Detection of BGP Routing Leaks Using Historical Baseline Profiling

Rifqy Hakimi*‡ and Martin J. Reed*
*School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom
‡School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia
*Email: {rh21215, mjreed}@essex.ac.uk

***Abstract*—Border Gateway Protocol (BGP) prefix hijacking and route leaks (RFC 7908 Type 5 incidents) pose critical threats to inter-domain routing stability. Despite expanding RPKI deployment reaching 56% IPv4 coverage by July 2025 and minimal BGPsec/Autonomous System Provider Authorization (ASPA) adoption, such large-scale events continue to bypass existing cryptographic protections.**

We present a dual-signal detection framework addressing complementary manifestations of Type 5 incidents. Our approach builds stable baselines from Routing Information Base (RIB) snapshots to generate Baseline-Deviation (BD) signals that identify incidents through origin AS deviations on known prefixes, and New-Prefix (NP) signals that identify incidents through previously unseen address space. These signals are aggregated across multiple vantage points using MAX aggregation, scored with Z-score thresholds, and fused via logical disjunction.

Cross-incident validation on six diverse real-world Type 5 incidents achieved F1-scores ranging from 0.48 to 0.88 (mean 0.69) on entirely unseen incidents, demonstrating that effective detection requires both signal types to capture the full spectrum of manifestation patterns. The lightweight approach requires no machine learning, enabling rapid deployment in resource-constrained NOC environments and improving interpretability for incident response.

***Index Terms*—BGP anomaly detection, prefix hijacking, route leak detection, multi-vantage analysis, baseline profiling**

I. Introduction

Border Gateway Protocol (BGP) is the foundation of inter-domain routing on the global Internet, yet it lacks built-in security mechanisms. This vulnerability enables serious routing anomalies, including route leaks and prefix hijacks that can misdirect traffic, cause service outages, and undermine Internet stability. These incidents range from accidental misconfigurations involving legitimate prefixes to deliberate hijacking attacks using previously unseen address space.

Cryptographic protection deployment remains incomplete. By July 2025, only 56% of IPv4 prefixes had RPKI Route Origin Authorisations (ROAs), leaving nearly half without origin validation [1]. Path validation mechanisms like BGPsec and Autonomous System Provider Authorization (ASPA) remain experimental with low adoption [2], [3], allowing large-scale anomalies to bypass existing protections.

These challenges underscore the need for complementary detection methods capable of operating effectively in partially secured routing environments. Most existing BGP anomaly detection approaches focus on either route leaks or prefix hijacking, not both, and either require expensive machine learning (ML) infrastructure or extensive manual tuning. Moreover, many methods rely on single monitoring vantage points (VPs) or coarse time-window labelling, limiting operational utility and generalisation across diverse incidents.

To address this gap, we propose: Can a lightweight, interpretable framework detect both route-leak and prefix-hijacking anomalies with fine temporal precision across diverse incidents? This work presents a dual-signal detection framework combining *Baseline-Deviation (BD)* signals for changes to known prefix-origin mappings and *New-Prefix (NP)* signals for previously unseen prefixes. Signals are aggregated across multiple VPs using MAX aggregation, scored with Z-score thresholds, and fused through logical disjunction. The approach requires no ML, facilitating operational deployment in resource-constrained environments.

The main contributions of this paper are:

1) We develop a dual-signal detection framework combining BD and NP indicators to capture complementary manifestations of RFC 7908 Type 5 prefix hijacking, addressing single-signal approach limitations. BD signals identify incidents through origin AS deviations on known prefixes, whilst NP signals capture incidents through previously unseen prefix announcements.
2) We establish rigorous per-update ground truth construction methodology using expert-verified incident reports, enabling precise evaluation reflecting operational detection requirements versus existing approaches using arbitrary time-window labelling.
3) We demonstrate effective multi-perspective anomaly detection through MAX aggregation across multiple VPs, preserving localised anomalies whilst providing noise reduction via statistical thresholding and fusion.
4) We conduct comprehensive cross-incident validation using leave-one-out methodology on six diverse real-world Type 5 incidents, achieving F1-scores ranging from 0.48 to 0.88 (mean 0.69), demonstrating practical generalisation across diverse incident manifestation patterns.

The remainder of this paper is organised as follows. Section II reviews related work. Section III describes our proposed framework. Section IV presents experimental results. Section V discusses implications and limitations. Section VI

concludes the paper.

II. Literature Review

Inter-domain routing relies on BGP, yet its design leaves it susceptible to Type 5 route leaks [4], including prefix hijacking events that can lead to widespread network disruption, making detection essential for Internet stability.

Early BGP anomaly detection approaches relied on single-point monitoring with limited scope. PHAS [5] established foundational real-time prefix hijack alerting, whilst iSPY [6] pioneered self-operated detection using data-plane probing. Ballani et al. [7] conducted comprehensive empirical analysis of prefix hijacking and interception, demonstrating global traffic redirection patterns. However, these approaches were constrained by single-point limitations or notification-only capabilities.

As the Internet matured, researchers recognised that multi-vantage aggregation was necessary for comprehensive detection. Statistical analysis [8] revealed BGP protocol weaknesses during worm attacks, whilst Mai et al. [9] demonstrated that wavelet transforms effectively capture short-lived abnormal patterns traditional methods miss. ARTEMIS [10] achieved significant advances with real-time detection and automated mitigation within one minute. Scott et al. [11] applied Matrix Profile data mining for parameter-light detection across BGP event categories. More recently, Scott et al. [12] reframed the challenge as group dynamics, applying Multidimensional Recurrence Quantification Analysis to capture collective Autonomous System (AS) behaviour patterns.

Recent advances exploit baselines built from RIB snapshots, maintaining stable prefix-origin pairs to flag deviations reliably. Signal fusion techniques enhance coverage without excessive computational overhead. However, many systems require extensive configuration or ML components, limiting practical deployment [13]. ML approaches increasingly enhance BGP anomaly detection. Recent surveys [13], [14] outline promising results alongside explainability and resource challenges. Multi-scale LSTM architectures [15] demonstrate effective BGP traffic classification, whilst graph neural networks such as BGNN [16] show strong performance in large-scale anomaly detection. However, operational deployment remains limited due to computational complexity. Efforts addressing evolving challenges include forged-origin attacks [17] and MOAS detection [18], underscoring growing attack complexity. Our previous work [19] showed that careful feature selection improves accuracy and reduces latency.

In summary, BGP anomaly detection has evolved from single-point solutions to multi-vantage strategies enhanced by ML. However, partial security deployment and operational constraints maintain the need for practical detection techniques providing semantic classification of anomaly types. This paper builds on these advances with a resource-efficient solution designed for real-world deployment.

III. Proposed Method

We present a per-message anomaly detection framework that analyses raw BGP updates from multiple VPs against dynamic prefix-origin AS baselines. Each update is classified as *benign*, *anomalous*, or *no_baseline*, producing two anomaly signal types: BD (deviation from stable mapping) and NP (prefix absent from any baseline). Unlike prior work that labels anomalies over coarse time intervals, we employ per-update, expert-verified ground truth from trusted sources, enabling fine-grained evaluation. Signals are aggregated across VPs using the MAX function, analysed with Z-score detection, and combined via logical disjunction to produce final detection decisions.

A. Data Acquisition

Our framework operates on two primary data sources: periodic RIB snapshots and continuous BGP update streams, both collected from multiple VPs. RIB snapshots provide complete routing table views, enabling construction of stable and unstable baselines for prefix-origin AS mappings. BGP updates capture incremental routing changes between snapshots, including announcements, withdrawals, and modifications.

We utilised data from all available RIPE RIS Remote Route Collectors (RRCs) during each analysed incident, covering diverse VPs distributed across multiple continents [20]. This ensures the dataset reflects widest possible visibility of global routing dynamics, capturing anomalies observable from specific geographic or topological locations. RIB snapshots were collected every 8 hours, whilst BGP updates were ingested with per-message granularity for precise classification.

B. Ground Truth Construction

Accurate ground truth is essential for evaluating BGP anomaly detection. Most prior studies aggregate BGP updates into coarse time windows and assign anomaly labels to all updates within reported incident intervals. This approach treats all updates inside the interval as anomalous regardless of actual semantics, introducing noise into evaluation.

We construct ground truth at the per-update level, focusing on confirmed RFC 7908 Type 5 incidents where the origin AS illegitimately claims prefix ownership. Each incident is identified using trusted reports from BGPmon, RIPE Labs, MANRS report, and ThousandEyes [21]–[23].

We extract all BGP announcements for each prefix announced by the hijacker AS from all available RIPE RIS collectors, covering **24 hours before** the reported start time (W_{pre}), the **anomaly window** (W_{anom}), and **24 hours after** the reported end time (W_{post}). Prefixes observed in either W_{pre} or W_{post} are considered *legitimate*, representing normal routing behaviour. During W_{anom}, announcements from the hijacker AS not in the legitimate set are labelled *anomalous*.

Let P_{pre}, P_{post}, and P_{anom} denote prefix sets announced by the hijacker AS in respective windows. The legitimate and

TABLE I
PER-MESSAGE CLASSIFICATION AND SIGNAL GENERATION RULES

Condition	Label	Signal
Prefix in Stable Baseline, origin matches	0 (Benign)	None
Prefix in Stable Baseline, origin differs	1 (Anomalous)	BD
Prefix in Unstable Baseline	1 (Anomalous)	BD
Prefix not in Stable or Unstable Baseline	No_baseline	NP

anomalous prefix sets are defined as: $LP_{\text{legit}} = P_{\text{pre}} \cup P_{\text{post}}$ and $LP_{\text{anom}} = P_{\text{anom}} \setminus LP_{\text{legit}}$

For each BGP update u observed in W_{anom}:

$$\text{label}(u) = \begin{cases} 1, & \text{if origin}\text{AS}(u) = \text{hijacker}\text{AS} \\ & \quad \wedge \ \text{prefix}(u) \in LP_{\text{anom}}, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

This ensures only updates corresponding to true hijacked prefixes receive the anomalous label (1), whilst legitimate announcements from the hijacker AS, even during the anomaly window, receive the benign label (0) and are excluded from positive detection.

C. Baseline Construction

We construct reference baselines of prefix-origin AS mappings for each VP from the most recent RIB snapshots over a fixed *stability window* of N snapshots (e.g., the last four snapshots). A prefix-origin pair is classified as *Stable* if it appears in at least a predefined *stability threshold* number of snapshots (e.g., ≥ 3 out of 4). Stable mappings form the VP's *Stable Baseline*, representing routing relationships consistently observed over time.

Prefix-origin pairs present within the stability window but failing to meet the stability threshold are placed in the *Unstable Baseline*. These capture mappings that appear intermittently or exhibit frequent changes, representing benign fluctuations or early signs of anomalous activity. Pairs absent from both baseline sets are treated as entirely new and trigger NP signals when observed. This conservative approach prioritises sensitivity at the cost of potentially flagging legitimate routing fluctuations.

By maintaining both stable and unstable baseline sets per VP, our framework distinguishes between highly reliable routing relationships and those that are less certain. In our evaluation, we used $N = 4$ snapshots with a stability threshold of 3.

D. Real-Time Per-Message Classification

With baselines established, the framework classifies every incoming BGP update according to the rules in Table I. This classification determines whether an update represents expected routing behaviour or should trigger an anomaly signal.

For each update, the system checks whether the announced prefix-origin pair appears in the VP's stable baseline. If present with matching origin AS, the update receives a *benign* label and generates no anomaly signal. If the prefix exists in the stable baseline but with different origin AS, this indicates potential hijacking and generates a BD signal.

Updates involving prefix-origin pairs in the unstable baseline also generate BD signals, reflecting our conservative approach treating deviations from any known mapping as potentially suspicious. Updates announcing prefixes absent from both baseline sets receive *no_baseline* classification and generate NP signals, indicating entirely new prefix announcements requiring investigation.

E. Multi-Perspective Signal Aggregation

$\text{BD}_v(t)$ and $\text{NP}_v(t)$ signals are generated from per-message classification according to Table I for each individual VP $v \in V$ using fixed-length time bins (typically one minute) at each time point t. These individual signals are aggregated at each time point t using the maximum classification count in each signal formally defined as:

$$\text{BD}_{\max}[t] = \max_{v \in V} \text{BD}_{\text{v}}[t] \tag{2}$$

$$\text{NP}_{\max}[t] = \max_{v \in V} \text{NP}_{\text{v}}[t] \tag{3}$$

This temporal aggregation reduces computational overhead, provides natural noise filtering by smoothing brief signal spikes, and enables statistical analysis over meaningful time intervals.

The MAX aggregation operator ensures anomalies visible from even a single VP are preserved in the global signal, avoiding dilution of viewpoint-specific events that can occur with averaging approaches. This reflects the reality that many BGP anomalies exhibit limited visibility and may only be observable from particular network VPs.

F. Statistical Anomaly Detection

To identify statistically significant deviations in aggregated time series, we employ Z-score [24] thresholding applied independently to $\text{BD}_{\max}$ and $\text{NP}_{\max}$ signals. For each incident, we compute the empirical mean μ and standard deviation σ from the 24-hour pre-incident reference window (W_{pre}) for each aggregated series.

The Z-score for each time interval t is calculated as:

$$z(t) = \frac{S[t] - \mu}{\sigma}, \quad S \in \{\text{BD}_{\max}, \text{NP}_{\max}\} \tag{4}$$

If $z(t)$ exceeds a predefined threshold τ_z, interval t is flagged as *anomalous* for that signal type. This statistical approach provides principled anomaly detection grounded in observed baseline behaviour whilst allowing threshold adjustment based on operational requirements.

G. Signal Fusion and Final Decision

The final stage combines independently detected anomalies from BD and NP channels into a unified decision timeline. This fusion ensures events detected by either signal type are reflected in the final output, maximising detection coverage across different anomaly categories.

Let $\text{BD}_{\text{flag}}[t]$ and $\text{NP}_{\text{flag}}[t]$ denote binary anomaly indicators obtained after Z-score thresholding. The fused anomaly decision for interval t is:

$$\text{Final}_{\text{Anomalies}}[t] = \text{BD}_{\text{flag}}[t] \vee \text{NP}_{\text{flag}}[t] \tag{5}$$

This logical disjunction fusion preserves anomalies attributable to either BD or NP signal appearances, providing comprehensive coverage whilst maintaining operational simplicity. The resulting unified anomaly series enables direct evaluation against per-update ground truth and computation of standard metrics, including precision, recall, and F1-score.

IV. Experimental Results

We evaluate our hybrid BGP anomaly detection framework using six well-documented real-world incidents that represent different anomaly characteristics and operational contexts. Our experimental design emphasises rigorous cross-incident validation to demonstrate generalisation capability across diverse threat scenarios. Our methodology will report performance using standard performance evaluation (precision/recall); however, it should be pointed out that in practical anomaly detection, it is important to get a strong positive signal even if not all of the positive events are detected. In Section V, we will discuss the precision/recall values with regard to practical deployment.

A. Experimental Setup and Dataset Characteristics

We evaluate our method on six well-documented Type 5 incidents spanning diverse geographic regions and manifestation patterns. These include Turk Telekom (2004) [21], a large-scale route leak affecting global routing tables; Vodafone India (2021) [22], involving misconfiguration with 30,000+ incorrect prefixes; Rostelecom (2020) [23], a prefix hijacking incident affecting 8,000+ prefixes from major CDNs and cloud providers; and historical incidents: Indosat 2011 [25], Indosat 2014 [26], and China Telecom (2010) [27].

Each dataset comprises three complete days: 24 hours before the incident, the incident day, and 24 hours afterwards, yielding approximately 10^6 BGP update messages per dataset. Available viewpoints (VPs) vary based on RIPE RIS infrastructure active during each incident period, ranging from 7 to 24 collectors.

The incidents exhibit markedly different temporal characteristics. Turk Telekom, Indosat 2011, and China Telecom demonstrate scattered, bursty patterns with multiple distinct peaks separated by quiet intervals. In contrast, Vodafone India and Rostelecom show continuous, persistent anomalous behaviour throughout their windows. Indosat 2014 exhibits clustered bursts with multiple peaks closely spaced in time. These varied manifestation patterns necessitate hybrid BD and NP detection mechanisms to capture the full diversity of Type 5 incident behaviours.

B. Signal-Type Analysis and Anomaly Characterisation

To understand how different Type 5 incidents manifest through our dual-signal approach, we analyse the temporal patterns of ground truth anomalies alongside aggregated $\mathrm{BD}_{\max}$ and $\mathrm{NP}_{\max}$ signal counts. This analysis reveals distinct signal signatures across incidents that validate our hybrid detection strategy, as summarised in Table II.

TABLE II
Evaluation Dataset Characteristics

Incident	VP	Minutes	Primary Signal	Temporal Pattern
Turk Telekom	24	204	BD	Scattered, bursty
Vodafone	14	13	BD	Continuous
Rostelecom	13	11	NP	Continuous
Indosat 2011	20	20	NP	Clustered bursty
Indosat 2014	18	149	BD	Clustered bursty
China Telecom	7	17	NP	Continuous, multi-peaked

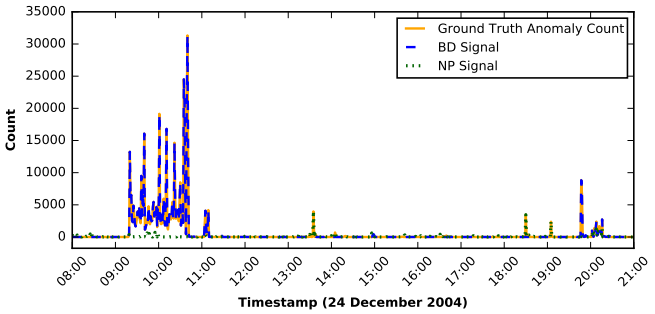

Fig. 1. Per-minute ground truth anomalous messages, $\mathrm{BD}_{\max}$ signals, and $\mathrm{NP}_{\max}$ signals for Turk Telekom incident.

Our signal analysis highlights three key patterns in anomaly characteristics:

BD-Dominated Events: Turk Telekom, Vodafone India, and Indosat 2014 exhibit strong correlation between ground truth anomalies and $\mathrm{BD}_{\max}$ signals (e.g., Figure 1), whilst $\mathrm{NP}_{\max}$ signals remain near baseline levels. These incidents primarily involve origin AS changes for known prefixes, where legitimate address space is announced with incorrect origin attribution.

NP-Dominated Events: Rostelecom, Indosat 2011, and China Telecom exhibit the opposite pattern, with $\mathrm{NP}_{\max}$ signals closely tracking ground truth anomalies (e.g., Figure 2) whilst $\mathrm{BD}_{\max}$ signals remain negligible. These incidents involve announcements of previously unseen address space rather than claiming ownership of existing prefixes.

This analysis clearly demonstrates why relying on either signal type alone is insufficient for comprehensive detection. BD-only detection achieves strong performance on origin AS deviation events but fails on new prefix announcements. Conversely, NP-only detection misses origin AS manipula-

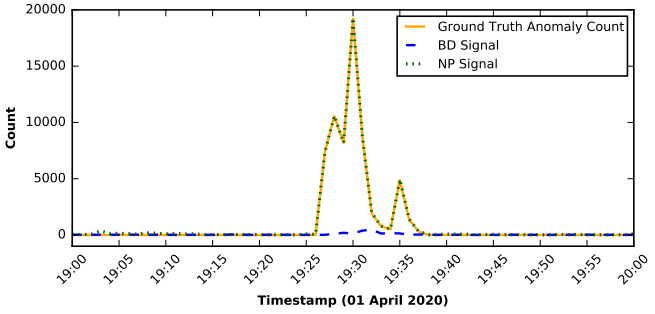

Fig. 2. Per-minute ground truth anomalous messages, $\mathrm{BD}_{\max}$ signals, and $\mathrm{NP}_{\max}$ signals for Rostelecom incident.

TABLE III
LEAVE-ONE-OUT CROSS-INCIDENT VALIDATION RESULTS

Testing	Optimal τ_z	Precision	Recall	F1
Turk Telekom	10	0.96	0.54	0.69
Vodafone	10	0.92	0.85	0.88
Rostelecom	10	0.88	0.64	0.74
Indosat 2011	10	0.54	0.48	0.51
Indosat 2014	11	0.96	0.71	0.82
China	9	0.36	0.71	0.48
Mean	-	**0.77**	**0.65**	**0.69**

tion incidents. Our hybrid detection framework overcomes these limitations by combining both signal types through principled fusion, thereby maximising detection coverage across diverse Type 5 manifestation patterns.

C. Cross-Incident Validation and Generalisation Assessment

The fusion system achieves mean F1-score of 0.69, with performance ranging from 0.48 to 0.88 across all six folds, demonstrating effective generalisation across diverse incident manifestations. High-performing incidents (Vodafone 0.88, Indosat 2014 0.82) clearly distinguish anomalous signals from normal behaviour, whilst low-performing incidents (China 0.48, Indosat 2011 0.51) reveal detection challenges.

Low recall cases like Turk Telekom (0.54) involve sparse anomalous messages that struggle to exceed the threshold. As anomalous activity subsides, signals climb but remain below τ_z before dropping, leaving tail anomalies undetected. This is inherent to statistical detection on sparse, low-magnitude signals. Operationally, this limitation is less critical since strong positive signals during active anomaly periods provide sufficient detection capability for incident response.

Conversely, low precision cases like China Telecom (0.36) indicate the detector flags intervals not captured in our ground truth, likely representing other BGP disruptions undocumented in trusted reports. These are probably legitimate anomalies rather than false positives. Operationally, higher recall with moderate false positives is often preferable to missing true incidents, and a more comprehensive ground truth would improve precision scores.

These patterns reflect the inherent characteristics of different incident manifestations and the challenges of maintaining consistent performance across diverse scenarios using fixed parameters. The variation in optimal thresholds (9 to 11) indicates that different incidents require different sensitivity levels, suggesting future work might benefit from adaptive thresholding strategies.

V. DISCUSSION

Our dual-signal framework addresses a fundamental limitation in BGP anomaly detection: single-signal approaches inevitably fail to capture all manifestations of RFC 7908 Type 5 prefix hijacking incidents. We demonstrate that BD and NP signals provide complementary detection mechanisms, with BD signals identifying incidents through origin AS changes on known prefixes and NP signals capturing incidents through previously unseen prefix announcements, providing theoretical justification and empirical proof for hybrid detection architectures. Although our results did not detect every single anomalous event, they provided strong signals during anomaly periods, offering actionable insight for network operators to alert on BGP anomalies. The cross-incident evaluation demonstrates the framework's ability to detect anomalies across diverse scenarios. Although performance varies by incident type, the framework consistently provides strong signals and good precision for operational use, with false positives mostly attributable to undocumented BGP anomalies rather than detector failures.

The per-update ground truth methodology establishes new evaluation standards, replacing coarse time-window labelling with semantic precision, ensuring only genuinely hijacked announcements receive positive labels. This rigorous labelling enables meaningful cross-incident generalisation assessment.

Compared to volume-based statistical methods, our prefix-origin semantic approach provides direct anomaly interpretation with clear operational meaning. Unlike data-hungry ML approaches requiring extensive labelled datasets, our hybrid design uses unlabelled RIB data for baseline construction whilst adapting through statistical thresholding, crucial for organisations with limited historical incident data. The MAX aggregation strategy addresses BGP's asymmetric visibility problem, preserving strong anomalies from any monitoring location whilst providing noise reduction. Signal-type analysis reveals complementary detection capabilities: BD signals effectively identify Type 5 incidents manifesting through origin AS deviations on known prefixes, whilst NP signals prove essential for detecting Type 5 incidents manifesting through previously unseen prefix announcements. Our per-message labelling and dual-signal framework create a foundation for future ML integration to learn adaptive thresholding strategies or optimised signal fusion, combining rule-based interpretability with ML's pattern recognition.

Our framework tackles critical deployment barriers preventing academic BGP detection from reaching production environments. Dual-signal categorisation enables immediate incident characterisation: BD anomalies indicate route leaks requiring upstream coordination, whilst NP anomalies suggest hijacking attacks demanding rapid filtering responses. The system provides interpretable detection rationales with traceable decision logic, supporting regulatory compliance and operator training, which are essential for NOCs requiring clear explanations during incident response.

Performance variation across incidents reveals important deployment considerations. Low recall on sparse incidents reflects a fundamental challenge: when anomalies are distributed sparsely across multiple minutes, they struggle to exceed statistical thresholds. As anomalous activity subsides, signals may climb but remain below threshold before dropping, leaving tail anomalies undetected. Addressing this limitation requires balancing threshold sensitivity against

false positive rates. Additionally, geographic concentration of RIPE RIS collectors may limit generalisation to regions with different routing topologies or fewer monitoring points.

Future work should investigate adaptive thresholding strategies that adjust sensitivity per incident type, expand detection to encompass all RFC 7908 route leak types (currently focused on Type 5), and integrate ML to learn optimised fusion strategies beyond logical disjunction.

VI. Conclusion

This paper demonstrates that effective BGP anomaly detection requires hybrid multi-signal architectures capable of identifying both route leaks and prefix hijacking attacks with fine temporal granularity. We present a per-message detection framework combining BD and NP signals, establishing new evaluation standards through per-update ground truth construction that replaces coarse time-window labelling with semantic precision.

Experimental validation on six diverse real-world Type 5 incidents reveals complementary detection mechanisms: BD signals identify incidents manifesting through origin AS changes on known prefixes, whilst NP signals prove essential for detecting incidents manifesting through previously unseen prefix announcements. Cross-incident leave-one-out validation achieves a mean F1-score of 0.69, with individual incident performance ranging from 0.48 to 0.88 depending on incident manifestation characteristics. This demonstrates effective generalisation across diverse, entirely unseen incident types, a strong result for a lightweight approach requiring no machine learning.

By addressing BGP's asymmetric visibility problem through MAX aggregation, our method preserves anomalies visible from any monitoring viewpoint whilst reducing noise. The rule-based, interpretable design supports regulatory compliance and operator training, which are critical requirements for operational NOC environments.

This academic research establishes foundations for future research which will lead to efficient operational detection. Multi-stage detection architectures combining our dual-signal framework with ML could leverage our per-message ground truth for feature extraction and adaptive thresholding. Expanding detection beyond RFC 7908 Type 5 to encompass the full route leak taxonomy would create comprehensive BGP security. As RPKI, BGPsec, and ASPA deployment continue, complementary anomaly detection provides essential defence layers, positioning this framework for production BGP monitoring systems.

VII. Acknowledgement

The author acknowledges the financial support provided by the Indonesian Education Scholarship (BPI) from the Center for Higher Education Funding and Assessment (PPAT) and the Indonesian Endowment Fund for Education (LPDP).

References

[1] National Institute of Standards and Technology, "NIST RPKI Monitor Analysis Report." https://rpki-monitor.antd.nist.gov/ROV/20250731.18/All/All/4, 2025.

[2] L. Bruder, M. Müller, and R. Koning, "BGPsec — could you run it if you wanted to?." https://blog.apnic.net/2025/05/23/bgpsec-could-you-run-it-if-you-wanted-to/, 2025.

[3] RIPE NCC, "RPKI Quarterly Planning." https://www.ripe.net/publications/documentation/quarterly-planning/rpki/, June 2025.

[4] K. Sriram, D. Montgomery, and et al., "Problem Definition and Classification of BGP Route Leaks." RFC 7908, June 2016. Informational.

[5] M. Lad and et al., "PHAS: a prefix hijack alert system," in *USENIX Security Symposium*, USENIX, 2006.

[6] Z. Zhang and et al., "iSPY: detecting IP prefix hijacking on my own," *SIGCOMM Computer Communication Review*, vol. 38, no. 4, p. 327–338, 2008.

[7] H. Ballani, P. Francis, and X. Zhang, "A study of prefix hijacking and interception in the internet," in *SIGCOMM*, pp. 265–276, ACM, 2007.

[8] L. Wang and et al., "Observation and Analysis of BGP Behavior under Stress," in *IMW*, IMW '02, pp. 183–195, ACM, 2002.

[9] J. Mai, L. Yuan, and C.-N. Chuah, "Detecting BGP anomalies with wavelet," in *NOMS*, pp. 465–472, 2008.

[10] P. Sermpezis, V. Kotronis, P. Gigis, and et al., "ARTEMIS: Neutralizing BGP Hijacking Within a Minute," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2471–2486, 2018.

[11] B. A. Scott and et al., "Matrix Profile data mining for BGP anomaly detection," *Computer Networks*, vol. 242, p. 110257, 2024.

[12] B. A. Scott and et al., "BGP anomaly detection as a group dynamics problem," *Computer Networks*, vol. 257, p. 110926, 2025.

[13] N. H. Hammood and et al., "A Survey of BGP Anomaly Detection Using Machine Learning Techniques," in *Applications and Techniques in Information Security*, pp. 109–120, Springer Singapore, 2022.

[14] B. A. Scott and et al., "A Survey of Advanced Border Gateway Protocol Attack Detection Techniques," *Sensors*, vol. 24, no. 19, 2024.

[15] M. Cheng and et al., "MS-LSTM: A multi-scale LSTM model for BGP anomaly detection," in *ICNP*, 2016.

[16] K. Hoarau and et al., "BGNN: Detection of BGP Anomalies Using Graph Neural Networks," in *ISCC*, pp. 1–6, 2022.

[17] T. Holterbach and et al., "A System to Detect Forged-Origin BGP Hijacks," in *NSDI*, pp. 1751–1770, USENIX, Apr. 2024.

[18] K. Chen, "BGP prefix hijack detection algorithm based on MOAS event feature," in *CNSSE*, vol. 13175, p. 131750P, SPIE, 2024.

[19] R. Hakimi and M. J. Reed, "Feature Analysis and Selection for BGP Anomaly Detection," in *ICIN*, pp. 99–106, 2025.

[20] "RIS Data Access." RIPE RIS Database, https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-data-access. Accessed: 2025-04-01.

[21] A. C. Popescu, B. J. Premore, and T. Underwood, "The Anatomy of a Leak: AS9121." https://archive.nanog.org/meetings/nanog34/presentations/underwood.pdf, May 2005. NANOG 34 Presentations.

[22] A. Siddiqui, "A Major BGP Hijack by AS55410-Vodafone Idea Ltd." MANRS report, https://manrs.org/2021/04/a-major-bgp-hijack-by-as55410-vodafone-idea-ltd/, April 2021.

[23] A. Siddiqui, "Not just another BGP hijack." MANRS report, https://manrs.org/2020/04/not-just-another-bgp-hijack/, April 2020.

[24] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, July 2009.

[25] A. Toonk, "Hijack by AS4761 Indosat: A quick report." BGPmon report, https://www.bgpmon.net/hijack-by-as4761-indosat-a-quick-report/, January 2011. Accessed: July 11, 2025.

[26] R. Wilhelm, "BGP leaks in indonesia." RIPE Labs, 4 April 2014, Available online https://labs.ripe.net/author/wilhelm/bgp-leaks-in-indonesia/.

[27] A. Toonk, "Chinese ISP hijacked 10% of the Internet." BGPmon report, https://www.bgpmon.net/chinese-isp-hijacked-10-of-the-internet/, April 2010. Accessed: July 10, 2025.