Journal of Information Systems Engineering and Management
2025, 10(29s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Cybersecurity through the Analysis of BGP Protocol for Detecting and Classifying Security Threats

Jenan Jader 1, Nada Jabbar Dubai 2, Ihab Hamza Ali 3
Computer systems technologies/ karbala technical institute /al-furat alawast technical university
jenan.jader@atu.edu.iq
division of insurance technologies/ al diwaniyah technical institute /al-furat alawast technical university
nada.jabbar.idi@atu.edu.iq
Computer systems technologies/ karbala technical institute /al-furat alawast technical university
Ihab.ali@atu.edu.iq1

ARTICLE INFO

Received: 28 Dec 2024

Revised: 18 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

The scrutiny and identification of anomalies within the Border Gateway Protocol (BGP) stand as pivotal focal points in contemporary cybersecurity research. This paper navigates this intricate terrain, exploring diverse anomaly detection methodologies, including and historical-based analyses, and machine learning applications, all applied to comprehensive BGP datasets. Drawing from BGP update messages sourced from Reseaux IP Européens and Route Views, the study specifically investigates anomalies induced by the Moscow blackout.

The research unveils insights into the dynamic landscape of BGP anomalies, shedding light on the impact and characteristics of incidents caused by specific threats. Leveraging real-world datasets enhances the authenticity of the analysis, contributing to a nuanced understanding of the vulnerabilities within the BGP protocol. By the Moscow blackout, this paper offers a tangible and contextualized exploration of BGP anomalies, advancing our comprehension of cybersecurity threats and fortifications.

Furthermore, the paper proposes enhancements and solutions aimed at fortifying the BGP protocol against emerging threats. The evaluation and validation section critically assesses the proposed solutions, offering insights into their practical applicability and efficacy. The discussion section contextualizes the findings within the broader realm of cybersecurity, emphasizing the significance of proactive measures in mitigating potential risks. In conclusion, this research contributes valuable insights into the evolving landscape of cybersecurity, offering tangible enhancements to fortify BGP against emerging threats.

Keywords: BGP anomalies, Transmission Control Protocol (TCP),Border Gateway Protocol BGP, cybersecurity research, Moscow blackout

INTRODUCTION

The Border Gateway Protocol (BGP), a pivotal incremental path vector routing protocol, intricately manages network reachability among Internet autonomous systems (ASes) [1]. ASes, delineated collections of BGP routers, are uniquely identified by numbers allocated through regional Internet registries (RIRs) such as AFRINIC, ARIN, APNIC, LACNIC, and RIPE NCC [2]. This protocol heavily relies on the Transmission Control Protocol (TCP) for secure router-to-router communication, employing open, keepalive, update, and notification message types [3]. The consequential role of BGP update and withdrawal messages cannot be overstated, serving as essential components in conveying alterations in network topology and reachability. These messages contribute indispensable data for the analysis of Internet topology, the inference of AS relationships, and the evaluation of intrusion and anomaly detection mechanisms [2] [3]. Data acquisition is facilitated through BGP trace collectors like RIPE and Route Views, route servers, looking glasses, and Internet routing registries, often amalgamated for a more comprehensive representation of the Internet topology [4] [5] [6]. Despite its critical role, BGP is susceptible to anomalies that impede the successful exchange of reachability messages, resulting in a proliferation of anomalous update messages [2] [7]. Numerous proposed modifications seek to enhance BGP security [8] [9]. Anomalies encompass diverse incidents, ranging from worms (e.g., Slammer) and ransomware attacks (e.g., WannaCrypt) to routing misconfigurations, IP prefix hijacks, and link failures, including significant events such as the Moscow blackout [10] [11] [12].

This research focuses on the Moscow blackout. BGP update messages from data collection sites, including RIPE and Route Views, are extracted for classification, with a comparative evaluation of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) recurrent neural network (RNN) classification algorithms [13] [14] [15] [16]. The subsequent sections of this paper delineate the BGP data collection sites, detail BGP anomalies, describe extracted datasets, and outline the experimental procedure, concluding with a summary of findings and contributions [13] [14] [15] [16].

LITERATURE REVIEW

The intricate structure of the global internet, comprising a diverse array of interconnected administrative domains known as autonomous systems (ASes), relies on the orchestration of information exchange facilitated by routers within an AS through the Internal Gateway Protocol (IGP) and between routers in different ASes through the Exterior Gateway Protocol (EGP). At the forefront of EGPs stands the Border Gateway Protocol (BGP) [26], a pivotal protocol entrusted with the meticulous maintenance of updated information among routers pertaining to selected paths leading to specific routing prefixes.

In the realm of BGP, routers, colloquially referred to as BGP speakers, manage a threefold Routing Information Base (RIB), encompassing the Adj-RIB-In, Adj-RIB-Out, and Loc-RIB tables. The intricacies of these tables are profound, with Adj-RIB-In storing unedited routing information received from neighboring routers, Loc-RIB retaining optimal routes derived from internal routing policies, and Adj-RIB-Out storing routes designated for dissemination to neighboring routers. The BGP collection process unfolds through the operation of route collectors simulating routers, establishing BGP peering sessions with authentic routers, and meticulously collecting update messages reflecting alterations in the Adj-RIB-Out.

Initiatives such as Route Views [23] and RIPE RIS [22] play a pivotal role in enhancing accessibility to dumps, providing a comprehensive perspective on observable routing dynamics [25].

However, the envisaged utility of BGP, conceived as an instrument for the exchange of information indicative of genuine changes in inter-domain infrastructure, encounters empirical challenges. Scholarly inquiries [20] [21] into BGP behavior underscore its inherent dynamism, instability, and propensity for anomalous behavior. Large-scale anomalous behavior in BGP updates deviates notably from the anticipated normal distribution of BGP dynamics, transiently affecting multiple ASes before reverting to a standard operational state. The taxonomy proposed by Al-Musawi et al. [24] delineates BGP anomalies into four primary categories: direct intended, direct unintended, indirect, and network failures.

Direct intended anomalies, typified by BGP hijacking attacks, manifest when an assailant falsely asserts ownership of a prefix belonging to another AS, thereby redirecting routes to either intercept or discard traffic. Direct unintended anomalies stem from inadvertent misconfigurations that cause routers to announce anomalous prefixes or sub-prefixes. Notably, these anomalies tend to be of brief duration, given the shared interest among AS origin and affected peers in promptly rectifying such deviations.

Indirect anomalies encompass events not directly tethered to BGP or internet routing but wield significant influence over BGP dynamics and AS reachability. Common catalysts include attacks generating heightened traffic volumes, precipitating congestion and unresponsiveness in AS routers, with potential repercussions such as route flapping.

The spectrum of network failures embraces various entities, including routers, ASes, and operators, whose malfunctions can yield diverse impacts on the internet landscape [25].

The BGP domain has witnessed the emergence of various methodologies and systems dedicated to the identification of anomalies and the elucidation of their origins. Noteworthy contributions have probed the intricacies of anomaly detection within the BGP infrastructure. To facilitate a comprehensive and systematic comparative analysis, these approaches are systematically classified into 2 primary classes: time series analysis, machine learning

A. Approaches Grounded in Time Series Analysis

Pioneering efforts in time series analysis for BGP anomaly detection can be traced back to early endeavors, utilizing the Fast Fourier Transform (FFT) on routing update rates. Subsequent advancements expanded this work, incorporating five distinct BGP features to identify instability. The introduction of the Wavelet Transform inspired the creation of the BAlet framework, leveraging Daubchies5 (db5) Wavelet transform and Single-Linkage for clustering. Similarly, the BGP-lens applied the Haar Wavelet transform and median filtering. An approach capitalized on Recurrence Quantification Analysis (RQA), emphasizing the deterministic, recurrent, and non-linear characteristics of BGP updates[17].

B. Approaches Rooted in Machine Learning

Internet Routing Forensic (IRF) framework introduced a machine learning-based approach[18], employing the C4.5 algorithm to construct a decision tree for anomaly detection. A comparable framework presented, incorporating diverse data mining algorithms like decision trees, Naive Bayes, and Support Vector Machine (SVM). Another mechanism operated in two distinct phases: advanced feature extraction utilizing Fisher and mRMR scoring algorithms, and classification employing SVM and Hidden Markov Models (HMMs). Furthermore, a system leveraged prefix visibility and a machine learning winnowing algorithm.

PREPROCESSING

Before model training, a meticulous preprocessing phase was undertaken to ensure the data's quality and relevance. This included a strategic column renaming process, enhancing interpretability and facilitating a more intuitive understanding of the dataset. The class imbalance issue, a common challenge in threat detection datasets, was addressed using Synthetic Minority considered to ascertain the most effective approach for threat detection. The models included Support Vector Machine (SVM), Decision Tree, Random Forest, and XGBoost. Each model was chosen based on its suitability for binary classification tasks and its ability to handle both numerical and categorical features present in the dataset.

MODEL TRAINING AND EVALUATION

The dataset was split into training and testing sets to facilitate model training and subsequent evaluation. Feature scaling was applied to normalize the data, ensuring that each feature contributes uniformly to the model. Models were trained using the training set and evaluated on the testing set. Evaluation metrics such as accuracy, precision, recall, and F1 score were employed to gauge the performance of each model.

RESULTS ANALYSIS

The performance of each model was thoroughly analyzed, considering both overall accuracy and class-specific metrics. Visualization techniques, including bar plots and confusion matrices, were utilized to provide a comprehensive understanding of the models' strengths and weaknesses. Additionally, the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) were employed to assess the models' ability to discriminate between classes

METHODOLOGY

- Data Collection

The dataset used in this study is crucial for understanding the intricacies of the threat landscape. Acquired from IEEE DataPort website , it provides a comprehensive representation of features associated with potential security threats. The BGP datasets featuring five prominent Border Gateway Anomalies, namely WannaCrypt, Moscow blackout, Slammer, Nimda, and Code Red I, spanning the years 2001 to 2017. These datasets, procured from Reseaux IP Europeens (RIPE) BGP update messages, encapsulate both regular and anomalous data, with a specific emphasis on anomalies for robust threat detection and classification. The datasets, publicly accessible through the Network Coordination Centre (NCC), serve as a comprehensive repository of information, offering detailed features derived from BGP update messages. Encompassing parameters such as the number of announcements, withdrawals, various path length metrics, and packet-level details, these key features provide a nuanced comprehension of routing dynamics during anomalous occurrences.

The target variable is a pivotal component, discerning between regular data (labeled as -1) and anomalous data (labeled as 1). This dichotomy establishes the groundwork for supervised learning methodologies. The datasets, meticulously organized based on collection date, adhere to the multi-threaded routing toolkit (MRT) format, with extraction facilitated through a Perl script. Additionally, the inclusion of information regarding the date of last modification and dataset sizes augments the contextual understanding of the dataset characteristics.

This diversity in data allows for a holistic exploration of threat detection methodologie

RESULTS

- **Threat Detection and Classification**
- Support Vector Machine (SVM)

The Support Vector Machine (SVM) model showcased robust performance in threat detection, achieving an impressive accuracy of 96%. Precision, recall, and F1 score were equally notable at 96.58%, 95.81%, and 96.19%, respectively. This suggests the SVM model's efficacy in identifying and classifying threats within the datas

- Decision Tree

The Decision Tree model demonstrated superior accuracy, reaching 98%. Precision, recall, and F1 score were 96.73%, 98.34%, and 97.53%, respectively, indicating its advanced threat classification capabilities. Decision Trees proved adept at capturing complex relationships within the data.

- Random Forest

The Random Forest model outperformed others with an accuracy of 99%. Precision, recall, and F1 score were exceptional at 99.56%, 98.99%, and 99.28%, respectively. This emphasizes the Random Forest's effectiveness in detecting threats, showcasing its ability to provide robust classification.

- XGBoost

Similar to Random Forest, the XGBoost model achieved an accuracy of 99%. Precision, recall, and F1 score were also high at 99.28%, 99.57%, and 99.42%, underscoring its robustness and reliability in threat classification.

ENHANCEMENTS AND SOLUTIONS

The incorporation of Synthetic Minority Over-sampling Technique (SMOTE) provided a notable solution to the class imbalance issue within the dataset. By oversampling the minority class, SMOTE ensured a balanced representation, contributing significantly to improved model performance security challenges in various domains.

REFRENCES

[1] H. Yan, R. Oliveira, K. Burnett, D. Matthews, L. Zhang, and D. Massey, "BGPmon: a real-time, scalable, extensible monitoring system," in Proc. Cybersecurity Appl. Technol. Conf. Homeland Secur., Washington, DC, USA, Mar. 2009, pp. 212–223.

[2] B. Al-Musawi, P. Branch, and G. Armitage, "BGP anomaly detection techniques: a survey," IEEE Commun. Surv. Tut., vol. 19, no. 1, pp. 377–396, 2017.

[3] K. Sriram, O. Borchert, O. Kim, P. Gleichmann, and D. Montgomery, "A comparative analysis of BGP anomaly detection and robustness algorithms," in Proc. Cybersecurity Appl. Technol. Conf. Homeland Secur., Washington, DC, USA, Mar. 2009, pp. 25–38.

[4] RIPE NCC. (2020, Sept.). [Online]. Available: https://www.ripe.net

[5] University of Oregon Route Views project. (2020, Sept.). [Online]. Available: http://www.routeviews.org

[6] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS level topology," ACM Computer Communication Review (CCR), vol. 35, no. 1, pp. 53–62, Jan. 2005.

[7] Y. Song, A. Venkataramani, and L. Gao, "Identifying and addressing reachability and policy attacks in 'secure' BGP," IEEE/ACM Trans. Netw., vol. 24, no. 5, pp. 2969–2982, Oct. 2016.

[8] D. Dolev, S. Jamin, O. Mokryn, and Y. Shavitt, "Internet resiliency to attacks and failures under BGP policy routing," Comput. Netw., vol. 50, no. 16, pp. 3183–3196, Nov. 2006.

[9] A. Lutu, M. Bagnulo, C. Pelsser, O. Maennel, and J. Cid-Sueiro, "The BGP visibility toolkit: detecting anomalous Internet routing behavior," IEEE/ACM Trans. Netw., vol. 24, no. 2, pp. 1237–1250, Apr. 2016.

[10] R. Mahajan, D. Wetherall, and T. Anderson, "Understanding BGP misconfiguration," in Proc. 2002 Conf. Appl. Technologies Architectures Protocols Comput. Commun., Pittsburgh, Pennsylvania, USA, Aug. 2002, pp. 3–16.

[11] C. Testart, P. Richter, A. King, A. Dainotti, and D. Clark, "Profiling BGP serial hijackers: capturing persistent misbehavior in the global routing table," in Proc. ACM Internet Meas. Conf., Amsterdam, Netherlands, Oct. 2019, pp. 420–434.

[12] J. L. Sobrinho and T. Quelhas, "A theory for the connectivity discovered by routing protocols," IEEE/ACM Trans. Netw., vol. 20, no. 3, pp. 677–689, June 2012.

[13] Q. Ding, Z. Li, S. Haeri, and L. Trajkovic, "Application of machine ´ learning techniques to detecting anomalies in communication networks," in Cyber Threat Intelligence, A. Dehghantanha, M. Conti, and T. Dar gahi, Eds. Berlin: Springer, 2018, pp. 47–70 and pp. 71–92.

[14] Z. Li, A. L. Gonzalez Rios, G. Xu, and Lj. Trajkovic, "Machine learning ´ techniques for classifying network anomalies and intrusions," in Proc. IEEE Int. Symp. Circuits Syst., Sapporo, Japan, May 2019, pp. 1–5.

[15] C. Cortes and V. Vapnik, "Support-vector networks," J. Mach. Learn., vol. 20, no. 3, pp. 273–297, Sept. 1995.

[16] C. L. P. Chen and Z. Liu, "Broad learning system: an effective and efficient incremental learning system without the need for deep archi tecture," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 1, pp. 10–24, Jan. 2018.

[17] Fonseca, P., Mota, E. S., Bennesby, R., & Passito, A. (2019, June). Bgp dataset generation and feature extraction for anomaly detection. In 2019 IEEE Symposium on Computers and Communications (ISCC) (pp. 1-6). IEEE.

[18] Li, Z., Rios, A. L. G., Xu, G., & Trajković, L. (2019, May). Machine learning techniques for classifying network anomalies and intrusions. In 2019 IEEE international symposium on circuits and systems (ISCAS) (pp. 1-5). IEEE.

[19] Y. Rekhter, T. Li, and S. Hares, "A border gateway protocol 4 (bgp-4)," Internet Requests for Comments, RFC Editor, RFC 4271, January 2006, http://www.rfc-editor.org/rfc/rfc4271.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4271.txt

[20] C. Labovitz, G. R. Malan, and F. Jahanian, "Internet routing instability," in Proceedings of the ACM SIGCOMM '97 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, ser. SIGCOMM '97. New York, NY, USA: ACM, 1997, pp. 115–126. [Online]. Available: http://doi.acm.org/10.1145/263105.263151

[21] J. Li, M. Guidero, Z. Wu, E. Purpus, and T. Ehrenkranz, "Bgp routing dynamics revisited," SIGCOMM Comput. Commun. Rev., vol. 37, no. 2, pp. 5–16, Mar. 2007. [Online]. Available: http://doi.acm.org/10.1145/1232919.1232921

[22] RIPE, "Ripe network coordination centre," [Online]. Available: https://www.ripe.net/analyse/internet-measurements/routing information-service-ris. org, 1999, [Online; accessed 25-August-2018]

[23] O. RouteViews, "University of oregon routeviews project," Eugene, OR. [Online]. Available: http://www.routeviews.org, 2013, [Online; ac cessed 25-August-2018].

[24] B. Al-Musawi, P. Branch, and G. Armitage, "Bgp anomaly detection techniques: A survey," IEEE Communications Surveys and Tutorials, vol. 19, no. 1, pp. 377–396, Firstquarter 2017

[25] C. Orsini, A. King, D. Giordano, V. Giotsas, and A. Dainotti, "Bgp stream: a software framework for live and historical bgp data analysis," in Proceedings of the 2016 Internet Measurement Conference. ACM, 2016, pp. 429–444.

[26] G. Aceto, A. Botta, P. Marchetta, V. Persico, and A. Pescap, "A comprehensive survey on internet outages," Journal of Network and Computer Applications, vol. 113, pp. 36 – 63, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804518301139

[27] [14] A. Kausar, A. Jamil, N. Nida, and M. H. Yousaf, "Two-wheeled vehicle detection using two-step and single-step deep learning models," *Arabian Journal for Science and Engineering*, vol. 45, pp. 10755-10773, 2020.

[28]

[29] [15] S. Vasavi, N. K. Priyadarshini, and K. Harshavaradhan, "Invariant feature-based darknet architecture for moving object classification," *IEEE Sensors Journal*, vol. 21, pp. 11417-11426, 2020.

[30]

[31] [16] Q. Li, S. Garg, J. Nie, X. Li, R. W. Liu, Z. Cao, and M. S. Hossain, "A highly efficient vehicle taillight detection approach based on deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 4716-4726, 2020.

[32]

[33] [17] A. A. Qazzaz and A. Y. Abdulkadhim, "Car Detection and Features Identification Based on YOLOV5," *International Journal of Mechanical Engineering*, vol. ISSN 0974-5823, 2022.

[34]

[35] [18] D. L. Nguyen, X. T. Vo, A. Priadana, and K. H. Jo, "Car Detector Based on YOLOv5 for Parking Management," in *Conference on Information Technology and its Applications*, Cham: Springer Nature Switzerland, 2023, pp. 102-113.