JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Intrusion Detection System Using Machine Learning

[1]Anirudh S, [2]Anagha K T P, [3]Karthik A P, [4]Nivin K S, [5]Daya Naik

[1,2,3]Student, [4]Assistant Professor,[5]Associate professor

[1,2,3,4,5]Artificial Intelligence and Machine Learning,

[1,2,3,4,5] Srinivas Institute of Technology, Mangalore, India

Abstract : This paper introduces a Python and Flask-based Intrusion Detection System (IDS) designed for real-time cybersecurity by analyzing network traffic using machine learning to detect and alert users of potential intrusions. Key features include data collection, processing, and email notifications for alerts, demonstrating high accuracy, low false positives, and efficient scalability. Ethical considerations, such as data privacy, user consent, and legal compliance, are addressed to ensure responsible deployment. The system integrates emerging technologies, including machine learning, cloud computing, blockchain, IoT, and big data analytics, making it applicable across sectors such as finance, healthcare, government, and critical infrastructure. This IDS offers a robust and scalable solution to address evolving cyber threats.

IndexTerms **– Intrusion Detection System, Machine Learning, Network Security, Anomly Detection,Denial-Of-Service(DOS),**

I. INTRODUCTION

In the current digital environment, protecting essential systems and confidential information from the growing sophistication of cyber threats is crucial for individuals as well as companies. The ongoing advancement of threats requires the adoption of strong cybersecurity measures. An Intrusion Detection System (IDS) acts as a crucial instrument in this context, efficiently overseeing system operations and network traffic to detect and address security threats. The presentation concentrates on developing an IDS with Flask and Python, utilizing artificial intelligence to improve its functionalities. Flask, a minimalistic Python web framework, allows for the creation of web applications and APIs, offering a versatile and effective groundwork for the IDS. The wide range of libraries and tools available in Python for data analysis and machine learning makes it a perfect option for developing advanced intrusion detection systems.

This IDS deployment employs AI to assess network traffic in real-time, identifying unusual behaviors that may signal possible security threats. The system employs a variety of machine learning techniques to analyze network patterns and detect anomalies, improving its capability to spot emerging threats. Upon detecting an intrusion, the IDS quickly notifies users through email, enabling them to act swiftly to reduce the threat. This immediate alert system guarantees that security issues are handled promptly, minimizing the chances of major harm. In conclusion, utilizing Flask and Python to implement an IDS, enhanced by artificial intelligence, provides an effective means for overseeing and protecting networks. Byutilizing AI, the system can quickly identify and react to threats, giving users prompt notifications and allowing them to safeguard their vital assets efficiently.

II. METHODOLOGY

The system that enables mathematical calculations through real-time hand gestures is designed to be simple and efficient. By combining gesture recognition with computational tools, it ensures smooth and effective interaction, even in environments with limited resources.The process involves key steps and components that work together to allow users to easily perform calculations just by using hand movements, making it both intuitive and practical. In the following sections, the methodology will be outlined to develop a robust Network Intrusion Detection System (NIDS) using machine learning methodologies that could eventually be integrated into an available web application for use by the users. A description of each step is provided

A. DATA COLLECTION:

This system is based on the NSL-KDD dataset which has been an enhanced version of the original KDD'99 dataset. The NSL-KDD dataset overcomes the problems of redundancy and representation as found in the original dataset. So, this has improved the credibility of the benchmark for the intrusion detection research. The dataset represents a number of network intrusions in the controlled military network. A realistic and diverse context was provided to train and test intrusion detection models. It is divided into two types: training and testing subsets to train and test the models upon different data, allowing them to be evaluated properly on performance.

B. DATA PREPROCESSING

Good preprocessing ensures the fact that the data has been of good quality and pertinent for machine learning models. The following steps are as follows:

1. **Feature Selection**:
   This involves the selection of major features that are significant in terms of predicting network intrusions. Methods for example, correlation analysis, mutual information, and feature importance metrics such as Random Forest models, select those that have high predictive powers from the dataset. Eliminating unwanted and redundant features from the dataset decreases dimensionality, thus enhancing the performance and interpretability of the model

2. **Normalization:**
   Scaling to the same range is done in achieving uniformity by using methods like Min-Max Scaling and Z-Score Standardization. Normalization reduces the impact of features with wide numerical ranges on model predictions, hence speeding up learning processes and improving convergence of machine learning algorithms.

3. **Encoding:**
   Since protocol types and service types are categorical features, they will be encoded using one-hot encoding, where all categories will be converted to numerical representations equitably. This prevents the occurrence of ordinal relationships and tends to improve the accuracy of the machine learning models that work with categorical data.

C. MODEL TRAINING AND EVALUATION

There are various machine learning algorithms that have been applied in training the IDS, and all of them were developed for the exploitation of some special strength in intrusion pattern detection. Here is one:

1. **Decision Trees**: The model relies upon simple and interpretable criteria for partitioning the data into subsets on the basis of feature attributes. The model, hence, becomes understandable and can be visualized as well.

2. **Random Forests**: This ensemble method uses multiple decisions to improve precision, cut down overfitting, and give more reliable predictions instead of using one decision tree.

3. **SVM**: Efficient in high-dimensional spaces, it determines the best hyperplane that separates classes for accurate classification.

4. **KNN**: It is a non-parametric approach in which the data points are classified based on closeness to labeled data. This is suitable for simplicity and effectiveness in specific contexts.

5. **Neural Networks**: Since neural networks can model complex, nonlinear relationships, they are applied to detection of sophisticated patterns in the dataset.

All the models were trained with the training subset and then tested using performance metrics like accuracy, precision, recall, and F1 score. It gives a general overview regardingthe proper classification of types of intrusion with the number of false positives and negatives. The IDS system isimplemented using a web interface that has been built withFlask-a lightweight Python web framework which ensures accessibility and usability:

1. **Home Page**:
   This is a general overview of the project, its objectives, and the importance of intrusion detection,making users understand why the system is needed and how it works.

2. **IDS Page:**
   This is an interactive page where a user would input specific network features for analysis by the IDS. The form is interactive to make sure all relevant data points are entered appropriately

3. **Prediction Page:**
   This indicates the outcome of intrusion detection,such as alerting when an intrusion has been detected.It provides more information about the threat that has been detected to the user

4. **Email Notification System:**
   The alert mechanism is implemented so that security teams are timely alerted in case intrusion has been detected. How the system works is given below:

   - **Email Configuration**: The application configures the email account for authentication purposes and secure communication with the SMTP server.

- **Email Generation**: An email message is composed with a descriptive HTML table summarizing the network features, their values, and the type of intrusion detected.
- **Sending Email**: This email is sent securely via SMTP protocol to alert concerned teams about potential threats,enabling prompt action.

D. Deployment:

The local Flask framework is used for implementing the IDS application. The server reads in the user queries, analyzes the data, looks for possible intrusions, and sends out email notifications as appropriate. Users can interact with the system by accessing the local server on a web browser (http://127.0.0.1:5000/). This setup makes it very accessible and allows for immediate feedback about the network's security status.

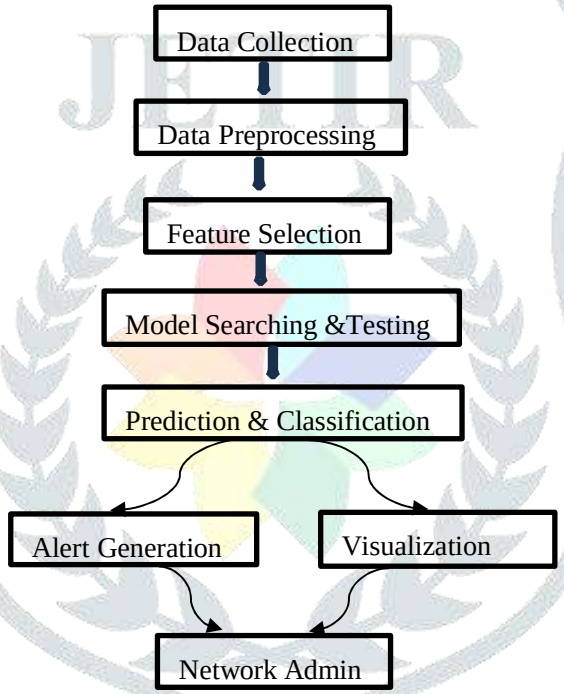

Fig.1. system architecture of intrusion detection system using ml

III. PERFORMANCE

This section evaluates the proposed Intrusion Detection System (IDS) based on its ability to classify network traffic with minimal errors. System performance is assessed using several key metrics. The first metric, accuracy, measures the correct classification of instances relative to the total number, though it can be misleading for imbalanced datasets.

Precision, defined as the ratio of true positive predictions to all predicted positives, is crucial for reducing false positives and avoiding unnecessary alerts. Recall, or sensitivity, measures the ratio of true positive predictions to all actual positives, ensuring most intrusions are correctly identified,thereby minimizing missed attacks. The harmonic mean of precision and recall, known as the F1-Score, provides a balanced measure of the model's effectiveness in detecting actual intrusions while minimizing false alarms.Additionally, a confusion matrix offers a detailed overview of true positives, false positives, true negatives, and false negatives, aiding in model refinement.

The machine learning models trained and tested include Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANNs). Decision Trees are simple and interpretable but are prone to overfitting, especially with complex datasets. Random Forests, comprising an ensemble of decision trees, exhibit robust performance in terms of precision and recall while reducing overfitting. SVMs perform well in high-dimensional spaces, providing robust precision and recall, particularly for linearly separable classes. KNN, though efficient in certain scenarios, relies heavily on the selection of neighbors (k) and is less effective for large datasets and high-dimensional data. Neural Networks achieve high performance by modeling complex, non-linear relationships but require substantial computing power and careful tuning of hyperparameters. Comparing the performance of these models using the aforementioned metrics determines the best performing algorithm.

IV. INTEGRATION WITH EMERGING TECHNOLOGIES

Improvement is a continuous process, and this principle applies to the development of intrusion management systems by integrating contemporary technologies. These technologies provide additional layers of security, refine the detection methods, and smooth out processes.

1. **Machine Learning and Artificial Intelligence**: Machine learning and artificial intelligence can improve the detection of complex and evolving intrusion patterns. With large datasets, the IDS can learn and identify potential intrusions and subtle anomalies. Analytical capabilities are also further improved with deep learning and neural networks

2. **Cloud Computing**: Cloud computing provides the scalability and flexibility required by IDS. Cloud services offer the necessary storage and computational power to handle large volumes of data. Moreover, deploying the IDS on the cloud facilitates easy updates and maintenance, ensuring the system remains current with emerging security threats.

3. **Blockchain Technology**: Blockchain technologies ensure the integrity and immutability of data collected by the IDS. Data tampering becomes virtually impossible because logs and alerts are recorded on the blockchain, which enhances the trustworthiness and transparency of the system's operations.

4. **Internet of Things (IoT):** It can integrate with the IoT networks, which could offer wide-ranging security protection due to the large number of devices connected. IDS can track traffic and activities conducted by IoT devices to identify possible threats that might affect such endpoints, thus securing the whole network, including the IoT devices.

5. **Big Data**: Big data analytics applications allow for processing and analyzing massive streams of real-time security data. This advanced analytics technique could determine whether there is an ongoing potential intrusion, by alerting the IDS beforehand for proactive threat detection and response.

6. **Automation and orchestration**: Tools can aid in automation and orchestration frameworks to make response efficiency to detected intrusions possible. Automated workflows will enable predefined actions in cases of intrusion, such as moving affected systems to a 'quarantine zone, notifying administrators, and kicking in remediation processes. In response, this approach takes out less response time from systems. This also minimizes the intruded aspects.

These technological integrations significantly enhance the capabilities of intrusion management systems, more robust security measures, and efficient handling of potential threats

V. ETHICS

The deployment and use of an intrusion detection system (IDS) must adhere to established ethical guidelines in order to ensure responsible usage and to protect the rights and privacy of individuals. Here are the key considerations:

A. Privacy:

An IDS gathers data that may include sensitive information about the users. Therefore, data anonymization should be performed whenever possible, and access to data should be restricted to authorized personnel only. This ensures that the privacy of the individuals is maintained, and their sensitive information is not exposed to unauthorized parties.

B. Consent:

Users should be notified of any monitoring and data collection activities by the IDS. Explicit consent must be obtained prior to gathering their data, ensuring transparency of the policies on data usage. Transparency helps in gaining the trust of the users and also in meeting the legal requirements

C. Data Security:

The data gathered by the IDS has to be secured. This could include measures like encryption, secure storage, and routine security audits to ensure that unauthorized access and breaches of the data are prevented. This ensures that the integrity and confidentiality of the data collected are ensured, which is not compromised.

D. Bias and Fairness:

Machine learning models in IDS should be checked for possible biases. This ensures that the models do not target or harm certain groups or individuals unfairly. Audits and updates of such models are necessary to mitigate any bias in the detection process to ensure fairness and equity.

E. Accountability:

Accountability measures should be put in place with detailed specifications for addressing any problems or incidents that may arise from the IDS. Roles and responsibilities for the monitoring of the system, response to alerts, and data breach management should be well defined. This ensures problems are identified and solved efficiently.

The IDS should respect current legal frameworks and regulations, including data protection laws—like the European General Data Protection Regulation—and those on cybersecurity. It would be proper to consult legal counsel to ensure that the IDS is compliant with all applicable legal requirements and functions within the boundaries set by law

By adhering to these ethical guidelines, the implementation and use of an IDS can be both effective and responsible in the protection of user privacy, ensuring fairness, security, and accountability.

VI. CHALLENGES

Some of the important challenges that have to be met in the development and deployment of an intrusion detection system (IDS) include ensuring the effectiveness and reliability of the system. Key among these challenges are:

A. Data Quality and Availability:
The quality and availability of data greatly affect the effectiveness of the IDS. Poor, noisy, or biased data can significantly deteriorate the accuracy of the detection outcome. Good quality and complete datasets are crucial for training and testing the IDS.

B. Evolving Threat Landscape:
Cyber threats evolve over time, with new attack vectors and techniques being discovered daily. Therefore, an IDS must be continually updated to recognize and respond to new threats. It calls for constant research and development efforts to keep the system current.

C. False Positives and False Negatives:
A good balance has to be found between the detection rate and false positives—normal activities wrongly flagged as attacks—and false negatives—missed intrusions. A high rate of false positives leads to alert fatigue, and false negatives result in undetected intrusions.

D. Performance and Scalability:
The IDS should be capable of processing large volumes of data in real-time without significant delays. However, the performance and scalability of the system must be ensured to handle greater data loads and complex detection algorithms.

E. Integration with Existing Systems:
Integration of the IDS with the prevailing IT infrastructureand security systems can be quite complex, includingcompatibility issues, inconsistent data formats, and multiple platforms and devices

F. Resource Constraints:
The deployment of IDS requires computational resources such as CPU, memory, and storage. The challenge here is to ensure that resources are utilized efficiently while maintaining high accuracy in detection, especially in constrained environments

G. User Awareness and Training:
The effective use of the IDS requires users and administrators to be aware of its capabilities and limitations. Thus, adequate training and documentation are necessary for ensuring the system is used appropriately and that alerts are dealt with accordingly.

H. Ethical and Legal Considerations:
Adhering to ethical guidelines and legal requirements is critical. Ensuring data privacy, obtaining user consent, and complying with regulations such as GDPR can be challenging but are necessary to maintain trust and avoid legal repercussions.

Addressing these challenges requires a multidisciplinary approach, involving expertise in cybersecurity, data science, software engineering, and legal and ethical considerations. By proactively identifying and mitigating these challenges, the IDS can be developed and deployed effectively

VII. APPLICATIONS

The Intrusion Discovery System(IDS) developed in this design has a wide range of operations across colorful disciplines, enhancing cybersecurity and guarding critical means from cyber pitfalls.

In network security, the IDS can be stationed to cover network business and descry implicit intrusions in real- time, relating and mollifying pitfalls similar as denial- of- service(DOS) attacks, probing, and unauthorized access attempts. For enterprise security , associations can use the IDS to cover their internal networks and systems from cyber pitfalls by covering stoner conditioning, system logs, and network business to detect suspicious activity and help data breaches.

The IDS also plays a vital part in critical structureprotection by securing essential services similar as powergrids, water force systems, and transportation networks from cyber pitfalls. In the fiscal sector , fiscal institutions can use the IDS to cover deals, descry fraudulent conditioning, and cover sensitive fiscal data, icing the security and integrity of fiscal systems.

In healthcare , the IDS can cover healthcare systems and patient data from cyber pitfalls by covering electronic health records(EHRs), medical bias, and network business to descry and help unauthorized access and data breaches. Government agencies and defense associations can work the IDS to cover sensitive information and critical systems from cyber spying and attacks, icing public security and guarding classified data.

Educational institution can emplace the IDS to guard their networks, systems, and data from cyber pitfalls, guarding pupil information, exploration data, and executive systems. By using the IDS in these colorful operations, associations can enhance their security posture, descry and respond to pitfalls more effectively, and cover their critical means from evolving cyber attacks.

VIII. FUTURE DIRECTIONS

The development of the Intrusion Detection System(IDS)opens several avenues for unborn exploration andadvancements. Then are crucial areas for enhancement

A. **Advanced Machine Learning ways**: Integrate advanced styles like deep literacy and ensemble ways to enhance delicacy and robustness in detecting complex attack patterns.

B. **Real- Time Data Processing** :Optimize the data processing channel to handle large volumes of data with minimum quiescence, enabling timely discovery and response to intrusions.

C. **Adaptive and tone- Learning Systems**: Develop tone-learning IDS that automatically modernize discovery models grounded on new data and arising pitfalls using nonstop literacy algorithms

D. **Integration with trouble Intelligence**: Connect the IDS with trouble intelligence platforms to gain fresh environment and perceptivity into detected pitfalls, enhancing the system's responsiveness.

E. **Enhanced stoner Interface and Reporting** : Ameliorate the stoner interface and reporting capabilities with dashboards, visualizations, and detailed reports for better understanding and response to security incidents.

F. **Cross-Domain operations**: conform the IDS for new disciplines like smart metropolises, independent vehicles, and critical structure, addressing unique challenges and conditions.

G. **sequestration- Conserving ways**: utensil ways like discriminational sequestration and securemulti-party calculation to cover sensitive data while maintaining effective intrusion discovery.

H. **Collaboration and Information participating**: Develop fabrics for secure and effective sharing of trouble intelligence and discovery models among associations to enhance overall intrusion discovery sweats.

By pursuing these directions, the IDS can continuously ameliorate to attack evolving cyber pitfalls, furnishing further robust and effective security results across colorful disciplines

IX. RESULT

A. Home Page:

Fig 2. home page

The home page of the Intrusion Detection System (IDS) serves as a comprehensive guide to understanding and mitigating network intrusions. It features an intuitive layout with sections such as "What is an Intrusion?" and "Types of Intrusions," offering a clear explanation of unauthorized access and malicious activities that compromise data integrity, confidentiality, and availability.

1. The "Types of Intrusions" section categorizes attacks into: Network-based Attacks: Targeting network infrastructure, including Denial-of-Service (DoS), Man-in-the-Middle (MITM), and unauthorized access attempts.

2. Host-based Attacks: Focusing on individual systems, such as malware infections, ransomware, and data exfiltration

3. Application-level Attacks: Exploiting vulnerabilities within applications, such as SQL injection and cross-site scripting (XSS).

With an easy-to-navigate interface and educational resources, the IDS homepage introduces users to key cybersecurity concepts while emphasizing the system's role in protecting against various threats.

B. Intrusion Detection System Page:

Network Intrusion Detection System Analysis:

The handed web runner appears to be a network intrusion discovery system analysis tool. It presents colorful attack scripts and corresponding criteria to help identify implicit security pitfalls. Some of the crucial criteria include:

- Number of connections to the same destination host
- Chance of connections to different services
- Chance of connections to the same source harborage
- Status of the connection
- Successfully logged in status
- Destination network service used(HTTP)

These criteria can be used to dissect and prognosticate implicit attacks, similar as a Denial of Service(DOS) attack, which is indicated by the" Attack Class Prediction DOS" field.

X. CONCLUSION

This project successfully designed and implemented an Intrusion Detection System (IDS) using Python and Beaker to enhance cybersecurity. It detects potential intrusions in real time, leveraging machine learning algorithms to analyze network traffic and identify suspicious activity.The IDS achieved high detection accuracy with a low false-positive rate, effectively identifying various cyber threats, including denial-of-service (DOS) attacks, probing, remote-to-local (R2L) attacks, and user-to-root (U2R) attacks. Optimized processing times and resource efficiency ensure smooth operation in diverse environments.Scalability tests confirmed its ability to handle increasing data volumes and evolving threats, making it suitable for network security, enterprise security, cloud security, IoT security, critical infrastructure protection, financial sector security, healthcare, government, defense, and education. Ethical considerations such as data privacy, user consent,and legal compliance were also addressed to ensure responsible implementation.By integrating emerging technologies like machine learning, cloud computing, blockchain, IoT, big data analytics, and automation, the system adapts to evolving cyber threats. Future enhancements could explore advanced machine learning techniques, real-time data processing, self-learning mechanisms, threat intelligence integration, improved user interfaces, cross-domain applications, privacy-preserving methods, and collaboration frameworks to further improve performance and security.This IDS offers a robust and scalable solution for real-time intrusion detection, helping organizations safeguard their critical assets from cyber threats.

REFERENCES

[1]Stallings, W., & Brown, L. (2018) Computer Security: Principles and Practice. Pearson.

[2]Scarfone, K., & Mell, P. (2007). Guide to Intrusion Detection and Prevention Systems (IDPS). NIST Special Publication 800-94. National Institute of Standards and Technology.

[3]Sommer, R., & Paxson, V. (2010) Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. IEEE Symposium on Security and Privacy.

[4]Denning, D. E. (1987). An Intrusion-Detection Model. IEEE Transactions on Software Engineering, SE-13(2), 222-232.

[5]Lippmann, R. P., Haines, J. W., Fried, D. J., Korba, J., & Das, K. (2000). The 1999 DARPA Off-Line Intrusion Detection Evaluation. Computer Networks, 34(4), 579-595.

[6]Buczak, A. L., & Guven, E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.

[7]Roesch, M. (1999). Snort - Lightweight Intrusion Detection for Networks. Proceedings of the 13th USENIX Conference on System Administration (LISA '99).

[8]Axelsson, S. (2000). Intrusion Detection Systems: A Survey and Taxonomy. Technical Report 99-15, Department of Computer Engineering, Chalmers University of Technology.

[9]Zuech, R., Khoshgoftaar, T. M., & Wald, R. (2015). Intrusion Detection and Big Heterogeneous Data: A Survey. Journal of Big Data, 2(1), 3.