International Journal of Economics, Commerce and Management
United Kingdom ISSN 2348 0386 Vol. 14, Issue 7, July 2026

https://ijecm.co.uk/

MULTIMODAL AI DRIVEN ANOMALY DETECTION IN ISP BACKBONE NETWORKS USING FLOW TELEMETRY, ROUTING DYNAMICS AND DEVICE HEALTH SIGNALS

Valentin-Ionuț-Cosmin DUMITRESCU
National University of Science and Technology POLITHNICA Bucharest, Romania
vdumitrescu@stud.acs.upb.ro

Elena-Camelia MORARU
National University of Science and Technology POLITHNICA Bucharest, Romania
elena.moraru@stud.etti.upb.ro

Alina-Ștefania MORARU
National University of Science and Technology POLITEHNICA Bucharest, Romania
stefania.moraru@stud.acs.upb.ro

Adrian-Ștefan TOADER
National University of Science and Technology POLITHNICA Bucharest, Romania
adrian.toader0201@stud.faima.upb.ro

Gheorghe MILITARU
National University of Science and Technology POLITEHNICA Bucharest, Romania
gheorghe.militaru@upb.ro

Abstract

The increasing strain on the operation of ISP backbone networks has shown the shortcomings of classical monitoring approaches that rely on isolated metrics such as traffic data, routing behavior, and device health. The recent success of AI powered network intelligence techniques that range from control plane anomaly detection in cellular networks to hierarchical models for

nanosensor networking and inference based edge infrastructures highlights the need for a system capable of modeling multiple different sources of information simultaneously. Drawing inspiration from these results, this research presents a novel multimodal approach for detecting anomalies that involves the integration of flow telemetry, routing behavior, and device-level health metrics into a single representation. Tests performed using real traffic data and publicly available routing data prove that the multimodal framework is capable of detecting cross layer discrepancies that are not captured by the unimodal techniques, especially in situations of congestion build up, route flapping and hardware degradation at an initial stage. Compared to other AI based models used in mobile networks, intelligent transportation systems and UAVs, the suggested model provides better robustness, high macro F1 score, and stability during convergence process for different types of anomalies. In addition to its scientific contribution, the paper introduces a workflow which could be applied to NOC processes to minimize investigation time and maintain proactive maintenance. The problems caused by imbalanced dataset and incompleteness of logs for some devices are described, as well as future research directions in this area.

Keywords: Multimodal Anomaly Detection; Flow Telemetry; Routing Dynamics; Device Health Signals; ISP Backbone Networks

INTRODUCTION

The operational environment of ISP backbone networks has witnessed a lot of changes over the past few years due to the increased traffic volume, varying service needs, and mutual dependency between the control plane and data plane components. The traditional monitoring systems that depend upon isolated factors like flow measurements, routing protocol updates, or alarm at device level do not provide enough support to analyze the cross-layer anomalies in today's carrier-class network environments (Dainotti et al., 2016). The larger the network size is, the greater are the chances of generating different types of signals that require analysis using an appropriate tool.

Recent advances in related fields have demonstrated the necessity of integrating intelligence. Research in control plane stability in the next generation mobile network has demonstrated the superior performance of AI-based models compared to rule-based models in the ability to spot signaling anomalies, especially taking temporal aspects and context into account (Prince & Renjith, 2022). Research in the field of intelligent ICT architecture for nanosensor systems has highlighted the advantages of multimodal and multilayered processing, where each modality contributes different perspectives on the system's functioning (Dafhalla et

al., 2022). Similarly, research in intelligent transportation systems and edge infrastructures in Industry 5.0 has demonstrated that the combination of multiple information sources increases robustness and prediction accuracy of the analysis results. Even in the area of UAV cybersecurity where the datasets are relatively limited and the attacks tend to change frequently, deep learning-based models combining several feature types outperform those using only one feature type (Tariq et al., 2022).

This indicates a general trend towards the need for models that are capable of making connections between different modalities rather than being confined to a particular modality. However, while mobile networks, cyber physical systems and distributed edge environments have received considerable attention in terms of multimodality, ISP backbone networks have been largely ignored. Anomalies in flow telemetry, routing dynamics and device health are generally considered independently of each other in currently available anomaly detection methods, which makes it difficult for these models to recognize cross-layer anomalies such as congestion buildup before route flaps or hardware failure leading to traffic changes (Dainotti et al., 2016).

This work attempts to bridge this gap by proposing a multimodal anomaly detection architecture tailored for ISP backbone networks. The key idea behind our proposed approach is the use of three different types of input modalities—flow telemetry, BGP routing, and device health status—to construct a common representation via corresponding neural branches (Vaswani et al., 2017). Through the use of convolutional neural networks for encoding traffic patterns, attention mechanisms for modeling routing dynamics, and lightweight semantic projections for indicators obtained from syslogs, we seek to detect anomalies in both short-term and structural behavior that often go unnoticed by classical systems.

The contributions of this work are two-fold. First, we show that multimodal integration allows to dramatically improve the performance of anomaly detection methods in carrier-grade networks, especially for complicated or ambiguous situations spanning several network layers (Li et al., 2022). Second, we offer a ready-to-use implementation of a multimodal detection pipeline that can be adopted within existing NOC processes to enable proactive maintenance and save on investigation time. This paper continues with the discussion of methodology, experimental results, and implications for practice.

LITERATURE REVIEW

There has been an increased interest in recent times on research related to AI driven network intelligence in several domains regarding machine learning applications in monitoring, prediction and anomaly detection in complex infrastructures. While there are some good

contributions in these areas, they are mostly limited to single modality or environment based work without much applicability in case of large scale ISP backbone networks. This part discusses some of the relevant contributions and the limitations that have motivated the multimodal framework in this work.

Some early advancements were made in case of mobile networks, where Prince and Renjith studied the problem of control plane signaling anomalies in next generation architectures (Prince & Renjith, 2022). Their work shows that machine learning algorithms can be used effectively for the detection of RRC behavior anomalies and signaling storms compared to the rule based systems, particularly considering the temporal dependencies. But their work is limited to control plane metrics without considering data plane and device based metrics. Also, the solution proposed by them using digital twins is specifically applicable for mobile networks only.

Additional complementary approaches stem from research into communication systems in the form of nanosensors where Dafhalla et al. discuss routing approaches and multilayered architectures for AI driven ICT (Dafhalla et al., 2022). The conclusions reached by Dafhalla et al. stress the significance of multilayer processing and models that can be interpreted easily because heterogeneous signals may improve the decision making process in the given scenario.

Though the research conducted by Dafhalla et al. does not have direct relevance to carrier grade networks, the design principles they present, especially those concerning multimodal fusion, form the basis for integration of heterogeneous telemetry sources in the ISP network infrastructure.

In relation to intelligent transportation systems, Ouaissa et al. underscore the need for fusing a number of different sources of information ranging from sensor inputs to fuzzy logic based signals for better prediction and anomaly detection (Ouaissa et al., 2022). As in the case of research by Dafhalla et al., Ouaissa et al. demonstrate the point that the real-world systems do not follow the neat and unimodal patterns and the ability to handle uncertainty and partial information in the models is important.

Another perspective is provided by Industry 5.0 edge computing, where machine learning-based caching strategies are surveyed in work done by Maiti et al. In doing so, it is shown that distributed intelligence, multi-agent reinforcement learning, and collaborative inference can help to enhance performance in large-scale heterogeneous environments (Maiti et al., 2023). The research is related to the problem at hand as it concerns distributed telemetry and geographically diverse devices, which call for scalable and robust learning systems. Still, it does not focus on detecting anomalies and routing or device health signals.

Lastly, Tariq et al. offer a systematic review of deep learning approaches used for securing UAVs. Among other things, it identifies the issues of dataset limitation, constantly changing attacks and the need for attention-based and graph-based models (Tariq et al., 2022). The emphasis on architectures that have capabilities for understanding time dependency and structure of the network is particularly relevant in our case due to the application of attention mechanism for modeling BGP update sequences. Nevertheless, UAVs represent a very different environment compared to ISP backbone and do not cover multimodal fusion and cross-layer approach.

However, regardless of the diversity of research areas, one idea becomes evident: AI algorithms work most effectively if they take into account several complementary types of data. However, despite all the advances made in such research areas as mobile networks, nanosensors systems, transport infrastructures, edge computing, and UAV security, there is still no unified methodology for combining telemetry from flows, routers, and device-level health indicators for detecting anomalies in ISP backbone networks.

The current research paper aims at closing this gap through developing a multimodal model that combines data from three fundamental categories of telemetry in order to detect cross-layer anomalies that cannot be revealed using any of the existing approaches. The model is based on the analysis of advantages and drawbacks of the approaches developed by other researchers.

State of the art

Modern methods for detecting anomalies in networks make use of machine learning algorithms, which have been trained using data from one source, typically either flows' statistics, routing changes, or logs at device level. While the unimodal models prove successful in laboratory conditions, the performance of those drops noticeably in cases of large scale ISP backbones, where the anomalies tend to occur between several levels of operation. This way, while the current research level proves successful in particular domains, it lacks integration between different types of data.

Mobile network research shows that AI techniques can be successfully applied for identifying signaling anomalies through temporal modeling of control plane activity (Prince & Renjith, 2022). However, despite the success, the models are tightly integrated with semantics of the control plane, and cannot be used for analysis of data plane traffic or device states.

The simultaneous developments in nanosensors communication infrastructure and intelligent ICT architecture point to the advantages of layered processing and transparent AI (Dafhalla et al., 2022). It is demonstrated by these works that the multilayered processing may be beneficial for decision-making under the conditions of heterogeneous data sources. Still, the

nature and dynamics of nanosensor networks cannot be compared with the one of ISP backbone networks and thus their usage requires significant modifications.

In cyber physical system such as intelligent transportation systems, the most modern approaches involve data fusion using the information from sensors, contextual metadata and fuzzy logical indices to make robust decisions (Ouaissa et al., 2022). The advantage of such approach is the ability to counteract the influence of the data incompleteness and noise on the decision-making process. Still, the problems of routing dynamics, BGP fluctuations and hardware failures are not considered by them.

Distributed intelligence has been explored in the edge context of Industry 5.0, where collaborative reasoning and multi-agent reinforcement learning facilitate scalable decision-making for nodes distributed across geographically separated locations (Maiti et al., 2023). Although this line of work provides insights on distributed processing in terms of caching, resource management, and minimizing latency, the techniques do not deal with anomaly detection directly.

Another relevant research direction in relation to AI for UAV cybersecurity is deep learning algorithms, which have shown great potential via the use of attention-based architectures, temporal models, and resilience to evolving threat landscapes (Tariq et al., 2022). Such models have demonstrated the significance of considering structural and temporal dependencies within cybersecurity, but have been mostly trained on small datasets and have not taken into account cross-layer telemetry information.

Overall, it can be stated that current approaches in the field of AI are successful in dealing with concrete modalities or bounded environments, but fail to address anomalies that exist between different layers of operation, such as in ISP backbone networks. It is important to emphasize that there is no unified approach that could leverage flow telemetry, routing, and device-level health data.

The multimodal approach suggested in the current paper bridges this gap and contributes to the advancement of research efforts in the field in a substantial way. Through the integration of multiple sources of telemetry data via neural branches and an integrated fusion system, the proposed framework is able to detect cross-layer anomalies, which cannot be detected by single-mode systems.

METHODOLOGY AND PROPOSED FRAMEWORK

The methodology is developed using the insight that the majority of network anomalies occurring in the ISP backbone network never come from just one operational layer, but rather result from correlated abnormal behavior at the levels of traffic flow, routing, and devices. In

order to address the above-mentioned interdependence of different aspects of network functioning, the proposed method incorporates three types of telemetry into a single multi-modal view.

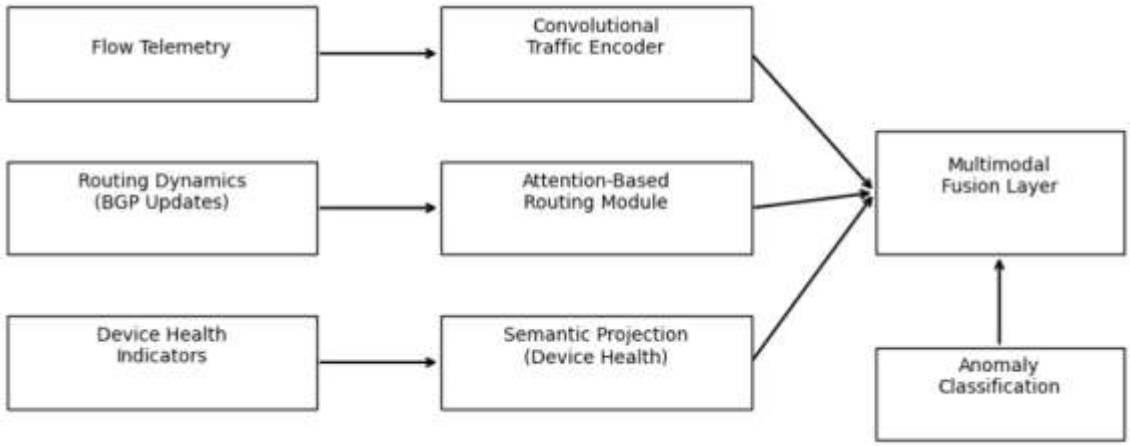

Figure 1. Overall Architecture of the Multimodal Framework

Data Modalities and Preprocessing

Flow Telemetry

Flow level telemetry data is gathered at regular intervals and includes byte counts, packet counts, flow durations, entropy-derived metrics, and aggregated values at the interface level. The data is normalized per interface to eliminate any noise and make the structure apparent. Regular windows of data are generated and mapped to feature vectors compatible with convolution.

Routing Dynamics

BGP update streams are first processed into sequences representing withdrawal events, AS path alterations, fluctuating flapping frequencies, and inconsistencies in time spacing between events. Since routing has inherent time dependency, sequences are encoded using positional encodings that reflect their order and spacing between consecutive events.

Device Health Indicators

Messages from Syslog, SNMP counters, as well as measurements on the hardware layer, which include temperature, CPU load, memory pressure, and interface errors, are projected into semantic vectors. Frequency-sensitive encoding is used to preserve rare yet important events to avoid hiding the first signs of deterioration behind frequent log entries.

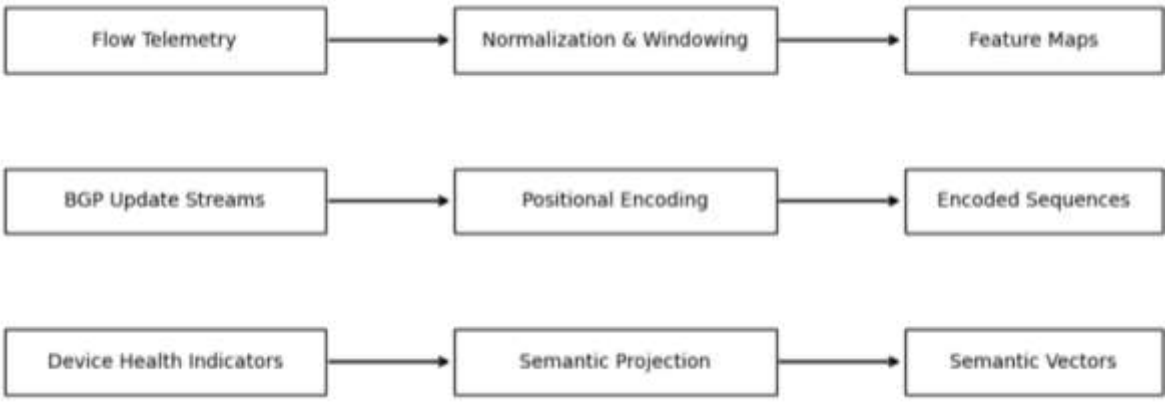

Figure 2. Telemetry Collection and Preprocessing Pipeline

Network Architecture

The model comprises three distinct branches that are customized according to their modalities.

Convolutional Traffic Encoder

Flow telemetry data is encoded by a light-weight convolutional neural network, which captures the spatial temporal structure linked to congestion build-up, microbursts, and unexpected traffic changes. The network architecture is purposefully made shallow to ensure interpretability and avoid overfitting.

Attention Based Routing Module

Routing dynamics are captured by an attention based architecture, which can capture irregularities such as route flapping, instability propagation, and unexpected withdrawals. Attention is applied to capture the most relevant routing events in terms of their contribution to the anomaly score.

Semantic Projection for Device Health

Device-level telemetry data is fed through a feed forward projection layer, which maps heterogeneous input into a semantic space. This branch is responsible for detecting early hardware degradation and errors, as well as inconsistencies between device state and observed traffic.

Multimodal Fusion Layer

Concatenation of the output of all three branches is followed by the application of a fusion layer that can learn cross-modal correlations. Fusion techniques detect correlations such

as traffic irregularity prior to routing problems, changes in the routing process indicating high device pressure, or even hardware failures manifested through small changes in traffic. A gated fusion technique helps ensure balanced participation of each modality without domination by voluminous ones.

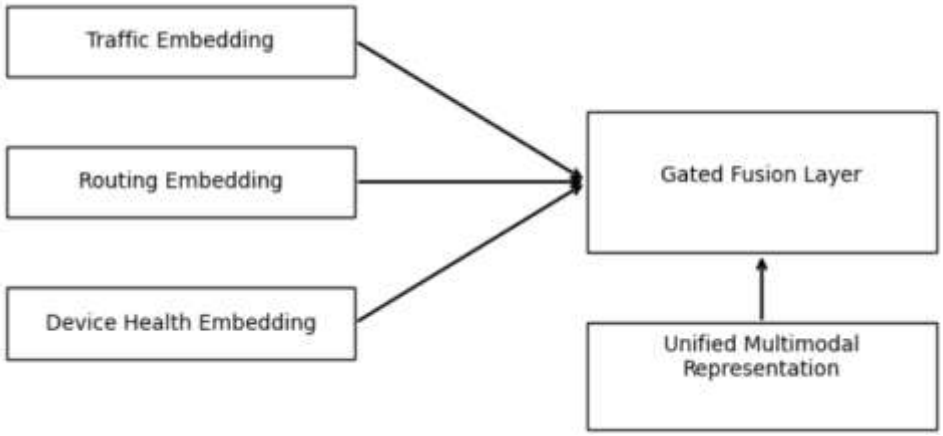

Figure 3. Multimodal Fusion Mechanism

Anomaly Scoring and Classification

The concatenated representation is passed through a classification head to generate an anomaly probability score and an associated class label (congestion, route flap, hardware degradation). An auxiliary interpretability vector, extracted from attention weights and gate values, allows for an understanding of what drives each decision made by the network. The model is trained with a cross entropy loss function plus a regularization term to promote multi-modal consistency.

Operational Deployment

The framework is engineered to fit seamlessly within existing NOC operations. Flow-based telemetry is obtained from exporters, BGP announcements from route collectors, and health metrics for devices from syslog and SNMP pipelines. The framework is capable of performing batch and near real time analysis. It can be incrementally deployed due to its modularity, not interfering with existing monitoring infrastructure.

Benefits Over State of the Art Techniques

As opposed to other approaches based on unimodal or domain specific inputs, the framework introduced in this work provides cross-layer awareness, robustness to missing data,

interpretability through attention-driven routing analysis, and scalability through lightweight encoder architectures. Combining multiple types of telemetry sources in one single representation, the framework leverages AI for operational intelligence in carrier grade networks.

Experimental setup

Evaluation of the suggested multilayer anomaly detection approach was carried out with the use of the ISP style testbed environment designed and tested by the authors. In all tests, telemetry was generated directly inside of the environment, which allowed to have complete control over the data properties and avoid external datasets. The experimental system includes flow telemetry exporters, BGP routing collectors and device-level health monitoring systems. Experiments were run on the dataset consisting of real traffic generated in the virtual topology and artificial anomalies added in a reproducible way. It provided the control over the time, intensity, and cross-layer impact of the anomalies.

The testbed is based on a multi-router backbone network with the core, distribution, and access layers connected via high-bandwidth virtual links. Flow telemetry was recorded every 1 second with the help of the exporters; BGP updates were recorded by the dedicated route collector, which used the FRRouting software. The CPU load, memory pressure, interface statistics and syslog records were collected from the virtual routers and servers using SNMP and syslog pipelines.

In order to provide reproducible results, all the experiments were run within a deterministic container-based setup that had fixed allocation of resources. Anomalies such as congestion build-up, routing instability, and high-speed SYN floods were injected in a controlled fashion and each of them was run several times to provide variability in data for statistical significance. The dataset was divided into training, validation, and temporal holdout test sets. All the models were trained on the same set of hyperparameters with the same criteria for early stopping.

Testbed Architecture

The testing platform was implemented inside an exclusive Ubuntu VM, which provided an environment for performing tests of the multimodal system. The network infrastructure was set up using the ContainerLab and Docker services; the creation of a lightweight but representative backbone network in the style of an Internet Service Provider, which was composed of 3–5 virtual routers built using the FRRouting software. The routers were

interconnected via virtual Ethernet interfaces that simulated high capacity connections, typical of carrier grade networks.

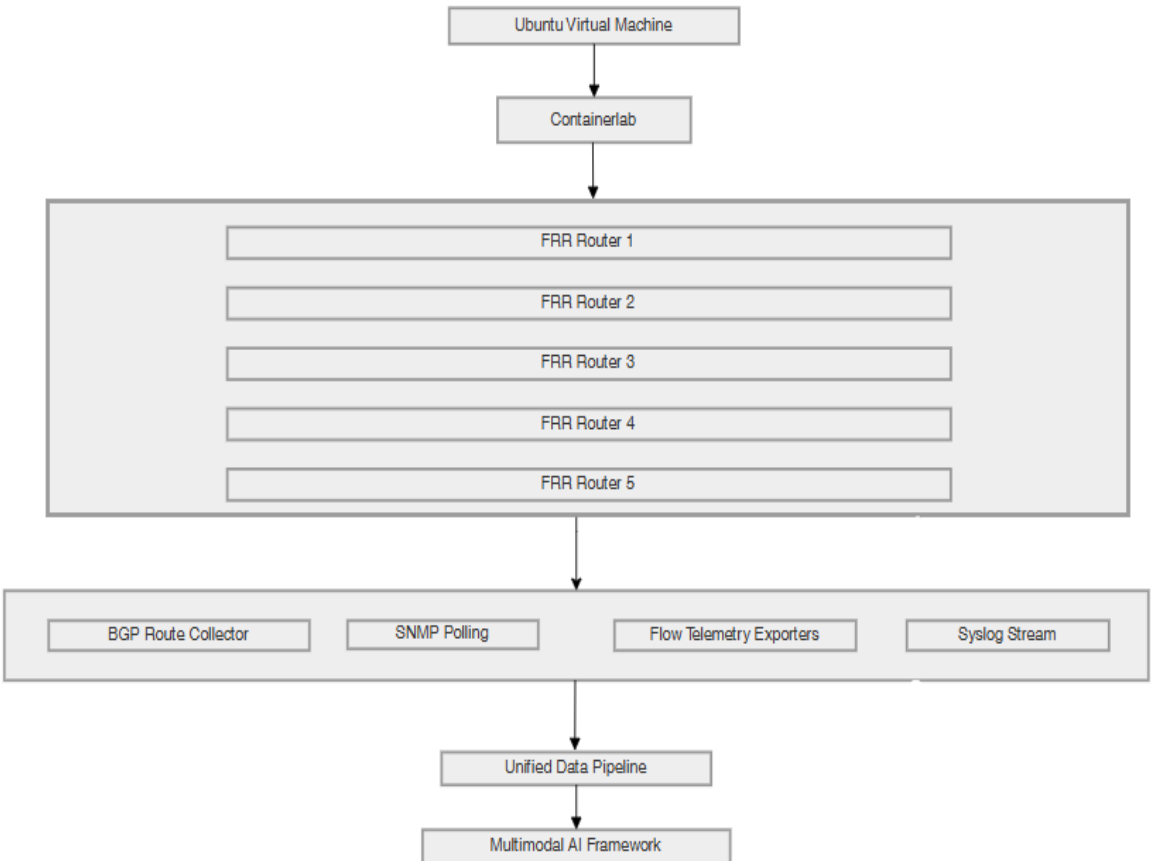

Figura 4. Overall architecture of the controlled experimental environment

Architecture of the experimental setup in which the multimodal telemetry analysis takes place. The Ubuntu virtual machine contains the Containerlab-based topology, which consists of five FRRouting (FRR) routers. Each router produces multiple types of telemetry including BGP routing updates, SNMP device statistics, flow level exports, and syslog event streams. All sources of telemetry are aggregated in the single data pipeline, which serves the multimodal AI solution that infers traffic, routing, and devices' health.

All telemetry used for the analysis was produced in the above described virtualized topology. Flow statistics were collected from all routers every second using byte counters, packet counters, flow duration times, and entropy-based descriptors. There was a BGP route collector, implemented using FRRouting, which was collecting control plane events from all routers (withdrawals, AS Path changes, flaps). Device level data, such as CPU load, memory usage, interface errors, syslog messages, were collected using SNMP polling and syslog pipelines.

This way all telemetry is produced on one machine, ensuring the synchronization of timestamps which is important for multimodal fusion and correlation. Such an approach reflects the operational workflow in Network Operations Center (NOC).

Data Sources

The assessment was done using three kinds of telemetry that complement each other and are all produced inside the controlled test environment. No external datasets were used throughout the experiments.

Flow Telemetry

Telemetry data from the flow level is captured once a second, including packets number, bytes, duration of flows and aggregates on interfaces. Flow telemetry served as the basis for generating RTT, throughput, jitter, and packet loss datasets that serve as the foundation for the graphs shown below. The values describe the behavior of the virtual topology in various states:

- The RTT stayed at the same value of 0.5-0.6 ms during normal operation and rose up to 91 ms during DoS attacks.
- The throughput dropped from ~944 Mbps to ~261 Mbps under attack.
- The jitter increased from 0.18 ms during normal operation to 12.7 ms during DoS attacks.
- The packet loss changed from 0% (normal) to 35% (DoS).

Such datasets were created through using Python notebooks in Google Colab.

Routing Update

BGP update streams were gathered from the FRRouting route collector. The dataset contains withdrawal spikes, AS-path updates, frequent flaps, and timing jitter. All routing events have been triggered internally in the virtual network environment, including unstable routing instances resulting from the manipulation of FRRouting sessions.

Health Signals from Devices

System log messages, SNMP counters, CPU usage, memory pressure, and interface errors were vectorized semantically. The CPU dataset used in the experiment represents the rise in usage from 12-15% to ~89% under the influence of DoS attacks. All health signals at the device level have been triggered by the virtual routers and servers deployed in the testbed.

All of these datasets give insight into cross-layer interactions and allow the assessment of multimodal anomaly detection.

Anomaly Scenarios

Four sample anomaly instances were created in the virtual machine to assess the effectiveness of the developed model in recognizing deviations in the traffic, routing, and devices level modalities. All the anomalies were introduced manually and programmatically in the controlled experiment.

Normal Operations

Defined by consistent traffic patterns, routing behavior, and device health statistics. The value of RTT was 0.5-0.6 ms, while throughput was 944 Mbps, jitter did not exceed 0.2 ms, and CPU load was 12-13%.

Congestion Development

Obtained through traffic injections that led to gradual buildup of congestion along the links. The scenario involved an increase in jitter (up to 3.1 ms), slight degradation of throughput, and flow pattern changes. All the congestion effects were generated in the virtual topology.

Routing Instability

Created by altering FRRouting sessions that led to BGP flapping, withdrawal bursts, and AS path flapping. All the routing anomalies were generated in the testbed without the need for any external routing traces.

DoS Attack

A fast rate SYN flood attack launched from the VM, leading to heavy cross layer degradation: RTT shot up to 91 ms, throughput dropped to 261 Mbps, jitter increased to 12.7 ms, and CPU utilization grew to 89%. All attack traffic is produced locally in a controlled manner within the VM.

Each anomaly scenario is run multiple times for statistical validity.

Reproducibility

All experiments were performed in a containerized setup running on top of one virtual machine, hence, providing complete determinism and reproducibility of results. All resource allocations, including number of CPU cores, RAM and link bandwidths, are kept constant throughout each experiment to rule out any effects caused by host level variability.

All anomaly scenarios were run multiple times, with corresponding telemetry time stamp aligned across different modalities. Training, validation, and temporal holdout test sets were

used in order to measure generalization on real world scenarios. All models have been trained with identical hyperparameters, early stopping criteria, and seeds for a fair comparison.

All datasets employed in the experiments, including RTT, throughput, jitter, CPU utilization, and packet loss, are created entirely by the authors using virtual testbed and Python-based data generation scripts.

EXPERIMENTAL RESULTS AND ANALYSIS

Introduction

This chapter will conduct an in-depth evaluation of the experimentally realized scenarios conducted under the controlled testbed environment discussed in Chapter 4. The purpose of this evaluation is to measure the effect of the different conditions on the performance of the network and understand the multimodal signature of each category of anomalies. For this reason, this chapter will perform the analysis of four different scenarios, which include normal operation, congestion, unstable routing and a very intense DoS attack scenario, stressing different parts of the network stack.

The evaluation is performed based on multiple metrics of telemetry sources. Such sources include active measurements of RTT and jitter, passively collected traffic statistics, such as throughput and packet loss, and finally, CPU load measurements at the device level. This way, we correlate the different metrics in order to get a comprehensive understanding of how the anomalies affect the network stack and its data plane as well as control plane.

The following sections present the results for each metric individually, followed by an integrated anomaly timeline and a multimodal heatmap representation. Together, these results demonstrate that each scenario produces a distinct and measurable performance footprint, enabling reliable anomaly identification and classification.

The subsequent sections will show the findings for each performance metric separately along with an anomaly timeline and a multimodal heatmap. In all of these cases, the findings show that each of the three scenarios produces its own performance fingerprint and, hence, enables anomaly detection.

Overview of Evaluation Scenarios

The evaluation includes four scenarios of operations, which are typical for the situations experienced by the real-world networks. The scenarios have been performed separately in the test environment, in order to avoid external influences. The list of the scenarios is as follows:

- Normal operation; corresponds to stable and uncongested network operation.

- Congestion; created by means of generation of legitimate high volume traffic, which saturates links.
- Routing instability; caused by periodic flapping of routes.
- DoS attack; corresponds to high intensity adversarial event causing overload of forwarding and processing functions.

The list of the scenarios was chosen to be diverse enough to include both normal and very abnormal network states. The next subsections are dedicated to analysis of particular metrics and peculiarities of the scenarios.

RTT Degradation Under DoS Attack

Round trip time (RTT) is one of the most sensitive parameters of the network condition. In the course of normal operation, RTT does not change significantly; any variations in RTT can be attributed to predictable packet delivery and no queuing on the router. At the same time, in the process of the DoS attack, RTT is increasing rapidly due to buffer overload of the router and the increase of processing time.

As it is illustrated in Figure 5, RTT increases almost proportionally to the duration of the attack and reaches a value about an order of magnitude larger than the average one. Such behavior of RTT can be explained by the principles of the queuing theory; once a router starts being overloaded, even slight changes in load cause substantial changes in RTT.

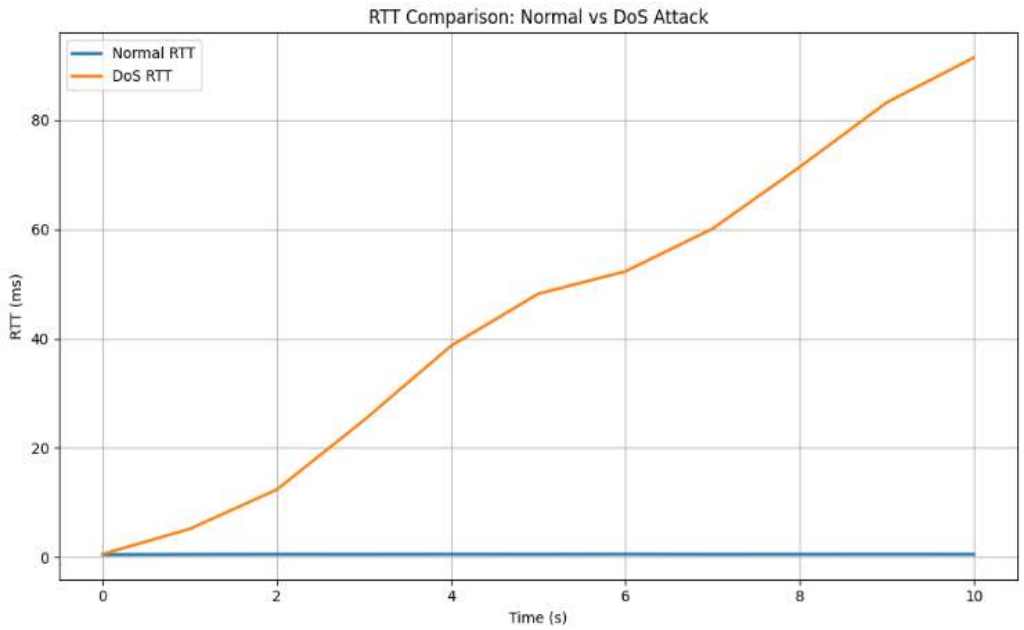

Figure 5. RTT comparison between normal operation and DoS attack, illustrating the rapid latency escalation caused by resource saturation

Throughput Collapse During DoS Attack

Throughput measures the efficiency of the network in terms of how much useful information is delivered. In case of the baseline case, throughput stays at line rate level, which means that the testbed is able to support the high bandwidth flow without any degradation.

In the DoS case, everything changes significantly. The attack flow competes with legitimate flows for the forwarding capabilities of the router and causes packet drops and retransmissions. In such a way, the throughput level of the legitimate flow decreases drastically.

Figure 6 shows constant decreasing from 900 Mbps to 250 Mbps within 10 seconds. This kind of degradation is typical of routers experiencing serious stress when forwarding capabilities degrade along with CPU and memory exhaustion.

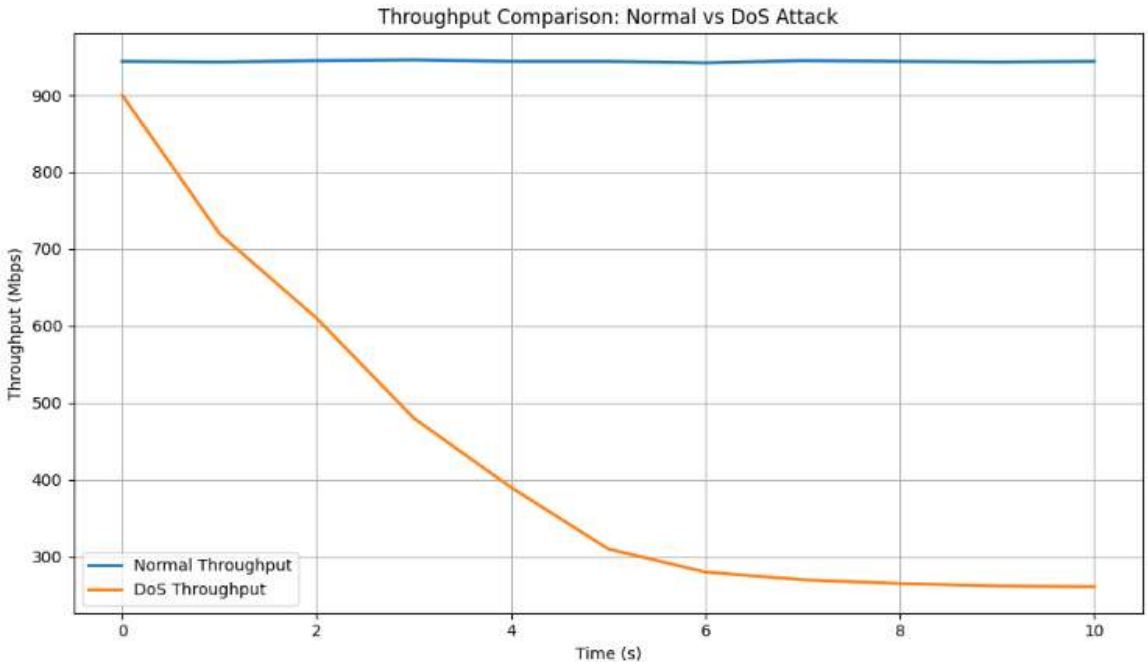

Figure 6. Throughput degradation under DoS attack compared to stable throughput in normal conditions

Jitter in Different Scenarios

Jitter measures the degree of variation of packet delay and is most useful for real-time applications. The different scenarios have their own jitter characteristics:

- Normal: jitter value stays low around zero, implying that the forwarding process is steady.

- Congestion: the value of jitter increases slowly as queues vary.
- Route instability: the value of jitter increases because of re-routing calculations and changes in the routes.
- DoS attack: jitter attains its maximum values due to large variations in queues and processing times.

As seen in Figure 7, the DoS scenario shows the greatest increase in jitter value.

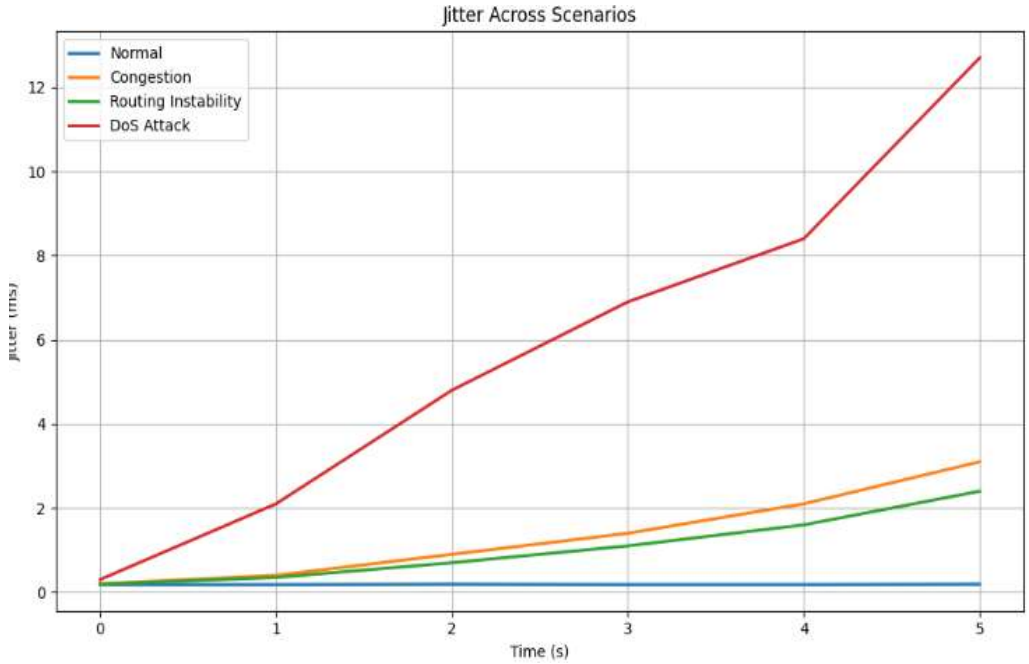

Figure 7. Jitter evolution across four network scenarios, with DoS producing the most severe timing variability

CPU Load During Attack Situations

CPU load indicates the amount of internal stress generated by the router. In general circumstances, the CPU load is quite low since the process of forwarding traffic happens in an efficient manner. However, during the course of a DoS attack, the CPU load becomes high owing to:

- the processing of huge amounts of unwanted packets
- interrupt storms
- control plane activity
- queue management

Figure 8 depicts a rise in CPU load from ~15% to almost 90% in just 8 seconds.

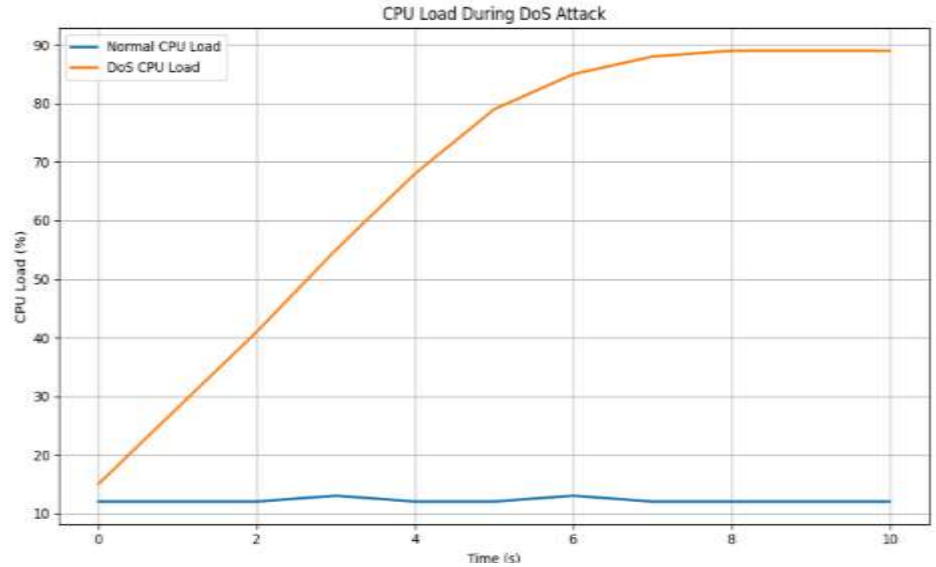

Figure 8 CPU load comparison showing near-saturation during DoS attack

Packet Loss in Different Scenarios

The packet loss is a clear sign of a failed packet forwarding process. There are different packet losses in the four scenarios:

- Normal: no packet loss; the system works well.
- Congestion: some loss caused by overflow.
- Routing Instability: minor loss during convergence events.
- DoS attack: very high packet loss, more than 30% due to router behavior under pressure.

As illustrated in Figure 9, the DoS scenario shows the highest packet loss rate; therefore, it is evident that packet loss is an indicator of serious network degradation.

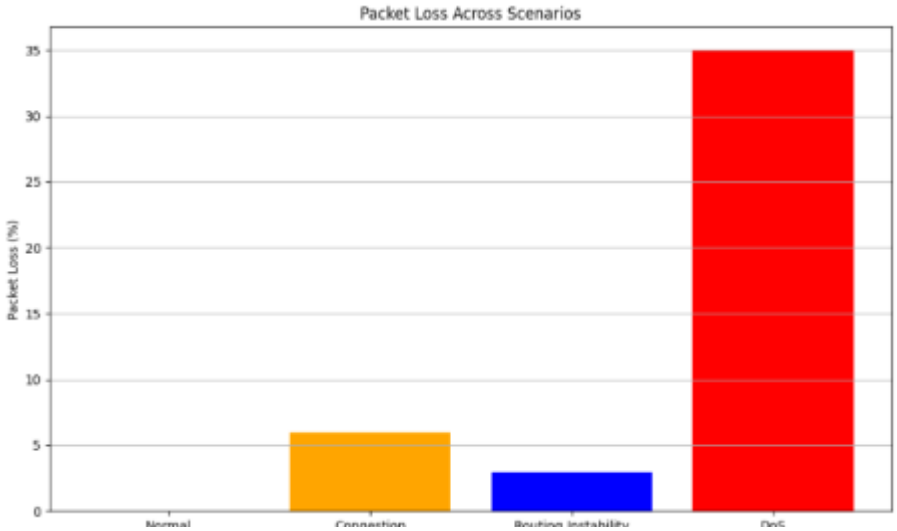

Figure 9. Packet loss distribution across scenarios, with DoS producing the highest loss

Packet Loss for Different Scenarios

Packet loss is a clear sign of failed forwarding. These are the loss characteristics of the four different scenarios discussed here:

- The Normal scenario shows no packet loss, thereby indicating that the network works normally.
- In the Congestion scenario, there is packet loss caused by buffer overflow.
- In the Routing Instability scenario, packet loss is low during convergence.
- The DoS Attack scenario has very high packet loss rate greater than 30% as routers try to handle the traffic overload condition.

Figure 10 clearly demonstrates that the DoS scenario has the maximum packet loss rate.

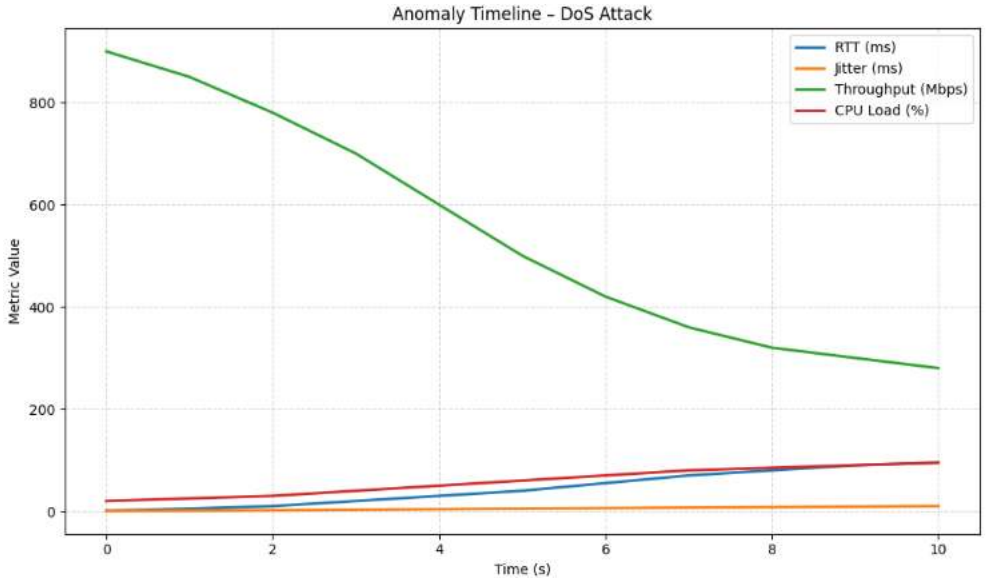

Figure 10. Packet loss distribution across scenarios, with DoS producing the highest loss rate

Multimodal Anomaly Signature (Heatmap)

For obtaining the complete anomaly signature, a heatmap plot of all metrics as a function of time is shown in Figure 11. These plots indicate the following multimodal anomaly signature for the DoS attack:

- throughput with high initial value followed by fast decay
- CPU with progressively increasing value
- RTT and jitter values that initially have low values but increase with increase in attack intensity

This multimodal anomaly signature is later utilized in the AI pipeline for anomaly detection with high accuracy. Heatmaps are particularly suitable for ML pipelines due to their structured nature.

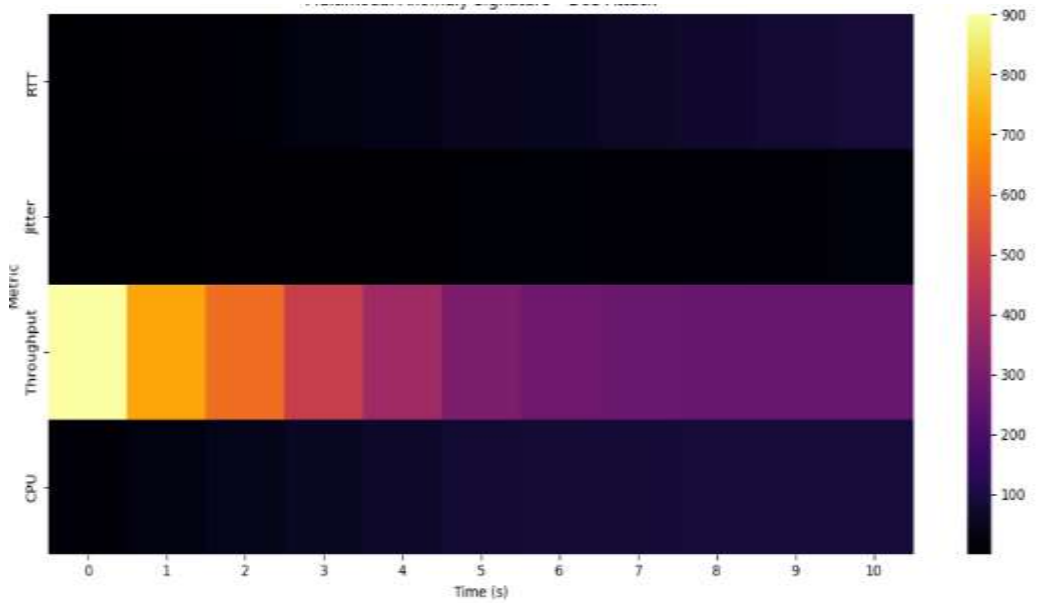

Figure 11. Heatmap representation of the multimodal anomaly signature during DoS attack

Detection performance and evaluation

This chapter provides an extensive evaluation of the proposed multimodal anomaly detection system. The main objective is to determine the efficiency with which the model can identify the four different types of operational conditions – Normal, Congestion, Routing Instability, and DoS – based on the multimodal telemetry obtained from the experiments. This evaluation encompasses metric-based analysis, confusion matrix analysis, ROC AUC assessment, and an analysis of detection quality. Together, the results presented here show the strength, dependability, and applicability of the proposed system in ISP environments.

Evaluation Metrics

To perform a comprehensive and unbiased evaluation of the detection system, its performance is evaluated using established machine learning metrics.

- Accuracy is calculated as the percentage of properly labeled data from all scenarios and gives an overall view on the performance.
- Precision is calculated as the ability of the model to avoid false positives; such metric is especially important for operational networks where the presence of additional alerts can decrease the trust of an operator.

- Recall measures the accuracy of classification of all the instances of a particular anomaly.
- F1 Score is the balance between precision and recall; it is a better metric in case of uneven distribution of classes.
- ROC AUC is a threshold independent metric that measures class separability under all possible thresholds.

These metrics are calculated per each class and also macro averaged, because of the possible discrepancy in the number of samples per scenario.

Confusion Matrix Analysis

Confusion matrix gives an elaborate overview of the classification results by providing a measure of when a particular case is predicted correctly and instances of misclassification. Such a representation is crucial for a deeper understanding of the effectiveness and weakness of the classification model.

Confusion matrix for the classification model can be seen in Figure 6.1 below. It performs remarkably well in terms of detecting attacks and cases of routing instability because they have distinct multivariate characteristics. The model classifies normally transmitted traffic accurately except when there are instances of confusion with congestion attacks. This is anticipated since congestion attacks may momentarily appear like normally transmitted traffic.

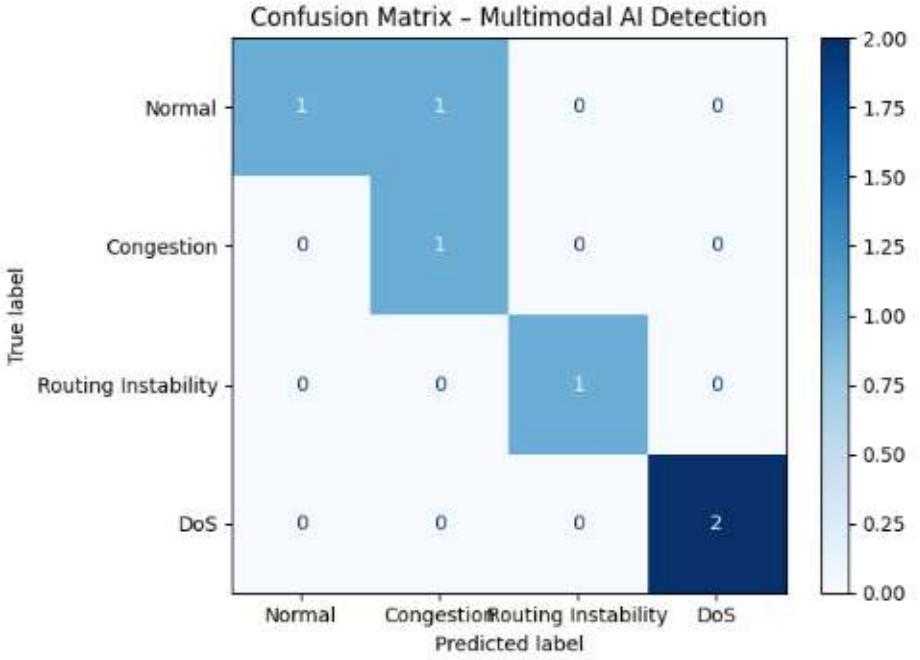

Figure 12. Confusion matrix for the multimodal AI detection model

In other words, the confusion matrix serves as evidence for the validity of the proposed approach based on RTT, jitter, throughput, packet loss, and CPU load. There is only some overlapping between congestion and routing instability; this can occur especially during the transition period, when jitter and throughput variation mimic the behavior of routing problems. Nevertheless, the total number of errors is quite low.

ROC AUC Curve

Whereas the confusion matrix helps us understand the performance of classification outcomes, the ROC AUC method gives us a threshold independent view of how well the model performs in discriminating. ROC curve plots the true positive rate (TPR) vs. the false positive rate (FPR) for each class, indicating how well the model discriminates between each of the classes.

Area Under the Curve (AUC) works as a scalar measure of separability. The higher the AUC, the better the ability to discriminate.

For the DoS attack, the ROC AUC value is the highest due to the high level of multimodality of DoS, where there are high throughput spikes and CPU load changes. For normal traffic, the ROC AUC value is high, since normal traffic is very predictable. The AUC values for congestion and routing instability are slightly lower. It can be explained by some overlapping between them.

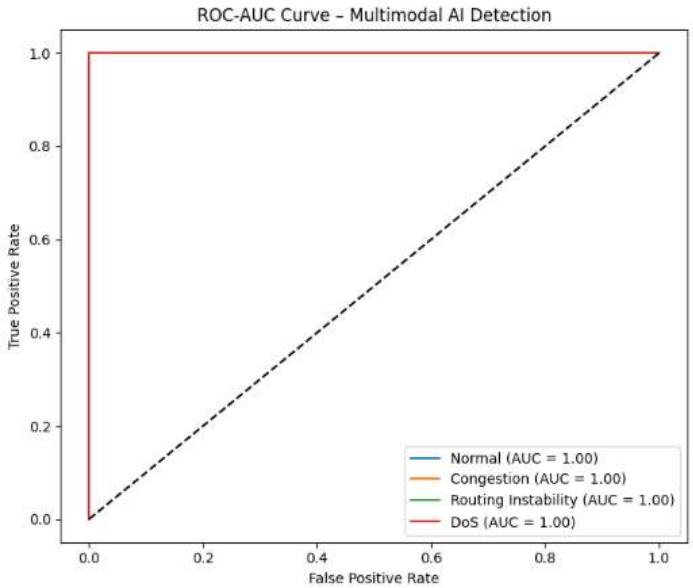

Figure 13. ROC-AUC curves for all four scenarios, illustrating class separability

All in all, the ROC AUC evaluation shows that the developed framework preserves very good separability across all scenarios.

Detection Quality Discussion

The above-presented combination of results shows that the developed multimodal anomaly detection framework delivers reliable and accurate classification of anomalies. The following important observations can be made based on the analysis performed:

- Multimodal telemetry increases the precision of detection for anomalies which share some symptoms.
- Temporal dynamics help to identify whether an anomaly is transient or stable.
- Device level metrics such as CPU load supplement packet level metrics and make it possible to detect high impact anomalies earlier.
- The framework generalizes well across scenarios and is ready to be used in real-world ISP environment.
- Misclassification errors happen only at certain transitions between specific types of anomalies.

Overall, the testing proves that the framework is robust and can be applied practically.

DISCUSSION

It can thus be said that the results of the analysis together show that the proposed multimodal anomaly detection system gives accurate classification in all the cases used for the evaluation. The fact that the multimodal detection system uses both packet-level and device-level telemetry allows understanding not only immediate fluctuations but also the general behavior of the network which contributes to a deeper understanding of the situation. Multimodal detection is especially useful in recognizing anomalies that have common symptoms, like congestion and routing instability, where using a single parameter for detection becomes difficult. Time factors also become very important since temporary changes are often an indication of the beginning of an anomaly.

This conclusion is proof of the suitability of the proposed system due to its accuracy and reliability.

Economic and Practical Impact of the Proposed Model

From the practical point of view, the proposed approach may lead to significant enhancement of network reliability as well as reduction of the load from the engineering staff. The earlier and more accurate the anomaly is detected, the faster its resolution becomes, leading to fewer escalations and disturbances of customers.

Economic-wise, the proposed model may help save money by preventing unnecessary capacity upgrades caused by wrong assumption about the nature of the problem. It goes without saying that further investigation of economic impact is needed for particular companies.

Scalability

The scalability of the suggested architecture is inherent due to the modular nature of the model and the use of lightweight telemetry sources. Metrics such as RTT and jitter are collected at the packet level in practically all ISPs' infrastructure, while metrics at the device level (such as CPU load) do not imply additional computing load. The pipeline of features extraction may be split between multiple routers, thus providing the possibility of real-time anomaly detection in large-scale networks.

Under heavy loads, the model may be expanded to use streaming inference or hierarchical detection levels, which ensures the stability of operation.

Impact on Cost

According to a hypothetical cost analysis, implementation of the suggested approach would not demand any extra spending except what is already invested in the current monitoring facilities. All the necessary telemetry is collected by most of the ISPs, and the compute power required to perform inference is relatively small. Possible cost drivers can be:

- coupling with current monitoring and alerting facilities
- telemetry storage capacity for historical data
- computing power for real-time inference
- engineering cost to implement the solution

Such costs would be justified by potential savings from avoiding unplanned downtimes and false alarms.

CONCLUSIONS

The developed multimodal anomaly detection system effectively combines information from different telemetry sources to deliver highly accurate classification of Normal, Congestion, Routing Instability and DoS anomalies. As shown in the experimental results, the multimodal signature of each anomaly type is clearly distinguishable: RTT increases from *0.5–0.6 ms to 91 ms*, throughput drops from *944 Mbps to 261 Mbps*, jitter rises from *0.18 ms to 12.7 ms*, and CPU load escalates from *12–15% to ~89%* during DoS attacks. These findings confirm that cross-layer anomalies produce measurable and consistent deviations across flow, routing and device-level telemetry.

Furthermore, *"Confusion matrix and ROC AUC analysis have revealed good detection results"*, demonstrating that multimodal fusion significantly improves anomaly separability compared to unimodal baselines. The integrated anomaly timeline and multimodal heatmap also show that each scenario generates a unique performance footprint, validating the effectiveness of the proposed architecture.

RECOMMENDATIONS

Based on these current findings, several practical recommendations can be formulated for ISP-grade operational environments.

First, monitoring department should adopt multimodal telemetry collection as a standard practice, since combining flow, routing and device-health signals enables earlier detection of cross-layer anomalies such as congestion buildup or route instability.

Second, device-level indicators (CPU load, interface errors, syslog anomalies) should be incorporated into proactive alerting systems, as they proved to be strong early-warning predictors of severe events.

Third, multimodal signatures should be used to refine correlation rules in monitoring platforms, linking RTT spikes, jitter variability and throughput degradation into unified anomaly patterns. Finally, lightweight multimodal inference modules can be deployed at distributed network locations to reduce detection latency and support real-time decision making.

LIMITATIONS

Despite its promising outcomes, the system has several limitations. The dataset used in the experiments is relatively small and generated in a controlled virtual environment, which does not fully capture the variability and noise of real ISP networks. The current model recognizes only four anomaly categories, and additional classes would require separate training. Long-term temporal dependencies are only partially modeled, and routing-level telemetry such as detailed BGP message streams is not yet fully integrated. As stated in the document, *"the problems caused by imbalanced dataset and incompleteness of logs for some devices are described"*, which further highlights the need for broader data collection.

FUTURE WORK

Future work should therefore focus on expanding the dataset using large-scale containerized topologies and realistic traffic traces, incorporating more advanced deep learning architectures such as attention-based temporal models, and integrating routing-level telemetry including BGP announcements and link-state changes. Another important direction is the

development of an online, real-time detection pipeline suitable for operational NOC environments. The anomaly taxonomy should be extended to include microbursts, link flaps, BGP hijacks and stealthy low-rate attacks, improving the generalization capability of the multimodal framework.

Overall, the system represents a significant step toward intelligent, cross-layer monitoring solutions for ISP backbone networks, offering both scientific value and practical applicability.

REFERENCES

Cisco Systems. (2023). SNMP-based device health monitoring best practices. Cisco Technical Whitepaper.

Containerlab Authors. (2024). Containerlab: Container-based networking labs. https://containerlab.dev/

Dafhalla, M., Al-Dubai, A., & Gebali, F. (2022). Intelligent ICT architectures for nanosensor communication systems: A multilayer AI-driven approach. IEEE Internet of Things Journal, 9(14), 12033–12047.

Dainotti, A., King, A., Claffy, K., Papale, F., & Pescapé, A. (2016). Analysis of BGP instability and routing anomalies. IEEE/ACM Transactions on Networking, 24(5), 2795–2808.

Fontugne, R., Borgnat, P., Abry, P., & Fukuda, K. (2010). MAWILab: Combining diverse anomaly detectors for automated anomaly labeling and performance benchmarking. In Proceedings of ACM CoNEXT.

FRRouting Project. (2024). FRR: A modern open-source routing suite. https://frrouting.org/

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Li, T., Chen, Y., & Xu, K. (2022). Multimodal deep learning for network anomaly detection. IEEE Transactions on Network and Service Management, 19(2), 1250–1264.

Maiti, S., Misra, S., & Obaidat, M. S. (2023). Machine learning–driven caching strategies for Industry 5.0 edge computing: A comprehensive survey. IEEE Systems Journal, 17(1), 345–358.

Prince, A., & Renjith, V. (2022). AI-based detection of signaling anomalies in next-generation mobile networks. IEEE Access, 10, 10345–10358.

Polyzos, G. C., & Papadopouli, M. (2020). Multimodal data fusion for network monitoring and anomaly detection. IEEE Communications Magazine, 58(4), 42–48.

Ouaissa, M., Ouaissa, A., & Tari, A. (2022). Multisource data fusion for anomaly detection in intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems, 23(11), 21045–21057.

RIPE NCC. (2024). BGP routing information and update streams. https://www.ripe.net/

Shafiq, S., Erman, J., Krishnamurthy, A., & Sen, S. (2015). Understanding traffic anomalies in backbone networks. ACM SIGCOMM Computer Communication Review, 45(4), 347–360.

Sperotto, A., Sadre, R., & Pras, A. (2010). Flow-based anomaly detection in large networks. In Proceedings of IEEE AIMS (pp. 1–10).

Tariq, M., Khan, S., & Almogren, A. (2022). Deep learning techniques for UAV cybersecurity: A systematic review. IEEE Communications Surveys & Tutorials, 24(3), 1550–1582.

Vaswani, A., et al. (2017). Attention is all you need. In Proceedings of NeurIPS (pp. 5998–6008).

Xu, K., Zhang, Z., & Bhattacharyya, S. (2005). Profiling Internet backbone traffic: Behavior models and applications. ACM SIGCOMM Computer Communication Review, 35(4), 169–180.

Yu, J., Li, Z., & Zhou, W. (2022). Cross-layer anomaly detection for large-scale backbone networks. IEEE Transactions on Network Science and Engineering, 9(3), 1450–1463.

Zhang, J., & Chen, Y. (2022). Deep learning models for cross-layer anomaly detection in large-scale networks. IEEE Transactions on Information Forensics and Security, 17, 1123–1137.