BGP Anomaly Detection: From Feature Engineering to Synthetic Traffic Generation

Shadi Motaali, Jorge E. López de Vergara, Luis de Pedro
Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain
{shadi.motaal, jorge.lopez_vergara@uam.es, luis.depedro}@uam.es

***Abstract*—Border Gateway Protocol (BGP) underpins inter-domain routing but remains vulnerable to anomalies such as prefix hijacks and route leaks. Machine learning offers promising detection capabilities to deal with them. However, it faces two barriers: the lack of interpretable feature representations and the scarcity of labeled data. This Ph.D. research addresses both through a four-phase approach. Phase 1 developed a hybrid feature selection framework that reduces 48 features to 25 (47.9% reduction) while achieving 89.45% accuracy with XGBoost and SHAP-based explainability. Phase 2 built an RFC-compliant Scapy-based traffic generator producing labeled BGP packets. Phase 3 benchmarked thirteen generators across five families using a protocol-aware fidelity framework, finding that SMOTE_kmeans achieves 97.5% in-distribution fidelity but degrades significantly under cross-collector shift (F1 dropping to 0.79). Phase 4 (ongoing) addresses this generalization gap by integrating graph-based features from RIB snapshots with domain adaptation for real-time, topology-aware detection.**

***Index Terms*—BGP security, anomaly detection, machine learning, synthetic data generation, feature selection, explainability**

I. Introduction

The Border Gateway Protocol (BGP) governs routing between over 70 000 autonomous systems (ASes) on the Internet. However, BGP original design relies on implicit trust between peers, lacking cryptographic validation in most deployments. This vulnerability has led to numerous incidents—the 2008 YouTube hijack, recurring route leaks, and state-level traffic interception—demonstrating that BGP security remains a critical operational challenge [1].

While cryptographic defenses such as RPKI [2] and BGPsec [3] provide origin and path validation, RPKI deployment covers only approximately 60% of announced IPv4 prefix-origin pairs [4], and BGPsec adoption remains negligible due to computational overhead. Consequently, Machine Learning (ML)-based anomaly detection through traffic analysis remains essential for operational security [5], [6].

However, deploying ML for BGP anomaly detection faces two intertwined challenges. First, the *feature engineering challenge*: BGP UPDATE streams generate high-dimensional feature spaces, where identifying the most discriminative attributes requires careful analysis across diverse anomaly types. Second, the *data scarcity challenge*: public collectors such as RIPE RIS [7] and RouteViews [8] provide raw UPDATE streams at massive scale, but ground-truth labeling remains incident-driven, expert-dependent, and severely imbalanced (anomalies typically represent less than 5% of traffic).

This Ph.D. research tackles both challenges through a systematic, four-phase approach:

1) *Phase 1—Feature Selection and Explainability* [9]: A hybrid feature selection framework combining six algorithms to identify the optimal feature subset, coupled with SHAP-based and Gini-index explainability analysis to understand model decisions.
2) *Phase 2—Synthetic Traffic Generation* [10]: An RFC-compliant BGP traffic generator using Scapy that produced protocol-valid packets with precise ground-truth labels.
3) *Phase 3—Generator Benchmarking* [11]: A systematic evaluation of thirteen generators from five families using a protocol-aware, 16-metric fidelity framework, including cross-collector generalization analysis.
4) *Phase 4—Graph Features, Real-Time Detection, and Domain Adaptation (ongoing)*: Incorporating Routing Information Base (RIB)-derived graph features, dual-stack IPv4/IPv6 support, and streaming-based inference for topology-aware AS-level anomaly detection that generalizes across collectors.

This bottom-up progression—from features to generation, evaluation and deployment—is detailed as follows. Section II reviews the state of the art. Section III identifies research gaps. Section IV presents the research questions. Section V details the methodology and progress. Section VI concludes and poses two questions for future discussion.

II. State-of-the-Art Review

Table I compares representative works across four key dimensions: use of real BGP data, statistical feature engineering, synthetic data generation, and graph-based feature extraction. Beyond these, our work provides hybrid feature selection with SHAP-based explainability (Phase 1), protocol-compliant generation validated against RFC 4271 (Phase 2), and cross-collector generalization evaluation (Phase 3). Notably, while Hoarau et al. [17] extract both statistical and graph features, no prior work has systematically analyzed their correlation,

This work is partially funded by a grant from the Dept. of Electronics and Communication Technologies at UAM, as well as by the R&D activity program with ref. TEC-2024/COM-504 and acronym RAMONES-CM, granted by the Comunidad de Madrid, Spain, through the Directorate General for Research and Technological Innovation via Order 5696/2024.

TABLE I
COMPARISON OF RELATED WORK. ✓ = SUPPORTED, — = **NOT ADDRESSED (GAP)**.

Work	Real BGP Data	Statistical Features	Synthetic Generation	Graph Features
Ammara et al. [12]	—[a]	✓[a]	✓	—
Nassir et al. [13]	—[b]	✓	—	—
Romo-Chavero et al. [14]	✓	✓	✓[c]	—
Allahdadi et al. [15]	✓	✓	—	—
Park et al. [16]	✓	✓	✓[c]	—
Hoarau et al. [17]	✓	✓	—	✓
Liu et al. [18]	✓	—	—	✓
Our Work (Ph. 1–4)	✓	✓	✓	✓[d]

[a]Evaluated on IDS datasets (NSL-KDD, CIC-IDS2017), not BGP-specific data.
[b]Uses simulated data; applicability to operational networks is limited.
[c]Uses SMOTE oversampling without protocol-aware post-processing for discrete BGP features.
[d]Phase 4 (ongoing): graph feature engineering from RIB snapshots.

complementarity, or optimal integration strategy for BGP anomaly detection.

III. GAPS AND LIMITATIONS

Our literature review reveals the following gaps:

Gap 1: Existing BGP anomaly detection studies use either full feature sets or ad-hoc feature selection, without systematic hybrid approaches that combine multiple selection algorithms. Furthermore, model explainability for BGP-specific decisions remains underexplored.

Gap 2: Synthetic BGP traffic generators either operate at the packet level without learning from data (Scapy, ns-3) or learn statistical patterns without enforcing protocol constraints (GANs, SMOTE). No study has systematically compared these approaches using protocol-aware evaluation metrics.

Gap 3: Cross-collector and cross-incident generalization—critical for real-world deployment—has not been evaluated for synthetic BGP data. Existing benchmarks focus on in-distribution performance.

Gap 4: Existing approaches treat statistical features (from UPDATE streams) and graph-based features (from AS-path trees, reachability graphs, neighbor connectivity) as independent modalities. No prior work has systematically analyzed their correlation structure and complementarity for BGP anomaly detection, nor has any study systematically combined them into a unified representation, despite BGP inherently graph-structured nature.

IV. RESEARCH QUESTIONS

Based on the identified gaps, this PhD addresses four research questions:

RQ1: How can a hybrid feature selection framework improve BGP anomaly detection accuracy while providing interpretable explanations for model decisions?

RQ2: Can synthetic BGP traffic generated through RFC-compliant packet construction and statistical methods effectively augment real datasets for ML-based anomaly detection?

RQ3: How do different families of synthetic data generators (rule-based, deep generative, oversampling, hybrid, statistical) compare in terms of fidelity, downstream detection utility, and cross-collector generalization?

RQ4: Can combining graph-based topological features from RIB snapshots with statistical UPDATE features, along with domain adaptation techniques, enable real-time AS-level anomaly detection that generalizes across heterogeneous BGP vantage points?

V. METHODOLOGY AND PROGRESS

Figure 1 summarizes the following phases.

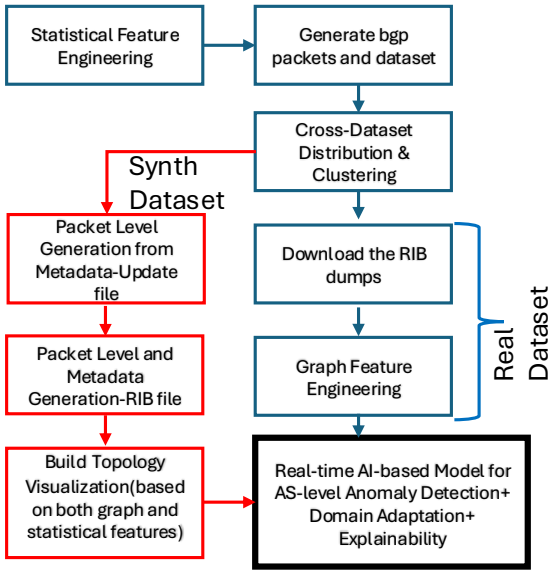

Fig. 1. Methodology steps

A. Phase 1: Hybrid Feature Selection and Explainability

We developed a hybrid framework combining six algorithms (ANOVA F-score, Mutual Information, Random Forest/XGBoost Importance, RFE, Lasso) [9]. Statistical features appearing in the top 30 of at least 75% of algorithms are retained, yielding 25 features from 48 (47.9% reduction). To

address dataset imbalance (95% normal), we adopted real-sample balancing anchored to the smallest class (1,217 instances) instead of SMOTE, which generates unrealistic floats for discrete BGP features.

XGBoost achieved 89.45% accuracy (89.72% F1 after Bayesian tuning). SHAP and Gini importance showed consistent top features (`edit_distance_dict_1`, `dups`, `nlri_ann`). Inference speedup: 36.9%. Ablation confirmed top-10 features yield ≈85% accuracy; removing `dups` causes 2.1% F1 drop.

Status: Published at ICCIDA 2025 [9].

B. Phase 2: RFC-Compliant Synthetic Traffic Generation

Building on Phase 1 finding about SMOTE limitations with discrete features, we developed a Scapy-based BGP traffic generator [10] producing protocol-compliant sessions (OPEN, UPDATE, KEEPALIVE, NOTIFICATION) with full RFC 4271 compliance and IPv4/IPv6 support via MP-BGP. The framework generates three anomaly categories—prefix hijacking, path manipulation, and UPDATE flooding—with process-ID-based labeling providing perfect ground-truth labels. Inter-arrival times follow Pareto distributions ($\alpha = 2.5$ normal, $\alpha = 1.8$ attacks). During development, we identified and corrected a Scapy KEEPALIVE bug producing 38-byte packets instead of the RFC-mandated 19 bytes. Validation through GNS3, Cisco traces, and bidirectional PCAP-to-Scapy conversion confirmed byte-level correctness. A proof-of-concept dataset of 45 928 labeled UPDATEs achieved 96.88% accuracy with Random Forest.

Status: Accepted at IEEE/IFIP NOMS 2026 [10].

C. Phase 3: Systematic Generator Benchmarking

Phase 2 revealed that Scapy does not capture correlated feature patterns from network-wide routing dynamics, motivating a comparison of thirteen generators from five families [11]. We developed a consensus-based labeling pipeline combining five unsupervised detectors on raw RIPE RIS streams (achieving a silhouette score of 0.965). All generators operate on a normalized feature space (derived from a separate feature extraction pipeline [9], [19]). To ensure SMOTE produces valid BGP values for discrete features, we implement: (i) log1p transformation before generation, (ii) inverse transformation and valid-range clipping (≥ 0) after generation, and (iii) integer rounding for routing-count dimensions.

Key findings: SMOTE_kmeans achieves 97.5/100 in-distribution fidelity and F1 > 0.99 under matched conditions. Cross-collector evaluation (RRC05) reveals degradation: F1 drops to 0.79, FNR = 0.35–0.41. Copula shows better cross-dataset stability (42.4% vs. 33.6%). Neural generators overfit collector-specific patterns (20–52/100). Scapy scored lowest (19.8–29.6/100) despite RFC compliance.

Status: Submitted to EuCNC/6G Summit 2026 [11].

D. Phase 4: Graph Features, Real-Time Detection, and Domain Adaptation (Ongoing)

The cross-collector degradation observed in Phase 3 (FNR = 0.35–0.41) and the broader thesis objective of improving the quality and security of future network technologies [20] motivate the current phase. The core hypothesis is that combining topological information from RIB snapshots with statistical UPDATE features will yield collector-invariant representations that generalize across vantage points. The phase is structured along two parallel tracks that converge into a unified detection model.

Track A—Synthetic enrichment. We extend the Phase 3 generators to produce packet level traffic from real metadata-UPDATE files and RIB snapshots, thereby inheriting realistic AS-path structure, prefix distributions, and peer diversity from actual routing tables. Rather than generating synthetic attributes independently, this approach conditions packet generation on observed RIB state, ensuring that AS-path lengths, origin ASes, and community attributes reflect real topological constraints. Leveraging our protocol-agnostic finding [20]—that common statistical features detect anomalies effectively across both IPv4 and IPv6—the extension to dual-stack synthesis adapts existing methods rather than rebuilding from scratch.

Track B—Graph feature engineering. We download full RIB dumps from multiple RIPE RIS collectors (captured at 8-hour intervals) and construct AS-level graphs for each snapshot. From these graphs we extract topological features in four categories: (i) *structural*—AS connectivity degree, clustering coefficient, betweenness centrality; (ii) *path-based*—AS-path tree depth, path diversity (number of distinct paths per prefix), maximum and mean path length; (iii) *temporal*—neighbor stability, new/withdrawn peer count; and (iv) *reachability*—prefix reachability ratio, origin AS concentration. These features complement the statistical UPDATE features from Phases 1–3 by capturing structural properties that volume-based metrics cannot represent: a prefix hijack, for instance, may introduce only a small UPDATE volume spike but drastically alter AS-path tree structure and origin concentration.

Feature integration and domain adaptation. A key research question is whether statistical and graph features are complementary or redundant (Gap 4). We will quantify cross-correlation (Pearson, Spearman, mutual information) between feature families and systematically evaluate three integration strategies: (i) *early fusion*—concatenating both feature vectors; (ii) *multi-view learning*—separate encoders per modality with a shared classifier; and (iii) *GNN-based fusion*—encoding AS topology as a graph with node features combining both statistical and structural attributes. To address the 25–30% cross-collector accuracy gap, domain adaptation techniques will train the model to learn collector-invariant representations, evaluated on held-out collectors not seen during training.

Evaluation plan. The integrated model will be evaluated on: (i) in-distribution detection (comparison with Phase 1–3 baselines); (ii) cross-collector generalization; (iii) detection latency under realistic UPDATE rates; and (iv) explainability quality via case studies on documented BGP incidents (e.g., the 2025 Iberian Peninsula outage).

Status: In progress; RIB download pipeline operational.

E. Discussion

Two key insights emerged across the four phases. First, Phase 1 initially concluded that SMOTE is unsuitable for discrete BGP features, but Phase 3 demonstrated that with protocol-aware pre/post-processing (log1p/expm1 transformation, valid-range clipping, integer rounding), SMOTE_kmeans achieves 97.5/100 fidelity—showing the limitation lies in unconstrained application, not in SMOTE itself. Second, our preliminary analysis [20] revealed that BGP anomalies are detectable using common statistical features across both IPv4 and IPv6, with protocol-specific attributes (NDP, extended communities) serving as optional enhancers. This *protocol-agnostic* finding validates that our 25-feature subset generalizes to IPv6 without re-engineering and enables a single model for both protocol families.

Despite these advances, three open challenges remain: (i) cross-collector generalization—even the best generator degrades by 25–30% under distribution shift (FNR = 0.35–0.41), as no single feature representation generalizes across vantage points; (ii) our consensus-based labeling relies on unsupervised detectors without ground-truth validation from network operators; and (iii) the relationship between statistical and graph-based features remains unexplored, leaving potential complementarity or redundancy unquantified.

VI. Conclusion and Further Research

This Ph.D. research, part of the broader thesis "Improving the quality and security of future network technologies," addresses feature engineering and data scarcity challenges in ML-based BGP anomaly detection. Phases 1–3 have produced so far: (1) a hybrid feature selection framework reducing dimensionality by 47.9% with SHAP-based explainability; (2) an RFC-compliant dual-stack (IPv4/IPv6) synthetic traffic generator; and (3) the first systematic cross-collector benchmark of thirteen BGP generators, revealing synthesis limitations under distribution shift (F1 degradation from 0.99 to 0.79). Phase 4 (ongoing) incorporates RIB-derived graph features, domain adaptation, and real-time streaming inference to close the cross-collector gap. Longer-term extensions include EVPN/VXLAN overlay networks and integration with SDN/NFV frameworks for automated anomaly mitigation.

Questions for Future Work:

1) *From statistical to structural features*: Phases 1–3 rely on selected statistical features from UPDATE streams, yet BGP is inherently graph-structured. How should graph-based features from RIB snapshots (AS connectivity, path diversity, reachability topology) be combined with statistical features—through concatenation, multi-view learning, or GNNs—to maximize detection while preserving explainability?
2) *Real-time architecture*: Given that RIB snapshots provide 8-hour topology snapshots while BGP anomalies propagate in seconds, what is the optimal architecture for combining slow-changing graph features with fast-updating statistical features—should we use incremental graph updates, dual-speed feature pipelines, or cached topological embeddings refreshed periodically?

References

[1] D. Madory, "A brief history of the Internet's biggest BGP incidents," *Kentik Blog*, June 2023. [Online]. Available: https://www.kentik.com/blog/a-brief-history-of-the-internets-biggest-bgp-incidents/

[2] M. Lepinski and S. Kent, "RFC 6480: An infrastructure to support secure internet routing," IETF RFC 6480, 2012.

[3] M. Lepinski and K. Sriram, "BGPsec Protocol Specification," Internet Engineering Task Force, RFC 8205, Sep. 2017. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8205

[4] National Institute of Standards and Technology, "NIST RPKI monitor," https://rpki-monitor.antd.nist.gov/, 2025.

[5] P. Edwards, L. Cheng, and G. Kadam, "Border gateway protocol anomaly detection using machine learning techniques," *SMU Data Science Review*, vol. 2, no. 1, 2019.

[6] K. Hoarau, P. U. Tournoux, and T. Razaindralambo, "Unsupervised representation learning for BGP anomaly detection," *ITU Journal on Future and Evolving Technologies*, March 2024.

[7] RIPE Network Coordination Centre, "RIPE routing information service (RIS)," https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris.

[8] U. of Oregon, "Routeviews project," http://www.routeviews.org/.

[9] S. Motaali, J. E. López de Vergara, and L. de Pedro, "Hybrid feature selection and explainable machine learning for bgp anomaly detection," in *Proc. 4th International Conference on Computing, IoT and Data Analytics*, Madrid, Spain, 2025.

[10] S. Motaali, J. E. López de Vergara, L. de Pedro, and I. González, "Generating balanced and realistic BGP traffic for machine learning-based anomaly detection," in *IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 2026.

[11] ——, "SynthBGP: Synthetic BGP traffic generation for enhanced cybersecurity anomaly detection," in *EuCNC/6G Summit, 2026*, 2026, submitted.

[12] D. A. Ammara, J. Ding, and K. Tutschku, "Synthetic network traffic data generation: A comparative study," *arXiv preprint arXiv:2410.16326*, February 2025.

[13] N. S. Kadhim, N. F. Abdullah, and K. Chellappan, "Hybrid machine learning algorithm for enhanced BGP anomaly detection," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 24, no. 11, pp. 1–12, 2024.

[14] M. A. Romo-Chavero, G. de los Ríos Alatorre, J. A. Cantoral-Ceballos, J. A. Pérez-Díaz, and C. Martinez-Cagnazzo, "A hybrid model for BGP anomaly detection using median absolute deviation and machine learning," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 2102–2115, 2025.

[15] A. Allahdadi, R. Morla, and R. Prior, "A framework for BGP abnormal events detection," in *2017 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*. IEEE, 2017.

[16] H. Park, K. Kim, D. Shin, and D. Shin, "BGP dataset-based malicious user activity detection using machine learning," *Information*, vol. 14, no. 9, p. 501, 2023.

[17] K. Hoarau, P. U. Tournoux, and T. Razafindralambo, "BML: An efficient and versatile tool for BGP dataset collection," in *IEEE Int. Conf. Commun. Wkshps. (ICC Wkshps.)*, 6 2021.

[18] Z. Liu, H. Qiu, R. Wang, J. Zhu, and Q. Wang, "Detecting BGP anomalies based on spatio-temporal feature representation model for autonomous systems," *IEEE 22nd Int. Conf. Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2023.

[19] P. Fonseca, E. S. Mota, R. Bennesby, and A. Passito, "BGP dataset generation and feature extraction for anomaly detection," in *2019 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2019.

[20] S. Motaali, J. E. López de Vergara, and L. de Pedro, "Real-time anomaly detection in BGP: Challenges, IPv6 considerations, and machine learning opportunities," in *IEEE 11th International Conference on Network Softwarization (NetSoft)*, 2025.