WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Innovative Machine Learning Techniques for Identifying Ransomware through BGP Records

M. Kavitha[1], R. Jaya Priya[2], P. Sangeetha[3], Muhammed Jawad K[4]

[1]Assistant Professor, Artificial Intelligence and Data Science, Kathir College of Engineering, Coimbatore. kavitha@kathir.ac.in
[2]Assistant Professor, Computer Science and Engineering, Kathir College of Engineering, Coimbatore. jayapriya@kathir.ac.in
[3]Assistant professor, Artificial intelligence and Data Science, Kathir college of Engineering Coimbatore, sangeetha@kathir.ac.in
[4]2nd year Master of CSE Student, Kathir College of Engineering, Coimbatore. afadjawad@gmail.com

A B S T R A C T

The Border Gateway Protocol (BGP) is essential for routing data on the Internet, and any issues with BGP can disrupt connectivity and cause problems like route disconnections and oscillations. In this paper, we explore different machine learning techniques to detect BGP anomalies, focusing on gradient boosting decision trees and deep learning models. We tested these models using data from the West Rock ransomware attack.

We also introduce BGPGuard, a tool that uses these advanced algorithms to monitor BGP in real-time and detect any unusual activity. Additionally, we present the Ransomware Readiness Detection Application, which uses the Random Forest algorithm to find suspicious behavior in Bitcoin transactions. By looking at transaction details such as amounts, frequency, and destinations, this application helps identify potential ransomware threats early.

Our research shows that machine learning can significantly improve cybersecurity by detecting BGP anomalies and ransomware activities. These tools give organizations a better chance to protect themselves against cyber threats.

Keywords: Border Gateway Protocol (BGP), Machine Learning, Ransomware, Cybersecurity, Anomaly Detection, BGPGuard, Ransomware Detection, Routing Disconnections.

1. Introduction

The Internet is made up of many networks called Autonomous Systems (ASs). These ASs are groups of routers managed by different organizations. They use a protocol called Border Gateway Protocol (BGP) to share information about how to route data across the Internet. BGP was introduced in 1989 and is now in its fourth version, BGP-4, which has been used since 1994. BGP-4 improved the protocol with error fixes, clarifications, and updates to support better routing practices.

BGP helps routers talk to each other using four types of messages: open, keep alive, update, and notification. An open message starts the connection, a keep alive message makes sure the connection is still active, an update message shares routing information, and a notification message is sent if there is an error, which then closes the connection.

In today's world, ransomware attacks are a major threat to organizations. These attacks use malicious software to lock or damage data, demanding payment to restore it. Traditional security methods often struggle to detect and stop these attacks effectively. To tackle this problem, we are developing a new tool called the Ransomware Readiness Detection Application. This tool uses advanced algorithms and data analysis techniques to detect ransomware activities within Bitcoin transactions. It aims to help organizations protect themselves from these threats.

Our Ransomware Readiness Detection Application is a significant step in improving cybersecurity. By using technologies like the Random Forest algorithm, we aim to help organizations find and stop potential threats before they cause harm. The application combines detailed data analysis with real-time monitoring to keep data safe and ensure business operations continue smoothly, even when faced with ransomware attacks.

As wireless networks become more common, they bring new challenges and potential security issues. These networks can be vulnerable to attacks, which makes it important to have strong security measures in place.

BGP, while essential for routing, can be complex and prone to errors. It lacks strong security features to verify route updates, making it vulnerable to anomalies. Anomalies in BGP can disrupt network performance and are often caused by issues like infrastructure failures, router misconfigurations, or

cyber-attacks such as IP hijacking or ransomware. To address these problems, machine learning techniques are being used to detect and analyze BGP anomalies.

This article explores the use of machine learning to understand and classify recent BGP anomalies, including those from the West Rock ransomware attack in January 2021. By analyzing BGP data and using machine learning algorithms like K-means, Isolation Forest, and various deep learning models, we aim to improve our understanding of network security and enhance defenses against emerging threats the paper, and put a nomenclature if necessary, in a box with the same font size as the rest of the paper. The paragraphs continue from here and are only separated by headings, subheadings, images and formulae. The section headings are arranged by numbers, bold and 9.5 pt. Here are further instructions for authors.

Nomenclature

Autonomous System (AS): A collection of routers under a single administration that uses a common routing policy and is identified by a unique number assigned by a regional Internet registry.

Border Gateway Protocol (BGP): A path-vector routing protocol used to exchange network reachability information between Autonomous Systems on the Internet. The current version, BGP-4, includes updates and improvements over earlier versions.

Ransomware: Malicious software designed to encrypt or damage a user's files and demand payment for decryption or restoration.

Ransomware Readiness Detection Application: A tool designed to detect ransomware-related activities, especially within Bitcoin transactions, using advanced algorithms and data analysis techniques.

Machine Learning: A branch of artificial intelligence that involves training algorithms to recognize patterns and make decisions based on data.

2. Requirement Specification

This aims to provide a detailed look at the Requirement Specification for the Ransomware Readiness Detection Application. This application is designed to help organizations stay ahead of ransomware attacks by analyzing Bitcoin transactions for signs of malicious behavior. The goal is to use advanced algorithms to detect any suspicious activity related to ransomware payments. By focusing on this specific area, the application aims to spot potential threats early and provide a proactive defense against them.

The application uses the Random Forest algorithm, a powerful tool for analyzing complex data, to examine Bitcoin transactions and identify patterns that might indicate a ransomware attack. This analysis is done in real-time, meaning the application can monitor transactions as they happen, allowing organizations to detect and respond to threats immediately.

Overall, the application is intended to improve cybersecurity by offering a robust mechanism for threat detection. By catching ransomware-related activities early, it helps organizations protect their critical data and maintain their operations without interruption. This proactive approach ensures that organizations are better prepared to handle and mitigate the impact of ransomware attacks, enhancing their overall resilience to such cyber threats.

BGP anomalies refer to unexpected changes in how routing information is shared across the Internet. These anomalies can disrupt BGP update messages and lead to harmful changes in how the Internet operates. This can slow down or even break Internet services. There are several causes for BGP anomalies:

- **Infrastructure Failures:** Problems like power outages or physical damage to network equipment (like cables and routers) can affect BGP.
- **Router Misconfigurations:** If routers are set up incorrectly, they can create issues such as:
    - Forwarding Loops: Data gets stuck in a loop between routers.
    - Oscillations: Routing paths keep changing back and forth.
    - Packet Loss: Data packets are lost during transmission.
    - Unintended Paths: Data might take unexpected routes.
    - Blackholing: Data is discarded without any notification.

These issues can cause packet loss and other problems, making the Internet less reliable.

3. Network Anomalies and Intrusions

3.1 BGP Anomalies

BGP anomalies refer to unusual or unexpected behaviors' in the Border Gateway Protocol (BGP) routing process that can indicate potential issues or security threats within a network. Common BGP anomalies include prefix hijacking, where an unauthorized entity advertises IP addresses it doesn't control, and route flapping, where routes frequently switch between available and unavailable states. These anomalies can disrupt network operations,

degrade performance, and expose the network to malicious attacks. Detecting and addressing BGP anomalies is essential for ensuring the security and stability of internet routing.

3.2 Network Intrusions

Network intrusions involve unauthorized access and theft of network resources. They can seriously impact network security. Here are some common types:

- **BGP Hijacking:** Attackers can redirect traffic by pretending to own certain IP address ranges. This can mislead data from its intended destination.
- **Worms:** These are types of malware that spread across networks. They can carry other harmful software like ransomware or viruses.
- **Ransomware:** This type of malware encrypts a user's files and demands payment to unlock them. During an attack, data may be stolen, encrypted, or deleted. Often, files are locked permanently until a ransom is paid.

Intrusions like worms and ransomware can cause an increase in the number of routing updates and changes in routing paths [4]. These attacks impact BGP routing even though routing, control, and management packets are kept separate from regular data traffic [2].

To combat these issues, machine learning algorithms are used to create systems that detect network intrusions. These systems are trained with datasets that include known anomalies. As cyber-attacks continue to evolve, these models must be updated with recent data to stay effective.

4. WestRock Ransomware Attack

In January 2021, WestRock, a large company that makes packaging, experienced a serious ransomware attack. Ransomware is a type of malicious software used by cybercriminals to lock or encrypt data and then demand money to unlock it. WestRock's IT systems, which manage information like emails and documents, and OT systems, which control machinery and production processes, were both affected by the attack. This caused significant disruptions to their operations.

The attack began on January 23, 2021. It wasn't just a quick hit; it lasted for several days, severely impacting the company's ability to function. Production slowed down, and shipments were delayed because the systems that normally manage these processes were compromised. To deal with the attack, WestRock had to shut down many of its systems to prevent further damage. They implemented a controlled remediation plan, which means they carefully followed steps to fix the issue. This plan included improving their existing security measures to prevent future attacks and gradually bringing their systems back online to ensure everything was safe.

The incident highlighted how disruptive ransomware attacks can be, not just for information systems but also for the physical operations of a company. It also showed the importance of having a robust response plan and strong security measures to handle such cyber threats.

5. Proposed System

5.1 Proposed Methods

Using machine learning algorithms like Random Forest and XGBoost can significantly improve the security and efficiency of BGP routing protocols. These algorithms help detect unusual patterns, predict routing paths, and identify potential threats.

Random Forest Algorithm:

- Anomaly Detection: Finds unusual BGP announcements that could indicate malicious activity.
- Routing Behaviour Prediction: Anticipates the likely path of a BGP announcement.
- Threat Identification: Spots potential prefix hijacking or route flapping.

XGBoost Algorithm:

- Anomaly Detection: Improves the identification of suspicious BGP announcements.
- Routing Behaviour Prediction: Enhances the prediction of BGP routing paths.
- Threat Identification: Detects security threats more effectively.

Both algorithms can handle large amounts of data and complex patterns, making them ideal for analyzing BGP routing protocols. Using these methods can lead to early detection of issues, better prediction of routing behaviour, and improved network security.

5.2 Data Pre-processing

Using machine learning algorithms like Random Forest and XGBoost can greatly enhance the security and efficiency of BGP routing protocols. These algorithms help by detecting unusual patterns, predicting routing paths, and identifying potential threats. The Random Forest algorithm finds unusual BGP announcements that might indicate malicious activity, predicts the likely path of a BGP announcement, and spots potential issues like prefix hijacking or route flapping. Similarly, the XGBoost algorithm improves the identification of suspicious BGP announcements, enhances the prediction of routing paths, and more effectively detects security threats. Both algorithms are capable of handling large amounts of data and complex patterns, making them ideal for analyzing BGP routing protocols. Using these methods can lead to early detection of issues, better prediction of routing behaviour, and improved network security.

5.3 GOALS

1. Anomaly Detection: Identify unusual BGP announcements that might indicate malicious activities.
2. Routing Behaviour Prediction: Predict the likely path a given BGP announcement will take.
3. Threat Identification: Spots potential prefix hijacking or route flapping.

Border Gateway Protocol (BGP) is crucial for routing decisions on the Internet, but it is susceptible to various security threats, such as prefix hijacking and route flapping. Applying the XGBoost algorithm to analyze BGP routing records can significantly enhance the detection of anomalies, prediction of routing behaviours, and identification of potential security threats. This approach leverages the capabilities of XGBoost, a powerful gradient-boosting algorithm, to improve the security and efficiency of network operations.

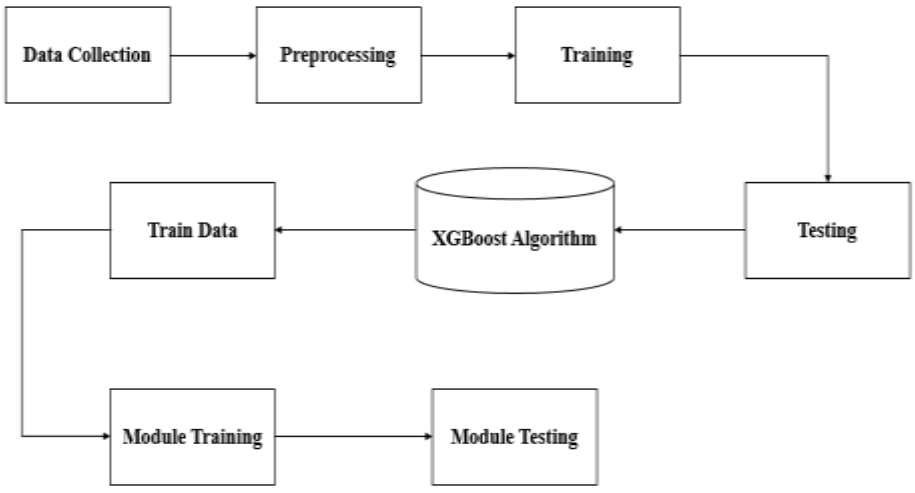
5.4 Advantages

To ensure the effectiveness of the proposed method, the data used for analysis must be diverse and representative of various network traffic patterns and behaviours. This diversity ensures that the system can accurately capture the complexities of real-world network environments and effectively detect anomalous activities, including those associated with the West Rock ransomware attack.

Both collected and generated data will undergo pre-processing to remove any irrelevant or redundant information while extracting relevant features. This pre-processing step is essential for optimizing the dataset for analysis, ensuring that the machine learning algorithms can effectively learn from the data and make accurate predictions. By eliminating noise and focusing on pertinent features, the system can enhance its detection capabilities and minimize false positives.

Once the data has been pre-processed and analyzed by the Border Gateway Protocol (BGP) system, any detected malicious activity will trigger alerts to notify system administrators. These alerts serve as early warnings, allowing administrators to take immediate action to mitigate potential security threats and safeguard the network infrastructure. By leveraging the output of the BGP system to alert administrators, the proposed method enables proactive response measures, thereby enhancing the overall network security posture.

5.5 System Architecture



5.6 System Environment

6. Results and Discussion

In this chapter, we discuss the outcomes of our proposed system for detecting anomalies and threats in BGP (Border Gateway Protocol) routing data. Our goal is to show how well our machine learning models, particularly Random Forest and XGBoost, perform in identifying unusual patterns and predicting routing behaviours. We start by looking at the effectiveness of the Random Forest algorithm. This algorithm helps us spot unusual BGP announcements, which might indicate malicious activities like prefix hijacking or route flapping. Prefix hijacking is when an attacker misleads routers into sending data through an incorrect path, while route flapping involves frequent changes in the data path, causing instability. The Random Forest algorithm uses multiple decision trees to improve accuracy, making it reliable for detecting these anomalies.

Next, we focus on the XGBoost algorithm. XGBoost is a more advanced machine learning technique that boosts the performance of our anomaly detection system. It helps us predict routing behaviours and detect threats more accurately. By analyzing large datasets, XGBoost identifies patterns that might be missed by simpler models. We assess its performance using various metrics such as accuracy, precision, recall, and F1-score. These metrics help us understand how well the model can correctly identify anomalies and avoid false alarms. We also discuss the data collection and preprocessing steps. We gather BGP routing records from public repositories like Route Views and RIPE NCC. This data includes information about how internet traffic is routed between different networks. To make the data usable, we clean it by handling missing values, normalizing features, and extracting important attributes. This preprocessing step ensures that our machine-learning models receive high-quality data for training and evaluation.

Once the data is prepared, we split it into training and test sets. The training set is used to teach our models to recognize patterns and anomalies in BGP routing data. The test set is then used to evaluate the models' performance. By comparing the predicted outcomes with the actual data, we measure how accurately our models can detect anomalies. After training and testing, we deploy our models for real-time monitoring. This means integrating the trained models into a network's monitoring system to continuously analyze BGP traffic. Our system can now detect anomalies and predict routing behaviours in real-time, providing ongoing protection against cyber threats. By identifying potential threats early, we can prevent attacks and maintain the stability and security of internet operations.

In the discussion section, we analyze the results in detail. We compare the performance of Random Forest and XGBoost, highlighting their strengths and weaknesses. We also consider the practical implications of our findings, discussing how our system can be used in real-world scenarios to enhance network security. Finally, we suggest future improvements and potential areas for further research to make our anomaly detection system even more effective.


0
Header Length:
54
Protocol Type:
6
Duration:
64
Rate:
0.329807153
Srate:
0.32980715
FIN Flag Number:
1


IOT Cyber Attack
Username
Password
Login

7. Conclusion

In this chapter, we summarize the key points and findings of our research on using machine learning algorithms to detect anomalies and threats in BGP (Border Gateway Protocol) routing data. We also discuss the implications of our work and suggest directions for future research.

Firstly, we highlight the importance of BGP in the functioning of the internet. BGP is the protocol that routes data between different networks, making it crucial for internet connectivity. However, it is vulnerable to various types of attacks and anomalies, such as prefix hijacking and route flapping, which can disrupt network operations and compromise security. Our research aimed to address these challenges by developing a system that uses machine learning to analyze BGP routing data. We focused on two main algorithms: Random Forest and XGBoost. The Random Forest algorithm, with its multiple decision trees, proved effective in detecting unusual BGP announcements and predicting routing behaviours. It helped us identify anomalies that might indicate malicious activities. The XGBoost algorithm, known for its high performance, further enhanced our anomaly detection capabilities by accurately predicting routine behaviours and identifying threats. We collected BGP routing records from public repositories like RouteViews and RIPE NCC and preprocessed this data to ensure its quality. The preprocessing involved cleaning the data, normalizing features, and extracting meaningful attributes. This step was crucial for training our machine learning models effectively. We then trained our models using a portion of the data and tested their performance on the remaining data. The results showed that both Random Forest and XGBoost performed well in detecting anomalies and predicting routing behaviours. We evaluated their performance using metrics like accuracy, precision, recall, and F1-score, which provided a comprehensive assessment of their effectiveness. One of the significant achievements of our research was the real-time deployment of our models. By integrating them into a network's monitoring system, we enabled continuous analysis of BGP traffic. This real-time monitoring allows for the early detection of anomalies and threats, enhancing network security and stability.

In our discussion, we compared the performance of the two algorithms, highlighting their strengths and areas for improvement. We also discussed the practical implications of our findings, emphasizing how our system can be used to protect networks from cyber threats in real-world scenarios.

Looking forward, we suggest several directions for future research. One area is to explore other advanced machine learning algorithms that could further improve anomaly detection. Additionally, we recommend extending our system to analyze other types of network data and integrate it with different network monitoring tools. By continuing to refine and expand our system, we can stay ahead of evolving cyber threats and contribute to the ongoing efforts to secure internet operations.

In conclusion, our research demonstrates the potential of machine learning algorithms in enhancing the security of BGP routing. By detecting anomalies and predicting routing behaviours, our system provides a proactive defence against various types of network threats. We believe that our work lays a strong foundation for future research and development in this critical area of network security.

Acknowledgements

The authors wish to thank everyone who has contributed to the success of this research work.

References

1. B. Al-Musawi, P. Branch, and G. Armitage, “BGP Anomaly Detection Techniques: A Survey,” IEEE Communications Surveys & Tutorials (Volume: 19, Issue: 1, Firstquarter 2017) pp. 377–396.
2. P. Mishra et al., “A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection,” IEEE Communications Surveys & Tutorials (Volume: 21, Issue: 1, Firstquarter 2019) pp. 686 - 728
3. V. Chandola, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” ACM Computing Survey (CSUR), vol. 41, Issue 3, July 2009, pp. 1–15
4. T. G. Griffin and G. Wilfong, “On the Correctness of IBGP Configuration,” Proc. ACM SIGCOMM (Pittsburgh, PA, Aug. 2002) pp. 17–29
5. W. Zhong, N. Yu and C. Ai, "Applying big data based deep learning system to intrusion detection", Big Data Min. Anal., vol. 3, no. 3, pp. 181-195, Sep. 2020.
6. M. H. Haghighat and J. Li, "Intrusion detection system using votingbased neural network", Tsinghua Sci. Technol., vol. 26, no. 4, pp. 484-495, Aug. 2021.
7. Y. Yang et al., "ASTREAM: Data-stream-driven scalable anomaly detection with accuracy guarantee in IIoT environment", IEEE Trans. Netw. Sci. Eng., Mar. 2022
8. I. Homoliak, K. Malinka, and P. Hanacek, “Asnm datasets: A collection of network attacks for testing of adversarial classifiers and intrusion detectors,” IEEE Access, vol. 8, pp. 112 427–112 453, 2020.
9. Y. K. Muhammad Ashfaq Khan, “Deep learning-based hybrid intelligent intrusion detection system,” Computers, Materials & Continua, vol. 68, no. 1, pp. 671–687, 2021
10. Van der Geer, J., Hanraads, J. A. J., & Lupton, R. A. (2000). The art of writing a scientific article.*Journal of Science Communication, 163*, 51–59.
11. Strunk, W., Jr., & White, E. B. (1979).*The elements of style* (3rd ed.). New York: MacMillan.
12. Mettam, G. R., & Adams, L. B. (1999).How to prepare an electronic version of your article. In B. S. Jones & R. Z. Smith (Eds.), *Introduction to the electronic age* (pp. 281–304). New York: E-Publishing Inc.
13. Fachinger, J., den Exter, M., Grambow, B., Holgerson, S., Landesmann, C., Titov, M., et al. (2004).Behavior of spent HTR fuel elements in aquatic phases of repository host rock formations, 2nd International Topical Meeting on High Temperature Reactor Technology. Beijing, China, paper.
14. Fachinger, J. (2006). Behavior of HTR fuel elements in aquatic phases of repository host rock formations.*Nuclear Engineering & Design,236*, 54.
15. A Border Gateway Protocol 4 (BGP-4),” IETF RFC 1654; https:// datatracker.ietf.org/doc/html/rfc1654, accessed Oct. 18, 2022