GEO-Honeypot-IDS: An AI-Based Intrusion Detection System using Distributed Honeypots

[1]Naveen D, [2]Rajesh L, [3]Mohammed Dhakeer N, [4]Mohammed Thaha, [5]Mrs.C.Subalakshmi, [6]Dr. T. Kumanan, [7]Dr. M Nisha.
1,2,3 Students, Department of CSE

5,7 Assistant Professor, Department of CSE 6 Professor, Department of CSE
Dr.M.G.R Educational and Research Institute, Maduravoyal, Chennai 95, Tamil Nādu, India

Abstract

As internet applications, cloud computing and cyber infrastructure continue to expand fast, so has the number and sophistication of cyber-attacks. The traditional cybersecurity tools like firewalls and signature-based Intrusion Detection Systems (IDS) are not effective anymore to detect complex attacks like the zero-day attacks, distributed denial of- service (DDoS), and advanced persistent threats that can be located in various locations and continuously evolve.

In this case we suggest Geo-Honeypot-IDS, which is a smart security system that incorporates AI-based intrusion detection and honeypots located in various geographical areas. Honeypots are virtual system which lures attackers and gives useful information regarding the attacks without compromising on real systems. By placing the honeypots in the various regions we can study the behaviour of the attacks depending on the origin of the attack, the origin of the attack and the rate of attacks. Machine learning is used to process the received information with the purpose of detecting malicious behaviour.

The system offers real time threat information, minimizes false alerts and supports early detection of attack. Geo- Honeypot-IDS presents an answer to modern cybersecurity challenges and uses AI, honeypots, and geographic data to provide a scalable and efficient solution**.** It has an accuracy in detection of malicious activity of 97.00 percent, indicating its reliability.

Keywords - *Artificial Intelligence, Cybersecurity, Distributed security, Honeypot, Intrusion Detection System, and Machine Learning.*

Highlights

1. A Geo-Honeypot-IDS platform that combines distributed honeypot sensors and AI-based intrusion detection.
2. Machine learning algorithms such as the Random Forest, Gradient Boosting and the Logistic Regression to detect intelligent attack.
3. Attack source identification of geographical distribution of cyber threats through Geo-IP.
4. Ensemble learning method which enhanced intrusion detection rate at 97% accuracy.
5. Scalable structure which allows monitoring and analysis of network attacks in real-time.

I INTRODUCTION

The rapid evolution of information technology, cloud computing, and web-based services has made organizations rely more on computer networks and, as a result, this has elevated their vulnerability to data breaches and cyberattacks. Due to this fact, network security is today a critical issue in modern computing settings [1],[2].

To identify bad behaviour, intrusion detection system (IDS) are used to track system action and network traffic. Two types of traditional IDS methods include signature-based systems and anomaly-based systems. Whereas anomaly-based

IDSs are capable of identifying the existence of unknown attacks at the cost of high false alarms rates, signature-based IDSs are effective in identifying the existence of known attacks and a poor response to zero-day threats [3], [4].

Honeypots enhance intrusion detection and provide security teams with a chance to study attack approaches and patterns by acting as decoys to attract attackers. Any activity that is directed towards honeypots is considered suspicious because legitimate users are not supposed to visit them [5], [6]. Nevertheless, the decentralized character of current cyberattacks that are executed through botnets and proxy networks restricts the effectiveness of singlelocation honeypots [7].

The limitations have been suggested to be addressed by distributed honeypot systems sollucombined with machine learning (ML) and artificial intelligence (AI) technologies. Visions AI-based IDSs are able to analyse large volumes of attack data and can also detect complex patterns and adapt to new threats [8], [9]. Geo- Honeypot-IDS framework enhances the general resilience of cyber security by integrating distributed honeypots, geographic intelligence and AI analysis to deliver accurate, real time intrusion detection and global monitoring of attacks [10].

A. Contribution of thisPaper

The key findings of this study can be summarized in the following way:

A new Geo-Honeypot-IDS architecture has been suggested, which can be used to identify cyber attacks through distributed honeypot systems.

In the research, the distributed honeypot data collection is combined with the intrusion detection models that is based on machine learning in order to enhance the accuracy of attack detection.

To detect the patterns of attacks and the geographical source of cyber attacks, a geo- location analysis module is proposed. 4. It is tested by means of several benchmark datasets such as NSL- databank.

KDD, UNSW-NB15 and CICIDS2017, provide
complete experimental validation.

II. RELATED WORK

This segment analyses some of the current studies on distributed data collection, honeypot- based security, intrusion detection systems, and machine learning techniques applied in cybersecurity.

A. Systems for detecting intrusions

Intrusion detection systems (IDS) have been analysed in detail and grouped under the literature. A very early taxonomy of IDS was provided by Satpute [1], who classified them into

signature-based and anomaly-based approaches. Signature-based intrusion detection systems (IDS) identify intrusions by comparing the network traffic with known attack to the signatures but, they are useless against zero-day attacks, and will require frequent signature updates to remain accurate [2].

Anomaly-based IDS model is a model that is typical of the system behaviour and it finds deviations of this behaviour to circumvent such limitations [3]. Although anomaly-based techniques can pick attacks that have never occurred, the false-positive rates can often be very high since defining normal behaviour in a dynamic network environment is often difficult.

B. Security Based on Honeypots

Honeypots are security devices that are based on deception and are designed to attract and analysis a malicious activity. Spitzner [4] demonstrated the usefulness of honeypots in capturing the behaviour of attackers and understanding the methods used by attackers. Honeypots may be configured as low-interaction or high-interaction systems depending on the level of interaction between the attacker and the honeynet. Despite the fact that honeypots provide valuable data regarding the activity of malware and the way it is executed, they have limited effectiveness when applied independently due to the issues of scalability and poor visibility of attacks.

C. Dispersed Data Gathering for IDS

One of the primary success factors in improving the work of the IDS is the access to realistic and large volumes of data. The UNSW-NB15 database has been introduced by Moustafa and Slay [5] and enlists the authentic network traffic trends and present-day attack situations to IDS evaluation. These datasets are useful in enhancing the accuracy of detection to enable enhanced intrusion detection model training and validation.Intrusion Detection Methods Using Machine Learning Machine learning techniques, such as decision trees, support vector machines, artificial neural networks, and deep learning models have been extensively used in intrusion detection [6], [7]. The approaches have the potential of significantly enhancing the accuracy of detection due to the ability to learn the attack patterns using large data sets. In a recent study, the AI-based intrusion detection system (IDS) is found to be more efficient compared to traditional methods of detecting new threats and reducing false alarms, given that the AI-based system operates with less false alarms.

These techniques train attack patterns on big data and may significantly enhance the precision of detection. As recent studies have shown, AI- based IDS systems are superior to traditional ones in the detection of new attacks with reducing false positives in case of having high- quality training data available [8].

D. Synopsis and Research Deficit

It has been revealed that machine learning algorithms, anomaly-based detection, honeypots, and signature-based IDS systems can all be useful as far as combating cyber threats is concerned. Current solutions are however mostly piece meal and more of individual component approach as opposed to a holistic approach. Surveys on the extensive frameworks that incorporate intelligent analysis, deception technologies, and intrusion detection, particularly those with geographical knowledge, are small meaning that complex strategies, such as Geo-Honeypot-IDS [9], are required.

III PROPOSED METHODOLOGY

A. Datasets

The datasets that are used in the exploration of intelligent honeypots are in the following. This experimentation is carried out on datasets obtained in implementing honeypot environment and NSL-KDD KDD Cup 1999 Dataset, UNSW- NB15 Dataset and CICIDS2017 Dataset.

The benchmark intrusion detection dataset in the proposed research that would be utilized in the proposed intrusion detection system includes the KDD Cup 1999 Dataset [1], the UNSW-NB15 Dataset [2] and the CICIDS2017 Dataset [3] to provide thorough, standardized, and scientifically reproducible evaluation of the proposed Geo-Honeypot based intrusion detection system. KDD Cup 1999 is a standard dataset that contains the labeled types of attacks, including DoS, Probe, R2L, and U2R, which can be employed as the solid ground truth to train and compare the performance. In order to address the weaknesses that have been identified in the earlier studies conducted on intrusion detection [19], [20], more realistic and contemporary patterns of traffic are suggested by using the UNSWNB15 and CICIDS2017 datasets, which enhances the viability and generalization power of the proposed system. Besides, application of these datasets is aligned with the machine learning-based IDS methods developed [21]-26]. The standardized benchmark datasets together with the geo-distributed honeypot traffic guarantee the rigorous validation procedure, the increased accuracy, and the greater adaptability to both the traditional and advanced cyber threats in the real-world network settings.

The training process was done on the NSL-KDD
[1] dataset to build and test the initial intrusion detection models with clear and balanced types of attacks, and the UNSW-NB15 Dataset [2] to incorporate modern and diverse types of attacks such as Fuzzers, Exploits, DoS and

Reconnaissance attacks, which are more representative of the actual network traffic patterns. To conduct the evaluation and compare the performance, the testing was carried out with the CICIDS2017 Dataset [3] and relevant test split of the UNSW-NB15 dataset [2]. Further, the obtained number of samples of network traffic of the geo-distributed honeypot network configuration comprised of 72,000, which is consistent with the current research practice on honeypot-based IDS systems [4]-[9] and machine learning-based systems [19]-[26], thus contributing to the improvement of the experimental verification procedure in field settings.

The feature sets used in this work are taken using three benchmark intrusion detection dataset in order to offer a well rounded coverage of network activities. The NSLKDD [1] dataset contains 19 features, which are network connection characteristics, including basic TCP/IP characteristics (e.g., connection time, protocol, service, flag, source and destination byte), content based characteristics (e.g., login attempt) and time based characteristics (e.g., connection and error rate), based on the best practices of a traditional IDS feature design [24], [22]. UNSWNB15 Dataset [2] comprises 23 features that reflect the modern network traffic pattern, including flow-related features (source/destination IP addresses and ports, protocol, state), packet-level and byte-level statistics, time-related features, and TCP-related parameters, according to the modern best practices of machine learning-based intrusion detection design [21], [25], [26]. Further, CICIDS2017 Dataset [3] contains well rounded flow-level features derived with CICID Flow Meter such as flow duration, packet length statistics, inter-arrival times, traffic rates, and TCP flag features, according to the current best practices in attack modeling research on modern cybersecurity analytics [28], [27]. The combination of these sets of features will enable the model to capture both classical and contemporary patterns of cyber-attack, enhancing the generalization of the intrusion detection system.

The datasets used in this study offer both conventional and contemporary network traffic patterns, which allows the detection and testing of the proposed intrusion detection system in their entirety.

Table 3.1: Dataset Description

Dataset	Type	No. of Samples	Features
NSL-KDD	Training Dataset	29,400	19
UNSW-NB15	Training Dataset	28,200	23
CICIDS2017	Testing Dataset	14,000+	14+

The datasets used in this research provide both traditional and modern network traffic patterns, enabling comprehensive training and evaluation of the proposed intrusion detection system.

B. Honeypot

The network design has two layers of security mechanisms as discussed adequately in the network design. Various levels of security application should be in place to provide better security. Even minor flaws in the design of the system can be used by attackers. The outdated software on the IoT devices forms the first tier of security. Outdated software is prone to malfunction compared to the new software. Hackers are likely to get into a network by trapping IoT devices. When one of the hackers catches an internet of things device, the IP will be blocked by a firewall. Since the honeypot deployed is a high interaction honeypot, a limited number of ports were opened that simulate the open ports of software that is executing within the honeypot like the port of Microsoft update plugin or port of database and TCP, UDP, TELNET, POP3, FTP, SMTP, etc..

C. Intrusion Detection System by Machine Learning

Normal intrusion and normal web traffic are not normally considered as dangerous to the common packets that the firewall blocks. The firewall can either be a hardware firewall or it can also be a software based firewall to filter the packets. The data of all the devices are combined at the HUB and transmitted to all machines. An intelligent attacker is able to detect the use or any other form of packet sniffing, and even the attacker is able to alter or delete log les and reports created by the Honeypot Server, yet the use of Hub allows it to appear as a natural device which is used to connect devices.

D. Threat Detection

Different algorithms of AI are used to define whether the relation is good or it is an invader. The information of the honeypot is analysed effectively using the AI methods that contain approximately 23 features prediction. They are compared with the help of Different machine learning models such as Random Forest, Gradient Boosting, and Logistic Regression as it is stated above. The one that is appropriate to intrusion detection provides a high value of the accuracy and the high accuracy model is chosen to be detected.

E. Mathematical Model Formulation

The Geo-Honeypot Intrusion Detection System (GH-IDS) may be designed as a monitored machine learning classification problem. The goal will be to categorize network traffic logs into normal or malicious attack.

Dataset Representation

Let the honeypot dataset be represented as: $D=\{(x1,y1),(x2,y2),(x3,y3),...,(xn,yn)\}$
where:

- xi represents the **feature vector** extracted from network traffic
- yi represents the **class label** $yi \in \{0,1\}$

Where:

- 0 → Normal Traffic □ 1 → Malicious Attack

Each feature vector contains attributes such as: $xi=(f1,f2,f3,...,fm)$
Where:

- f1 = Protocol Type
- f2 = Destination Port
- f3 = Packet Size
- f4 = Connection Duration
- f5 = Number of Failed Logins
- fm = Other network behavioural features

Logistic Regression Model

Logistic regression estimates the probability of an attack using the sigmoid function: $P(y=1 \mid x)=1/(1+e-(wTx+b))$
Where:

- w= weight vector
- b = bias term
- x = input feature vector The classification rule is:

$y\hat{}=\{1 \text{ if } P(y=1 \mid x)>0.5$

ISSN: 25830-6129
DOI:10.55041/ISJEM05734

{0 otherwise

Random Forest Model

Random Forest builds multiple decision trees and aggregates predictions.

For a forest with T trees:

$RF(x)=1/T\sum\ t=1\ ht(x)$

Where:

- $h(x)$ = prediction from tree t

The final prediction is determined by **majority voting**.

Gradient Boosting Model

Gradient Boosting minimizes prediction error by iteratively updating the model:

$Fm(x)=Fm-1(x)+\gamma mhm(x)$ Where:

- $Fm(x)$= final model at iteration mmm
- $hm(x)$ = weak learner
- γm = learning rate

Ensemble Prediction Model

The final prediction of the GH-IDS system is obtained using **ensemble voting**:

Y final=Mode(YLR,YRF,YGB) Where:

- YLR = Logistic Regression prediction
- YRF = Random Forest prediction
- YGB = Gradient Boosting prediction

This ensemble approach improves detection accuracy and reduces misclassification errors.

IV. SYSTEM ARCHITECTURE

The proposed Geo-Honeypot Intrusion Detection System is a layered system, which is a combination of distributed honeypot sensors, data processing modules, machine learning models, and geo-location analysis to identify and analyse different types of cyber attacks.

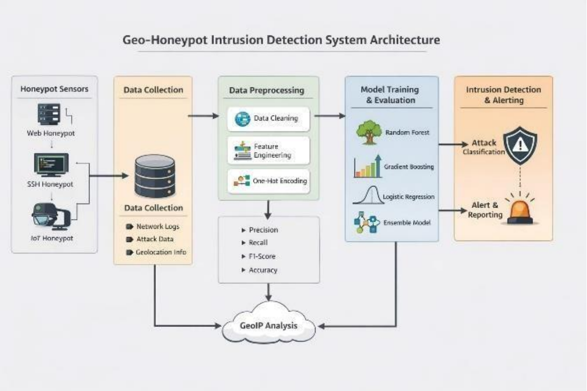

Fig 4.1 Architecture Diagram for Geo- Honeypot Intrusion Detection System Architecture

Fig. 4.1 the architecture of the proposed Geo Honeypot Intrusion Detection System (Geo Honeypot-IDS) is shown. The system is modeled as a description of a multi- layer cybersecurity system comprising of distributed honeypot sensors, centralized data processing, machine learning- based intrusion detection, and geographical analysis of attacks. The architecture allows the system to identify malicious activities in the network, analyse attack patterns, and have intelligent detection of cyber threats.

Fig. 4.1 illustrates that the system has several interconnected layers that collaborate in countering cyber attacks by detecting and analysing them. Honeypot Sensor Layer also deals with the deployment of distributed honeypots, which pretend to use vulnerable services like a web server, SSH services, and internet of things devices to lure attackers and capture malicious traffic. The obtained network logs are loaded into the Data Collection Layer, to which all the attack data are saved in a central repository to be analysed further. The obtained data is further transformed in the Data Processing Layer where data cleaning, feature extraction, coding of categorical variables, and normalization of numerical variables are carried out to hand over data to machine learning models. The trained dataset is then presented to the Model Training and Evaluation Layer where various machine learning methods such as Logistic Regression, Random Forest and Gradient Boosting are trained to identify network traffic as either being normal or malicious.

Intrusion Detection and Alerting Layer makes use of the trained model to detect potential cyber attacks and send alerts to the observed malicious activity. Lastly, the Geo-IP Analysis Module can be used to map the IP addresses of the attackers to their physical location, which helps security analysts to know the origin and spread of cyber attacks. This multi-layered design of architecture facilitates effective data gathering, smart attack recognition and better analysis of threats in contemporary networks..

A. Honeypot Sensor Layer

The Honeypot Sensor Layer is a distributed honeypot system that observes the activity of attackers by imitating vulnerable services to lure malicious traffic by attackers. The suggested system is composed of web honeypots, SSH honeypots, and IoT honeypots which identify various forms of attacks such as the web attacks, brute force attacks and the attacks by botnets. The honeypots observe the actions of the attacker and records the detailed network traffic.

B. Data Collection Layer

The Data Collection Layer receives the network traffic logs created by the honeypot sensors into a centralized location. Data Collection Layer logs the network traffic logs the source IP address, destination IP address, ports, protocols, and timestamps, attack payload, and other network-related data.

C. Data Preprocessing Layer

The Data Preprocessing Layer will have the task of processing the gathered network traffic data to be used in machine learning. Prior to the training of the machine learning models, there were various preprocessing processes conducted to guarantee the quality and consistency of the data. First, incomplete values and records were eliminated out of the data. The values of categorical characteristics like protocol type and service were transformed into numbers with the use of one-hot encoding. Standard Scaler was used to scale the features and normalize the numeric features and enhance model performance. A split ratio of 70:30 was then used to split the dataset into training and testing sets. To make the model robust, 10-fold cross- validation was used in the training of the model. The grid search methods were used to perform hyperparameter tuning of the machine learning models to identify optimal parameters of Random Forest, Gradient Boosting, and Logistic Regression models. These preprocessing and training procedures assist in enhancing stability of models, decreasing overfitting and enhancing the entire performance of the proposed intrusion detection system. The cleaned data is then delivered to the machine learning models after preprocessing to train and make attack classification.

D. Model Training and Evaluation Layer

Here, machine learning models are trained to classify network traffic and identify attacks. Different machine learning models, including Logistic Regression, Random Forest, Gradient Boosting, and others, are trained on the processed data. Ensemble model is also developed to merge several models that will give correct prediction. Intrusion Detection and Alerting Layer

The Intrusion Detection Layer performs the analysis of the data in real-time with the help of the machine learning model. In this case, different forms of attacks have been identified such as brute-force attacks, scanning attacks as well as exploitation attacks. An alert is provided when the attack is detected.

E. Geo IP Analysis Module

Geo IP Analysis Module can be applied in the mapping of the IP address of attackers and their respective geographical locations.

V. EXPERIMENTAL RESULTS

The proposed Geo- Honeypot-IDS system was evaluated using several machine learning models that are trained on information collected on scattered honeypots. As seen in the experimental results, the Ensemble Model had the best results among the Gradient Boosting (95.00%), Random Forest (96.00%), and Logistic Regression (90.00%) models since it had the highest accuracy (97.00%). The performance trend and accuracy comparison is indicated by corresponding figures.

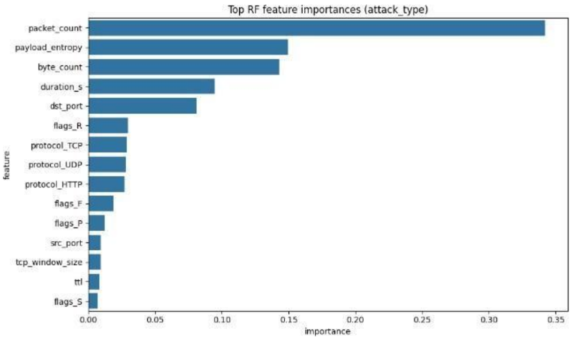

Fig 5.1 Random Forest Feature Importance

The figure above is the significance of features in a Random Forest model in the classification of

attack types. This indicates that the major feature is the packet count, then payload entropy, the count of bytes, and durations and hence it implies that the volume of traffic, the randomness of payloads, and the duration of the session are more relevant variables to determine the type of attack. Such characteristics as the destination port, network protocols (TCP, UDP, HTTP), and TCP flags can be considered relatively significant to the classification task, whereas source port, TCP window size, and TTL are not very significant. This discussion suggests that payloads of traffic behaviour are more significant than features of protocols needed to classify attacks in an efficient manner.

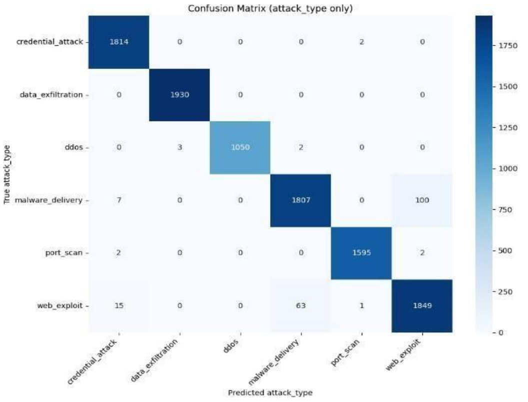

Fig 5.2 Confusion Matrix

The confusion matrix demonstrates how well the suggested intrusion detection model detects attacks of various categories. The classes of attacks that are examined in the present research are based on the benchmark datasets applied in the experiment, specifically NSL- KDD, UNSW-NB15, and CICIDS2017.

In the case of the NSL-KDD, there are major different attack types, such as Denial of Service (DoS), Probe attacks, Remote to Local (R2L) attacks, and User to Root (U2R) attacks, and also normal traffic. In the same vein, theUNSW-NB15 and CICIDS2017 datasets have more attack types like Exploits, Fuzzers, Reconnaissance and Web-based attacks and hence better represent the current cyber threats.

As can be seen in the confusion matrix, the majority of predictions are along the diagonal elements, which means that the model identifies the majority of the instances of attacks correctly. There are few cases of misclassifications that are seen between different types of attacks because of similarities in the traffic pattern of the particular types of attacks. On the whole, the findings indicate that the suggested Geo-Honeypot-IDS framework will be able to effectively identify the normal network traffic and various categories of computer attacks.

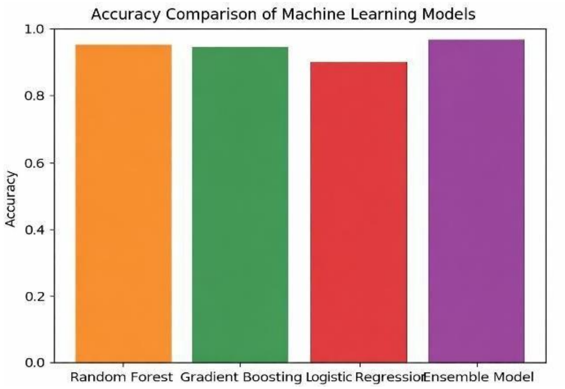

Fig 5.3 Accuracy Comparison

The accuracy comparison of the proposed Geo- Honeypot-IDS method compared to the current intrusion detection algorithms is presented in the figure 5.3. The current methods have relatively low detection performance because of their low learning capacity and absence of real time attack intelligence. Instead, the recommended algorithm is more accurate and its effectiveness proves that it is efficient in identifying malicious activity. The performance improvement is attributed to the integration of distributed honeypots, which collect different types of attack data, and AI based learning models, which

enhance the ability to classify. The comparison reflects the superiority of the proposed method in regards to detection accuracy and dependability.

Algorith m	precisio n	Recal l	F1-Scor e	Accurac y
Existing	0.95	0.94	0.95	95%
Proposed	0.97	0.96	0.96	97%

Table 5.1 is the performance of the proposed algorithm compared with the existing algorithm on the basis of the performance measures. The metrics of the proposed algorithm are better than the current algorithm, whose precision, recall, and F1-score value are 0.97. Also, the accuracy of the proposed algorithm is greater.

Instead of 95.00% which is the current algorithm. It would mean that the detection results of the suggested algorithm are more accurate.

Table 5.2 Performance Comparison of Models

Model	Precision	Recall	F1-Score	Accuracy	Training time
Random Forest	0.96	0.95	0.95	96%	2.8s
Gradient Boosting	0.95	0.94	0.94	95%	4.1s
Logistic Regression	0.90	0.89	0.89	90%	1.5s
Ensemble model	0.97	0.96	0.96	97%	0.8s

The performance of the applied machine learning models is compared in Table 5.2 according to the precision, recall, F1-score, accuracy, and training time. The individual models performed well with the Random Forest coming up with the best performance of 0.96 accuracy and the Gradient Boosting with 0.95 accuracy. Logistic Regression was found to be relatively low in performance with 0.90 accuracy.

The presented Ensemble Model has recorded the highest overall result of 0.97 precision, 0.96 recall, 0.96 F1-score and a 97 percent accuracy showing that even with multiple classifier combination, the process of detection will be enhanced and the Geo- Honeypot IDS will be more effective.

CONCLUSION

In this paper, I have described a Geo-Honeypot- IDS architecture that combines distributed honeypot systems and machine learning based intrusion detection methods. The proposed architecture will allow gathering actual attack data using honeypot sensors and will intelligently analyse the data to identify malicious behaviour in the network. Experimental assessment of the proposed system on benchmark datasets, including NSL-KDD, UNSW-NB15, and CICIDS2017, indicated that the proposed system can perform reliable intrusion detection.

The present study is limited in a number of ways even though the outcome is promising. The system was tested primarily on offline data and no actual real-time application to large-scale production was tested. Also, the research is centered on the classical machine learning models as opposed to deep learning models.

The future effort will be to expand on the framework to incorporate deep learning based intrusion detection model, real-time monitoring of the attacks, and implementation of this system in a cloud-based honeypot setting to examine large-scale cyber-attacks. Moreover, the introduction of automated response systems and visualization dashboards can also make the proposed system more practical

AUTHOR CONTRIBUTION

The author came up with the conceptualization and architecture of Geo-Honeypot Intrusion Detection System. The author designed the honeypot environment where the malicious activities are to be captured and applied the intrusion detection models. The author also incorporated the geolocation module that terminates the geographical origin of attackers by the use of IP addresses. In addition, the writer conducted data pre processing, model training and performance evaluation and analysis. The author drafted and completed the manuscript, which comprised of system design, methodology, results and documentation.

FUNDING

No external funding was provided in this research. The researcher worked on the study independently with the help of the available academic sources and cloud-based platforms to collect the data, process, and experiment.

ACVAILABILITY OF DATA MATERIALS

The attack records, geolocation mappings of IPs, and records of intrusion detection created in the honeypot set-up are safely stored. Because of the aspects of security and privacy, the datasets that will be utilized in this research are not publicly accessible, but they can be shared on the basis of the reasonable request and intended academic and research purposes.

DECLARATIONS

The author states that no conflict of interest exists that is connected with this research. The article is centered on the development of a Geo-Honeypot Intrusion Detection System that monitors and identifies malicious actions in real time, identifies the geographical location of attackers in order to monitor and analyse cybersecurity.

REFERENCES

[1]Satpute, A., Nikam, S., Gaikwad, V., Kakade, Y., and Mhaske, C. (2025). The intelligent intrusion detection system based on AI and SSH Honeypots. ICCK Transactions on Cybersecurity.

[2]Khan, A. H. (2023). A Framework of Generative AI-based Threat Detection and Cyber Deception: Honeypots in the Age of Generative AI. International Journal of Emerging Research in Engineering and technology.

[3]Gajjar, H., & Malek, Z. (2024). Honey pot based Network Intrusion Detection System in Cloud Environment. Smart systems and applications: International Journal of Intelligent Systems and Applications in Engineering.

[4]Abas, S. N., & Kaur, P. (2021). Honeypot: An Attacker Trap. International Journal of Scientific Research in Computer science, Engineering and IT.

[5]Shirsath, V. (2021). A Survey on Current States of Honeypots and Deception. Techniques of Attack Capture. International Journal of Engineering Research and. Technology.

[6]Dagar, M., & Popli, R. (2025). virtual Network Intrusion Monitoring System: Honeypots. International Journal of Scientific Research in Network security and communication.

[7]Čisar, P. (2025). The Position and the role of Honeypot Solutions in Network Intrusion

Detection System. Internet Research Transactions of IPSI.

Solanki, M., Petkar, J., Wanjari, R., Gajbhiye, S., and Kalode, V. (2025). Improving Cybersecurity through Dynamic Honeypots based on AI. International Journal of research Publication and Seminar.

Nagabhushanam, D., Kumar, U. J., Hemanth, K., Sai, A. N., and Saradhi Reddy, A. P. (2025). Honeypots with Deep Learning: ANN-based Detection of Sophisticated Cyber Threats. International Journal of Human.
Computations & Intelligence.

Franco, J., Aris, A., Canberk, B., and Uluagac, A.
S. (2021). A Survey of Honeypots and Honeynets to IoT, IIoT, and CPS. arXiv.

[11] Commey, D., Hounsinou, S., & Crosby,
F. V. (2024). Honeypots Strategic Deployment in IoT Blockchain-Based. Systems. arXiv.

[12] Trajanovski, T., & Zhang, N. (2021). A Framework: An Automated and Comprehensive Framework to detect and analyse IoT botnets (IoTBDA). arXiv.

Agiollo, A., Mahbooba, Z., and Palomares, I. (2022). Multi Agent Cyber Deception Framework of Adaptive Attacker. Engagement. IEEE Access.

[14] Singh, A., & Joshi, R. (2022). The Pervasive Survey of Cyber Deception.
Modern Threat Environment Technologies. Journal of Cybersecurity Technology.

[15] Gopireddy, S. R. (2022). Artificial Intelligence Honeypots: Improving Honeypots to Cyber Defense. International Journal of Advanced Computer Science.

[16] Garcia, J. B. (2023). High Interaction Honeypots with Docker: A Scalable Solution. International Journal of Cybersecurity Engineering.

Auti, A., Makwana, J., Pagar, S., Mishra, V., and Borade, S. (2023). HoneyTrack: A better Honeypot. Proceedings of

IEEE SCEECS.

Elsayed, A. (2021). Machine Learning to Zero-Day Attacks Detection: The Challenges and Opportunities. Cybersecurity Journal.

[18] Dehghantanha, A., & Gumaei, A. (2020). Cybersecurity Analytics: Future and Current Developments. Journal of Network and Computer Applications.

[19] Mitchell, R., & Chen, I. R. (2020). Incidences in the Behavioural Insider Threat Detection. IEEE
Information forensics and security transactions.

[20] Buczak, A. L., & Guven, E. (2020). Cybersecurity Intrusion Detection with machine Learning and Data Mining: Current Trends. IEEE Tutorials and Survey of communications.

[21] Javaid, A. Y., Niyaz, Q., Sun, W., & Alam,
M. (2020). Intrusion Detection through Deep Learning: Current Development. IEEE Access.

[22] Yin, C., Zhu, Y., Fei, J., & He, X. (2020).
Regular Neural Network Bases Intrusion Detection Systems: A|human|>Regular Neural Network Bases Intrusion Detection Systems: A|human|
Contemporary Study. IEEE Access.

[23] Liao, H. J., Lin, C. H. R., Lin, Y. C., &
Tung, K. Y. (2020). Intrusion Detection Pinpointing with Feature Selection and Ensemble Approaches: New Developments. The Expert Systems with Applications.

[24] Sommer, R., & Paxson, V. (2020). The Lessons and challenges in using Machine Learning in Network Intrusion Detection. IEEE Security & Privacy.

[25] Shabtai, A., Menahem, E., & Elovici, Y. (2020). Malware Detection of Behaviours in Current Networks. Journal of Smart Information Systems.

[26] Franco, J., & Canberk, B. (2022). IoT Networks: Cyber Deception: New.
Strategies. Internet of Things Journal, IEEE.

[27] Kumar, P., & Lee, S. (2023). Artificial intelligence-based intrusion detection on clouds. IEEE Transactions on Cloud computing.

[28] Zhang, Y., Wang, L., & Chen, X. (2024). Adaptive Honeypot Framework to detect Advanced Persistent Threats. Computers & Security.

[29] Sharma, R., & Patel, M. (2026).
Smart Geo-Distributed Honeypot Arch. to obtain Threat Intell. in Time. IEEE Access.

[30] R. Sommer and V. Paxson, “Outside the Closed World: On the Use of Machine Learning to Network Intrusion Detection Review IEEE Security and Privacy, vol. 8, no. 6, pp. 4251 2010.

[31] Buczak and Guven, A. L., and E. Guven, A Survey of Data Mining and Machine Learning Approaches to Cyber Security Intrusion Detection, IEEE Communications Surveys and Tutorials, vol. 18, no. 2, p. 1153-1176, 2016.M. Ring, S. Wunderlich, D. Grüdl, D. Landes, and A. Hotho, A Survey of Network- Based Intrusion Detection Data Sets, Computers &Security, Elsevier, vol. 86, pp. 147167, 2019.

[32] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, “Towards Generating a New Intrusion Detection Data set and Intrusion Traffic Characterization, ICISSP, 2018.

[33] N. Moustafa and J. Slay, N. Moustafa, J. Slay, UNSWNB15: A Comprehensive Data Set of Network Intrusion Detection Systems, Military Communications and Information Systems Conference, IEEE, 2015.