TIFDA JOURNAL
Technology Information and Data Analytic
Journal Website : http://jeptira.id/index.php/tifda
e-ISSN : 3064-0660

Detection of SSH Brute Force Attacks Using Naïve Bayes Classification on Cowrie Honeypot Logs in a Virtualized Environment

Arya Adhari Prasetyo[1], Herianto[2*], Yahya[3], Nur Syamsiyah[4]
[1,2] Information Technology Study Program, Darma Persada University
[3,4]Information System Study Program, Darma Persada University
[1,2,3] South Malaka Park Street, East Jakarta City, Special Capital Region of Jakarta 13450, Indonesia
*heri.unsada@gmail.com

Abstract — The increasing number of brute force cyberattacks targeting SSH services highlights the urgent need for effective early detection and mitigation systems. This study analyzes brute force attack patterns using the Naïve Bayes classification algorithm on log data generated by the Cowrie Honeypot. A virtual environment simulates attack scenarios and produces authentic SSH log data while maintaining server confidentiality. The system adopts the CRISP-DM framework, covering data preprocessing, model development, evaluation, and deployment. Experimental results show that the Naïve Bayes classifier achieves an accuracy of 91.8%, precision of 89.3%, recall of 90.7%, and F1-score of 90.0% in distinguishing brute force attempts from normal traffic. These results confirm the potential of combining Cowrie honeypot data with machine learning classifiers as an effective early warning tool for intrusion detection in enterprise network infrastructures.

Keywords—Brute Force Attack, Cowrie Honeypot, Naïve Bayes, Cybersecurity, Log Analysis, SSH

I. INTRODUCTION

In the era of digital transformation, cybersecurity has emerged as a critical concern for organizations and individuals alike. According to the 2023 Indonesian Cybersecurity Landscape report by the National Cyber and Crypto Agency (BSSN), over 430 million anomalous traffic events were detected throughout the year, with the highest volume occurring in August [1]. Among the most prominent threats were brute-force attacks targeting SSH (Secure Shell) services, which represent a significant portion of the cyber threats faced by public and private infrastructure [2, 3].

Brute force attacks involve systematically attempting a vast number of username and password combinations to gain unauthorized access. With the increasing availability of automated tools and open-source frameworks, this type of attack is not only persistent but also more sophisticated [4, 5]. Hence, early detection and effective classification of such attack patterns are essential to protect critical systems and data integrity [6].

Previous studies have demonstrated the potential of combining honeypot data with machine learning for detecting cyber threats. For example, [7] successfully applied the Naïve Bayes algorithm to Dionaea honeypot logs, achieving an F1-score of 84.9% in identifying port scanning attacks. Similarly, [8] integrated Cowrie honeypot with IDS Snort to detect SSH brute force and DoS attacks, confirming the effectiveness of honeypot-based threat intelligence. However, most existing works either lack a focused analysis on SSH brute force detection using probabilistic models or are limited to static log inspection without real-time or virtualized implementation [9, 10].

To address this gap, this study explores the deployment of a Cowrie honeypot, a medium-interaction security mechanism that emulates SSH services and logs all attempted intrusions. These log data are then analyzed using the Naïve Bayes classification algorithm, which operates on the principles of probabilistic inference and assumes conditional independence among features. The study is conducted in a virtualized environment that mirrors the

real network setting of PT Riwtech Workshop Indonesia to ensure safe and controlled experimentation.

In previous work, a room-level security system was developed using IoT infrastructure and two-factor authentication to prevent unauthorized physical access [11]. While this approach effectively addressed physical intrusion, it did not consider remote logical threats such as SSH brute force attacks. This study builds upon that foundation by shifting the focus to digital attack detection using honeypot-based data and probabilistic classification.

The objectives of this research are (1) to identify distinct brute force patterns from honeypot-generated logs, (2) to evaluate the effectiveness of Naïve Bayes in classifying normal vs. malicious activities, and (3) to assess the model's performance through standard metrics such as accuracy, precision, recall, and F1-score. By combining honeypot technology with machine learning classification, this study contributes a practical approach to enhancing proactive cybersecurity defenses, particularly in organizations transitioning toward digital operations.

Unlike previous works, this study focuses specifically on SSH brute force attack detection using Naïve Bayes applied to Cowrie honeypot logs generated within a fully virtualized and controlled enterprise-like environment. This research uniquely contributes by validating a lightweight, interpretable model that can be deployed as an early detection system, especially in infrastructure-limited organizations.

II. METHODOLOGY

This research adopts an experimental approach using a simulated cybersecurity environment to analyze SSH brute force attacks. The system architecture involves two primary components: the Cowrie honeypot, which acts as the attack trap and log generator, and the Naïve Bayes classifier, which is applied to analyze and classify attack patterns based on honeypot logs.

A. Data Collection

The primary dataset consists of log files generated by the Cowrie honeypot, a medium-interaction system designed to emulate an SSH server and record all unauthorized login attempts. Since real attack data from PT Riwtech Workshop Indonesia are confidential, a virtual environment was developed to simulate realistic attack scenarios using brute force tools. These logs are stored in structured JSON and text formats, capturing key attributes such as timestamp, attacker IP address, attempted usernames and passwords, and login status.

B. Research Framework: CRISP-DM

The methodology follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework[12], which consists of the following phases:

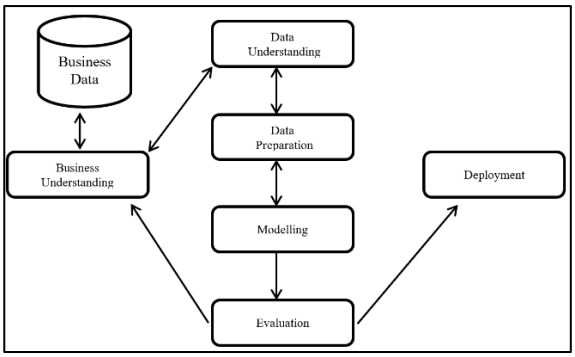

Figure 1. The Cross-Industry Standard Process for Data Mining (CRISP-DM)

a) Business Understanding
   Identification of cybersecurity risks, especially brute force attack vulnerabilities within the SSH service of a virtual enterprise server.
b) Data Understanding
   Exploration of Cowrie log formats, event types, and frequency of login attempts to gain insights into attacker behavior.
c) Data Preparation
   Preprocessing steps include log parsing, irrelevant feature removal, normalization, and transformation of categorical data into numerical format suitable for classification.
d) Modeling
   The Naïve Bayes algorithm is implemented to classify events as either malicious (brute force attack) or benign (normal access attempt) [13]. The model is trained on labeled samples derived from the prepared log data.
e) Evaluation
   Model performance is evaluated using a confusion matrix and related metrics: accuracy, precision, recall, and F1-score to determine the effectiveness of the classifier in distinguishing attack patterns [14].
c) Deployment
   Once validated, the model is deployed alongside the honeypot environment to function as an early detection system for SSH brute-force activity.

C. Data Preparation

After the collection and structuring phase, the Cowrie honeypot produced a total of 4,500 log records over the course of the attack simulation. Among these, 2,700 entries (60%) were identified as brute force attack attempts, while 1,800 entries (40%) were categorized as normal traffic.

The dataset was then divided into training and testing subsets using an 80:20 hold-out method, maintaining the same class distribution in both sets. Stratified sampling was applied to ensure balanced representation of attack and normal classes during training.

D. Software and Tools

The implementation and testing environment utilizes.

- Honeypot: Cowrie (running on Ubuntu Server)
- Development tools: Python, Jupyter Notebook
- Libraries: pandas, scikit-learn, matplotlib
- Virtualization platform: VirtualBox

By integrating Cowrie log analysis and Naïve Bayes classification within the CRISP-DM pipeline, this research aims to demonstrate an efficient and scalable approach to intrusion detection for small- to mid-scale enterprise systems.

III. RESULTS AND DISCUSSION

This section presents the experimental results from the simulation of SSH brute force attacks using the Cowrie honeypot, followed by analysis of classification performance using the Naïve Bayes algorithm. The dataset used for training and evaluation was generated in a virtual environment designed to mimic real-world enterprise server conditions:

A. System Implementation Results

The developed system was deployed on a virtual machine configured with Cowrie Honeypot to emulate a realistic SSH server. During the simulation phase, Cowrie recorded multiple unauthorized access attempts, which were stored in structured JSON log files. These logs were successfully extracted, cleaned, and preprocessed for model input.

Classification Performance Using the preprocessed dataset, the Naïve Bayes model was trained and tested to classify attack logs into two classes: "brute force attack" and "normal access." The model evaluation was based on confusion matrix metrics.
The summarized results are as follows:

Table 1. Confusion matrix-based metrics

Metric	Value
Accuracy	91.8%
Precision	89.3%
Recall	90.7%
F1-Score	90.0%

Figure 2 below shows the confusion matrix summarizing the classifier's performance:

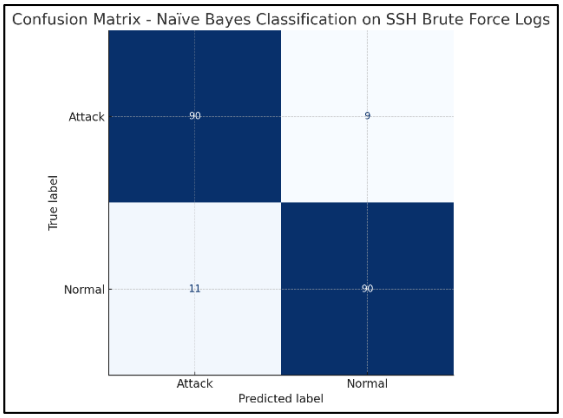

Figure 2. Confusion Matrix for Naïve Bayes

These results demonstrate that the Naïve Bayes classifier is highly effective at distinguishing between brute force attacks and legitimate login attempts, even under simulated real-world conditions. Classifier on SSH Brute Force Logs. These values indicate that the classifier performs well in detecting SSH brute force attacks with a high degree of reliability. The model was able to accurately distinguish between legitimate and malicious activities in most cases.

B. Comparative Analysis

To validate the effectiveness of the proposed system, its performance was compared with previous works such as the Naïve Bayes implementation on Dionaea honeypot logs [7], which reported an F1-score of 84.9%. In contrast, this study achieved an F1-score of 90.0%. The improvement is attributed to several factors:

a) Data specificity: This study focuses solely on SSH brute force attacks, enabling more targeted feature selection.
b) Honeypot quality: Cowrie, as a medium-interaction honeypot, generates more realistic and detailed logs compared to low-interaction systems like Dionaea.
c) Controlled environment: The use of a fully virtualized enterprise-like network allows for cleaner data and reduced noise compared to live traffic data used in other studies.
d) Preprocessing improvements: Structured log parsing, feature normalization, and categorical encoding enhanced model learning and reduced classification ambiguity.

These improvements underscore the practical advantage of combining Cowrie honeypot logging with probabilistic classification in structured, simulated environments.

C. Deployment Considerations

The trained model was integrated into the same virtual infrastructure, enabling real-time monitoring of Cowrie logs and automatic classification of potential brute force events. Future enhancements could include

integration with alerting systems or security dashboards to improve operational response time and automation.

IV. CONCLUSION

This research confirms that combining Cowrie honeypot log data with the Naïve Bayes classification algorithm offers a viable and effective method for detecting SSH brute force attacks. Through simulation in a controlled virtual environment, the system demonstrated high accuracy and reliability in distinguishing between normal and malicious SSH login attempts.

The success of this approach illustrates that even lightweight machine learning models, when paired with quality log data, can serve as effective tools for early intrusion detection. Unlike complex deep learning models, Naïve Bayes is computationally efficient and interpretable, making it suitable for deployment in environments with limited resources.

As a practical implication, this system can be implemented in small to medium-sized enterprises (SMEs), universities, or educational labs seeking low-cost and reliable cybersecurity monitoring. It provides a foundational framework for building automated alert systems and can be integrated into existing IT infrastructure with minimal overhead.

Future work may involve expanding the model to support multiclass attack detection, integrating real-time alert modules, or evaluating other probabilistic classifiers under varied network attack scenarios.

REFERENCES

[1] S. Mishra and S. Gochhait, "Emerging cybersecurity attacks in the era of digital transformation," in *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2023: IEEE, pp. 1442-1447.

[2] J.-K. Lee, S.-J. Kim, J. Woo, and C. Y. Park, "Analysis and response of SSH brute force attacks in multi-user computing environment," *KIPS Transactions on Computer and Communication Systems,* vol. 4, no. 6, pp. 205-212, 2015.

[3] M. D. Hossain, H. Ochiai, F. Doudou, and Y. Kadobayashi, "Ssh and ftp brute-force attacks detection in computer networks: Lstm and machine learning approaches," in *2020 5th international conference on computer and communication systems (ICCCS)*, 2020: IEEE, pp. 491-497.

[4] K. Apostol, "Brute-force attack," ed: SaluPress, 2012.

[5] M. M. Najafabadi, T. M. Khoshgoftaar, C. Kemp, N. Seliya, and R. Zuech, "Machine learning for detecting brute force attacks at the network level," in *2014 IEEE International Conference on Bioinformatics and Bioengineering*, 2014: IEEE, pp. 379-385.

[6] S.-E. Jeon *et al.*, "An Effective Threat Detection Framework for Advanced Persistent Cyberattacks," *Computers, Materials & Continua,* vol. 75, no. 2, 2023.

[7] D. K. NURILAHI, R. MUNADI, S. SYAHRIAL, and A. Bahri, "Penerapan Metode NaÃ¯ ve Bayes pada Honeypot Dionaea dalam Mendeteksi Serangan Port Scanning," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika,* vol. 10, no. 2, p. 309, 2022.

[8] T. Natanegara, Y. Muhyidin, and D. Singasatia, "Implementasi Honeypot Cowrie Dan Snort Sebagai Alat Deteksi Serangan Pada Server," *JATI (Jurnal Mahasiswa Teknik Informatika),* vol. 7, no. 3, pp. 1871-1877, 2023.

[9] H. Alosimy, J. AlZaidi, S. H. Alajmani, and B. Soh, "An Algorithm for Detecting Brute Force Attacks on FTP and SSH Services Utilizing Deep Learning with Probabilistic Neural Networks (PNN)," 2025.

[10] C. Menteng, A. Setyanto, and H. Al Fatta, "MODEL DETEKSI SERANGAN SSH-BRUTE FORCE BERDASARKAN DEEP BELIEF NETWORK," *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika,* vol. 17, no. 2, pp. 101-110, 2023.

[11] Herianto and E. M. Shamirah, "PERANCANGAN SISTEM KEAMANAN RUANGAN BERBASIS INTERNET OF THINGS DENGAN FITUR TWO-FACTOR ATHENTICATION (2FA)," vol. 13, ed, 2023, pp. 96-104.

[12] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, "The evolution of CRISP-DM for data science: Methods, processes and frameworks," *SBC Reviews on Computer Science,* vol. 4, no. 1, pp. 28-43, 2024.

[13] D. Berrar, "Bayes' theorem and naive Bayes classifier," 2025.

[14] S. Sathyanarayanan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," *African Journal of Biomedical Research,* pp. 4023-4031, 2024.