Network Intrusion Detection

1st Amer Mohamed
Autonomous Systems A
Hochschule Hamm-Lippstadt
Lippstadt, Deutschland
mohamed-ahmed-mohamed-ali.amer@stud.hshl.de

***Abstract*—**

I. Introduction

Networks safety is becoming a crucial topic in all types of enterprises and organizations. The importance of Network security and data confidentiality are growing rapidly due to the rapid evolution in technology and AI. One of the approaches taken to ensure network safety are network intrusion detection techniques [1]. Many techniques have been introduced through the evolution of technology. The goal of intrusion detection systems is to be able to detect unusual activities, unauthorized personnel, or misuse from insiders and external penetrators in a network, preferably in real-time. [2]. One of the emerging new technologies for network security is the use of machine learning and deep learning models to identify network attacks and intruders in a network. The abundance of big data has led to the ability to train machine learning and deep learning models efficiently. More research is being done on making those models extremely accurate and efficient for large deployment. [3]

II. Background

The main two types of Network detection system are Deployment method based and Detection method based. From the detection method perspective, it is further divided into two more categories "Signature-based intrusion detection (SIDS)" and "Anomaly detection-based intrusion detection (AIDS)".

The SIDS is based on the idea of defining a specific unique signature for network attacks. Those signatures are stored in a database. The system from there matches those signatures with the activity in the network and detects if there is a probable attack on the service. This type of approach lacks the ability to detect new types of attack as it lacks its signature and requires a huge carefully selected database which increases the computing resources needed for this algorithm [3].

The AIDS approach, also called the "behavior-based IDS," is based on the idea of defining a clear profile of normal users. Any deviation from this normal profile will be considered as an anomaly [4].The biggest advantages of using the AIDS approach are its ability to detect novel and new types of attacks. Though the only drawback of using this approach is the hard nature of classifying the difference between a normal and an abnormal profiles specially with the rising popularity of different IOT devices [5].In this paper, we aim to explore Random Forest machine learning approach along with data preprocessing techniques to improve the accuracy of network intrusion detection systems. The goal is to identify the most effective model parameters for detecting malicious activities within network traffic.

Identify applicable funding agency here. If none, delete this.

III. Dataset

The dataset used in this study is the KDD Cup 1999 Intrusion Detection Dataset, which was created by simulating attacks on a U.S. Air Force LAN to capture raw TCP/IP traffic. It contains 41 features (3 qualitative and 38 quantitative features) per connection and the target variable named class is labeled as normal and anomalous behaviour [6]. The features can be grouped by type into 4 categories.

A. Features

- **Basic Features of Individual TCP Connections:** These features are extracted from the basic TCP connection. Many attacks can be determined just by analysing how the connection behaves [6], [7].
- **Content Features within a Connection:** These are features that inspect the payload or command content of a connection to detect any suspicious behaviour. They go beyond the basic properties of a connection and look deeper into what is actually being transmitted during the session. [6], [7]
- **Time based Traffic features:** These features consider connections to the same host in the past two seconds. [6], [7]
- **Host-based Traffic Features:** It is similar to Time based traffic features but these use a larger time window to detect patterns in connections. [6], [7]

B. Types of attacks

- **Denial of Service Attack (DOS):** It is a type of attack where the attacker overloads computing and memory resource. That makes the service too busy to fully handle legitimate requests and denies legitimate users access to the machine [7], [8]

- **User to Root Attack (U2R):** This type of attack occurs when the attacker starts out with access to a normal legitimate account on the system and then becoming able to exploit vulnerability to gain root access to the system [7], [8]
- **Remote to Local Attack (R2L):** It occurs when an attacker can send packets to a machine over a network where the attacker does not have access as a user to that machine [7], [8]
- **Probing Attack:** It is an attempt to gather information about a network of computers for the purpose of breaching through their security and gaining root access [7], [8]

C. Potential Issues in the Dataset

Previous research shows that this specific dataset has some issues. They experimented with various machine learning models all of which showed a very high accuracy of approximately 98% which is a high accuracy for machine learning models and often means a problem with the model and it will not translate into real life deployment. The first important deficiency in the KDD data set is the huge number of redundant records. Analyzing KDD train and test sets, it was found that about 78% and 75% of the records are duplicated in the train and test set respectively. This will cause the model to be biased towards the more frequent records and prevent it from learning the novel records. [8].

D. Dataset Preprocessing

E. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

F. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

G. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	*Subhead*	*Subhead*
copy	More table copy[a]		

[a]Sample of a Table footnote.

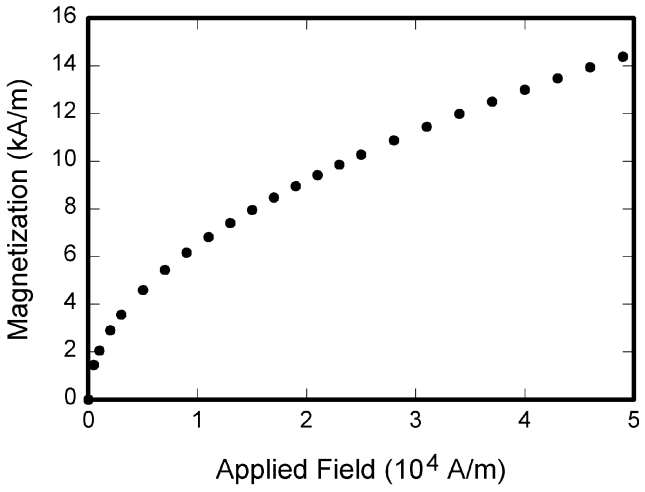

Fig. 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

References

[1] S. Kumar, Survey of current network intrusion detection techniques, Washington Univ. in St. Louis, pp. 1–18, 2007.

[2] B. Mukherjee, L. T. Heberlein and K. N. Levitt, "Network intrusion detection," in IEEE Network, vol. 8, no. 3, pp. 26-41, May-June 1994, doi: 10.1109/65.283931. keywords: Intrusion detection;Computer networks;Protection;Computer security;Data security;Computer science;Computer crime;Information security;Real time systems;Prototypes,

[3] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," Trans. Emerging Telecommun. Technol., vol. 32, no. 1, e4150, 2021. [Online]. Available: https://doi.org/10.1002/ett.4150

[4] W. Ma, "Analysis of anomaly detection method for Internet of Things based on deep learning," Trans. Emerg. Telecommun. Technol., vol. 31, no. 6, e3893, 2020. [Online]. Available: https://doi.org/10.1002/ett.3893

[5] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-Things-based smart cities: Recent advances and challenges," IEEE Commun. Mag., vol. 55, no. 9, pp. 16–24, Sep. 2017. [Online]. Available: https://doi.org/10.1109/MCOM.2017.1600514

[6] KDD Cup 1999, "KDD Cup 1999 Data," 1999. [Online]. Available: [Accessed: May 13, 2025].

[7] KDD Cup, "KDD Cup 1999 Data," [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/task.html, [Accessed: May 13, 2025].

[8] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, 2009, pp. 1-6, doi: 10.1109/CISDA.2009.5356528. keywords: Testing;Intrusion detection;Data security;Statistical analysis;Computer security;Computer aided manufacturing;Learning systems;Computational intelligence;Computer networks;Application software,