explores the importance of datasets for intrusion detection, highlights key requirements for such datasets, and discusses notable datasets and their features commonly used in the field. By understanding the value of high-quality data, researchers and cybersecurity professionals can better address the evolving landscape of network attacks and fortify their defense mechanisms.

Keywords: Intrusion Detection Systems, Network Security, Datasets, Cyber Threats, Machine Learning.

Introduction

With the proliferation of network-connected devices and the increasing complexity of cyber threats, the importance of effective Intrusion Detection Systems (IDS) cannot be overstated. IDS serve as the first line of defense in identifying and mitigating unauthorized access, malicious activities, and network anomalies. To develop robust IDS solutions and assess their performance, researchers and practitioners rely heavily on datasets that accurately represent real-world network traffic and attacks.

This article delves into the significance of datasets in the realm of intrusion detection, outlining essential requirements for datasets, discussing prominent datasets available for research and analysis, and highlighting their distinctive features.

Requirements for Datasets:

To be effective, datasets for intrusion detection must meet specific criteria:

Realistic Representation: Datasets should emulate real-world network traffic, encompassing various protocols, traffic volumes, and patterns. This realism is essential to train and test IDS models effectively [5].

Labeling and Ground Truth: Annotated labels identifying normal and anomalous network traffic are crucial. Ground truth data helps evaluate the accuracy of IDS solutions and measure their performance.

Diversity of Attacks: Datasets should cover a wide spectrum of attack types, including known and unknown threats. This diversity ensures that IDS can detect both common and emerging threats.

Scalability: Scalable datasets allow researchers to study network behavior at different scales, from small-scale local area networks to large-scale enterprise networks.

Privacy and Ethics: Adherence to privacy and ethical considerations is vital when creating or using datasets. Anonymization techniques should be employed to protect sensitive information [3].

Datasets and Their Features to Detect Network Attacks

Several datasets are widely used in the field of intrusion detection, each offering unique features and challenges. Some notable datasets include:

KDD Cup 1999: Derived from the 1998 DARPA Intrusion Detection Evaluation Program, this dataset is a benchmark for evaluating IDS. It contains a vast number of features and covers various attack scenarios [1].

NSL-KDD: An improved version of the KDD Cup 1999 dataset, NSL-KDD addresses some of its limitations, including redundancy and lack of representation of modern network attacks [2].

CICIDS2017: Focused on modern cyber threats, this dataset provides a comprehensive collection of benign and malicious traffic, including Distributed Denial of Service (DDoS) attacks and botnet activity [4].

UNSW-NB15: This dataset, designed for machine learning-based intrusion detection, comprises diverse network traffic, spanning multiple attack categories and protocols.

CTU-13: A dataset created from real network traffic captures, CTU-13 includes a range of contemporary attacks, such as botnet, DoS, and malware-related activities.

AWID: Focusing on Wireless Intrusion Detection Systems (WIDS), the AWID dataset includes both benign and attack traffic in Wi-Fi networks, making it valuable for WIDS research [5].

Table 1. Comparison of datasets by basic criteria's

Dataset Name	Source	Purpose	Diversity of Attacks	Anomaly Detection	Scalability	Year of Release
KDD Cup 1999	DARPA Intrusion Detection	Benchmark	Various attack types	Yes	Yes	1999

NSL-KDD	Improved KDD Cup Dataset	Benchmark (Improved)	Various attack types	Yes	Yes	2009
CICIDS2017	Created for Modern Threats	Contemporary Threats	DDoS, botnets, etc.	Yes	Yes	2017
UNSW-NB15	Machine Learning-Focused	Comprehensive	Various attack types	Yes	Yes	2015
CTU-13	Real Network Traffic	Modern Threats	Botnet, DDoS, etc.	Yes	Yes	2013
AWID	Wireless Networks	Wi-Fi Intrusion Detection	Wi-Fi attacks	Yes	Yes	2014

Conclusion

Intrusion Detection Systems are indispensable in safeguarding network security, and their effectiveness is heavily reliant on the quality of the datasets used for development and evaluation. By adhering to specific requirements and utilizing diverse and representative datasets, researchers and cybersecurity professionals can enhance the accuracy and robustness of their IDS solutions. As cyber threats continue to evolve, staying abreast of the latest datasets and their features is paramount in ensuring the resilience of network security systems.

References:

1. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (pp. 1-6).

2. Moustafa, N., & Slay, J. (2015). The significant features of the NSL-KDD dataset. In Proceedings of the 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (pp. 25-31).

3. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy (pp. 108-116).

4. Dainotti, A., Bernaschi, M., & Pescapé, A. (2014). Analysis of an anomalous Internet-wide TCP state-transition attack. In Proceedings of the 2014 ACM Conference on SIGCOMM (pp. 427-428).

5. Ring, M., Wunderlich, S., & Scheitle, Q. (2017). Evaluation of network-based intrusion detection system datasets. In Proceedings of the 2017 Network Traffic Measurement and Analysis Conference (TMA) (pp. 33-40).