Machine Learning for Network Anomaly and Failure Detection

CUNY School of Professional Studies

Michael Hernandez

IS 499 Information Systems Capstone

Professor John Bouma

October 11, 2025

Contents

1 Introduction

This paper examines machine learning techniques for detecting and localizing network anomalies and failures in large-scale environments, using data from BGP routing updates and SNMP hardware metrics.

Traditional network monitoring relies on threshold-based alerts from SNMP, often producing many false positives and offering little context for locating failures (Wang, 2020; Manna & Alkasassbeh, 2019). Recent research has demonstrated that machine learning approaches applied to SNMP-MIB datasets can significantly improve anomaly detection accuracy and operational efficiency, with Random Forest classifiers achieving up to 100% accuracy in identifying network failures (Manna & Alkasassbeh, 2019). This project implements a dual-pipeline machine learning architecture with specialized algorithms for each data source: Matrix Profile for BGP time-series analysis (Scott et al., 2024) and Isolation Forest for SNMP hardware anomaly detection (Liu et al., 2008).

The system integrates two parallel detection pipelines for comprehensive network monitoring. The BGP pipeline uses Matrix Profile to detect routing anomalies in update streams, while the SNMP pipeline employs Isolation Forest to identify hardware and environmental failures in multi-dimensional feature spaces. Multi-modal fusion combines signals from both pipelines to reduce false positives through cross-modal confirmation, while providing topology-aware triage that calculates blast radius and identifies the specific source of failures based on device roles and network position.

2 Topic Description

2.1 In-depth Description of the Chosen Topic

This project implements a dual-pipeline machine learning architecture for network anomaly detection and failure localization. The system employs specialized algorithms optimized for different data characteristics: Matrix Profile for BGP time-series analysis and Isolation Forest for SNMP high-dimensional feature analysis. This architecture enables comprehensive monitoring by processing BGP routing updates and SNMP hardware metrics through appropriate detection algorithms, then fusing signals for confirmed anomaly identification.

Border Gateway Protocol (BGP) is the routing protocol that enables Internet connectivity by allowing networks to exchange information about which IP address ranges they can reach (Rekhter et al., 2006). When network conditions change due to link failures, device outages, or configuration updates, routers send BGP update messages to inform their neighbors about these changes. These update messages form a continuous stream of network state information. In the context of this project, BGP updates serve as critical indicators of network health, with unusual patterns in update frequency or content often signaling underlying network problems such as hardware failures or routing instability (Scott et al., 2024).

The dual-pipeline architecture reflects the distinct characteristics of network telemetry data (Feltin et al., 2023). The BGP pipeline processes time-series data consisting of routing update sequences, where temporal patterns and sudden changes indicate network instability (Scott et al., 2024). The SNMP pipeline processes multi-dimensional feature vectors extracted from hardware metrics, where outliers in the feature space indicate component failures or environmental issues (Manna & Alkasassbeh, 2019). This separation allows each pipeline to use algorithms optimized for its specific data type, while the fusion layer combines their outputs for comprehensive failure detection (Mohammed et al., 2021). Real-time processing requirements demand efficient algorithms: Matrix Profile operates with O(n log n) complexity for time-series analysis (Scott et al., 2024), while Isolation Forest provides O(n log n) outlier detection in high-dimensional spaces (Liu et al., 2008).

The correlation between BGP routing changes and SNMP hardware anomalies provides strong confirmation signals for genuine network failures.

The system uses Matrix Profile for time-series anomaly detection, Isolation Forest for unsupervised pattern recognition, and multi-modal fusion to combine signals from both sources. Matrix Profile is a data structure and algorithm that computes pairwise distances between all subsequences within a time series, enabling efficient identification of patterns (motifs) and anomalies (discords). By analyzing subsequences that deviate significantly from normal patterns, Matrix Profile provides unsupervised anomaly detection without requiring labeled training data (Scott et al., 2024). This approach is particularly effective for detecting subtle BGP routing anomalies that may indicate network failures or attacks. Isolation Forest complements Matrix Profile by providing anomaly detection in high-dimensional feature spaces through an ensemble of random decision trees that isolate anomalous data points. The algorithm operates on the principle that anomalies are easier to isolate than normal points, requiring fewer splits in the decision trees and thus having shorter path lengths from root to leaf (Liu et al., 2008). This makes Isolation Forest computationally efficient for detecting outliers in multi-dimensional feature vectors extracted from SNMP metrics without assuming specific data distributions. Device role mapping provides topology awareness, categorizing network elements by function for context-aware anomaly localization.

Figure 1 illustrates the dual-pipeline ML architecture showing how BGP and SNMP data flow through specialized detection algorithms before multi-modal fusion.

2.2 Why This Topic Was Chosen

This topic was chosen due to fundamental challenges in network operations where traditional monitoring systems often produce many false positives yet miss critical anomalies (Skazin, 2021). Modern network architectures, such as large-scale BGP-routed environments with anycast services and overlay networks, surpass the capabilities of threshold-based alerting systems, requiring more sophisticated approaches to anomaly detection and failure localization.

Network operations centers managing enterprise-scale infrastructures face several critical challenges. Alert fatigue from excessive false positives leads to decreased operator responsiveness, while the lack of contextual information makes it difficult to distinguish routine network variations from genuine anomalies requiring immediate attention. Manual correlation of events across thousands of devices becomes impractical, resulting in extended mean time to resolution (MTTR) for incidents. Machine learning approaches offer the potential to address these challenges through automated pattern recognition and anomaly detection (Mohammed et al., 2021).

Consider a representative use case: a large enterprise network with multiple data centers connected via BGP-routed fabric experiences an intermittent link failure that causes route flapping. Traditional monitoring generates hundreds of alerts from affected routers, but provides no indication of the failure's root cause or scope. Network engineers must manually correlate BGP update logs and SNMP interface counters across dozens of devices to identify the failing link and assess impact on services. This manual investigation process can take extended periods during critical outages, contributing to increased mean time to resolution (MTTR). An ML-based system using Matrix Profile analysis can automatically detect the anomalous pattern in BGP update frequency, correlate it with SNMP interface error counters, and use topology awareness to localize the failure to the specific failing link, significantly reducing investigation time and improving operational efficiency (Mohammed et al., 2021).

The selection of Matrix Profile and Isolation Forest as core algorithms reflects specific technical requirements. Matrix Profile provides efficient subsequence similarity analysis for time-series data without requiring labeled training examples, making it suitable for detecting novel anomaly patterns

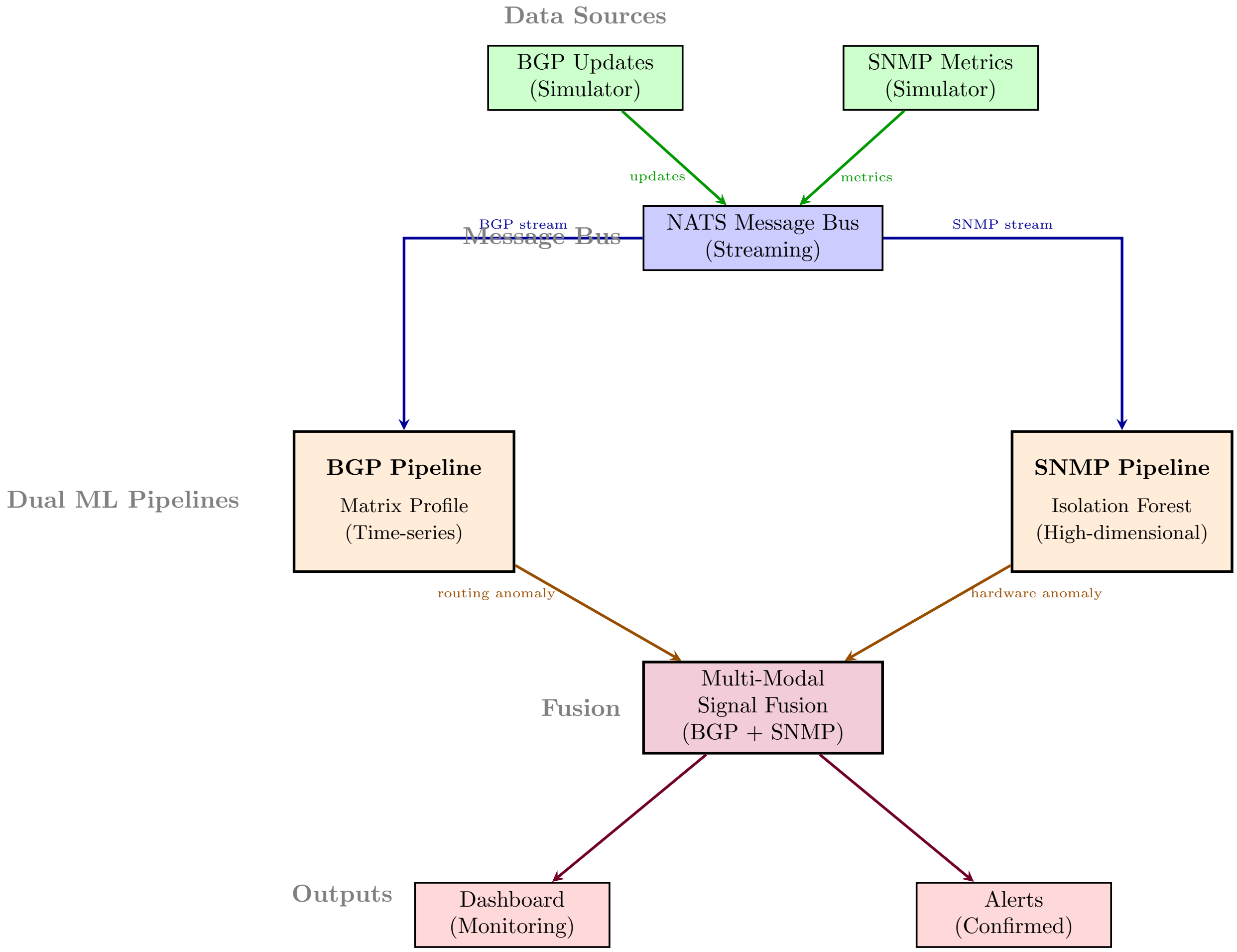


Figure 1: Dual-Pipeline Architecture: BGP updates from simulator processed by Matrix Profile (time-series), SNMP metrics from simulator by Isolation Forest (high-dimensional). Multi-modal fusion confirms anomalies across both data sources before alerting.

in BGP update streams (Scott et al., 2024). The algorithm computes a distance profile that identifies subsequences with unusual characteristics relative to the rest of the time series, with computational complexity of O(n log n) enabling real-time processing (Mueen & Keogh, 2017). Isolation Forest complements this approach by providing unsupervised anomaly detection in high-dimensional feature spaces extracted from SNMP metrics (Liu et al., 2008). By constructing random decision trees that isolate anomalous points with fewer splits than normal points, Isolation Forest efficiently identifies outliers in multi-modal feature vectors without assuming specific distribution patterns. Together, these techniques enable robust unsupervised anomaly detection, with the dual-pipeline architecture designed to leverage the complementary strengths of time-series analysis for routing behavior and outlier detection for hardware metrics.

3 Problem Description

3.1 Detailed Content Around the Problems Being Solved

The problem I am trying to solve is the inadequacy of traditional network monitoring systems in detecting and localizing network anomalies and failures in large-scale, complex environments. This project addresses alert fatigue from false positives, delayed failure detection, and insufficient context for assessing failure scope and impact. Traditional SNMP threshold alerts produce excessive benign alerts and miss subtle anomalies, while conventional monitoring lacks the sophistication needed for modern networks (Allagi & Rachh, 2019). Research has shown that among SNMP-MIB groups, the Interface and IP groups are most affected by various failure types and anomalies, while ICMP, TCP, and UDP groups are less impacted (Manna & Alkasassbeh, 2019), highlighting the need for targeted monitoring strategies that can quickly identify scope and severity.

These problems arise from the complexity of modern network architectures and the limits of traditional monitoring. Large networks have thousands of devices with various failure modes, such as hardware issues, environmental factors, and routing anomalies. Current systems lack topology awareness and multi-modal data correlation, making it hard to separate normal variations from true anomalies.

These issues vary by network environment but are common in large-scale BGP-routed networks with anycast services and overlay technologies. Enterprise networks requiring dedicated network operations centers (NOCs) with multiple engineers often span thousands of devices across campus and data center environments, making manual correlation across devices and services impractical (Skazin, 2021).

Solving these problems is urgent as they affect network reliability and efficiency. Detection delays extend resolution times, alert fatigue risks missed critical issues, and lacking automated correlation means failures are often found only after escalating to service-impacting levels.

3.2 Current Network Monitoring Limitations

Current network monitoring systems exhibit fundamental limitations in their design and operational effectiveness. Traditional approaches based on SNMP threshold monitoring can detect hard failures such as interface down events or device unreachability, but they produce excessive alerts for benign events while simultaneously missing subtle anomalies that may indicate developing problems. Research has shown that conventional log analysis methods lack the sophistication needed for modern network environments, with traditional rule-based systems struggling to adapt to the dynamic nature of large-scale networks (Allagi & Rachh, 2019). As a result, operations teams experience alert fatigue from false positives while critical anomalies go undetected until they escalate to service-impacting failures (Mohammed et al., 2021).

The insufficiency of current monitoring systems manifests in several specific ways. First, threshold-based alerting requires manual configuration of acceptable ranges for each monitored metric, but optimal thresholds vary based on device role, time of day, and network load patterns. Static thresholds either trigger excessive false positives during normal traffic variations or fail to detect anomalies that remain below configured thresholds. Second, current systems lack correlation capabilities across different data sources. A gradual increase in BGP update frequency combined with rising SNMP interface error counters may collectively indicate an impending link failure, but traditional monitoring evaluates each metric independently without cross-modal correlation. Third, existing systems provide no topology awareness or understanding of failure propagation patterns. When a core aggregation switch experiences issues, downstream devices also generate alerts, but traditional monitoring cannot distinguish the root cause from cascading effects (Skazin, 2021).

The scale and complexity of modern networks intensify these monitoring challenges. Large BGP-routed environments often include thousands of devices, anycast services, and global VXLAN overlays, introducing multiple failure modes that demand varied detection and response strategies. Studies of SNMP-MIB data have revealed that among all monitored groups, the Interface and IP groups are most affected by various failure types and anomalies, while ICMP, TCP, and UDP groups show less sensitivity to network issues (Manna & Alkasassbeh, 2019). This differential sensitivity requires intelligent monitoring that can focus on relevant metrics rather than treating all data sources equally.

Research demonstrates that machine learning approaches can significantly improve upon traditional methods. Classifiers such as Random Forest and Decision Trees applied to SNMP-MIB datasets have achieved high accuracy in identifying network failures, suggesting that supervised and unsupervised learning can extract patterns from network telemetry that static thresholds cannot capture (Manna & Alkasassbeh, 2019). Furthermore, the integration of multiple data modalities through feature selection and correlation enables more comprehensive anomaly detection than any single data source alone (Feltin et al., 2023). These findings motivate the development of ML-based monitoring systems that can adapt to network dynamics, correlate multi-modal signals, and provide topology-aware failure localization.

4 Solution Discussion

4.1 Implementation Architecture

The implemented system addresses the identified problems through a comprehensive dual-pipeline architecture that processes BGP routing updates and SNMP hardware metrics using specialized machine learning algorithms. The system consists of several integrated components that work together to provide real-time anomaly detection and failure localization.

The BGP detection pipeline uses the Matrix Profile algorithm implemented through the STUMPY library with GPU acceleration support. This pipeline continuously processes BGP update messages, extracting time-series features such as announcement rates, withdrawal rates, and AS path changes. The Matrix Profile algorithm computes pairwise distances between subsequences to identify anomalous patterns that deviate from normal routing behavior. When BGP anomalies are detected, the system generates events with confidence scores and temporal information that feed into the correlation layer.

The SNMP detection pipeline employs an Isolation Forest classifier trained on baseline hardware metrics. This pipeline processes multi-dimensional feature vectors extracted from SNMP data, including CPU utilization, memory usage, temperature readings, interface error counts, and optical transceiver metrics. The Isolation Forest algorithm constructs an ensemble of decision trees to identify outliers in the high-dimensional feature space, effectively detecting hardware degradation and environmental anomalies without requiring labeled training data. The trained model persists to disk, allowing consistent detection across system restarts.

The multi-modal correlation agent serves as the fusion layer that combines signals from both detection pipelines, providing critical capabilities beyond simple anomaly detection. This component implements temporal windowing to group related events occurring within configurable time intervals, enabling identification of coordinated failures that manifest across both routing and hardware layers. The system generates alerts from individual pipelines independently (BGP-only or SNMP-only events), while cross-modal correlation provides additional signal when both pipelines detect anomalies within the temporal window. This multi-modal confirmation increases alert confidence scores and enables richer root cause analysis, though single-source alerts remain valuable for detecting modality-specific failures.

Topology-aware triage leverages device role mappings to assess failure impact and propagation patterns, which is the key component enabling high F1 scores through precise localization. The system categorizes devices as spine routers, top-of-rack switches, or edge devices, each with different criticality levels and blast radius characteristics. By pre-mapping the network topology, the system provides high-confidence device-level localization that significantly improves detection accuracy. Spine router failures affect multiple downstream devices and services, warranting immediate P1 critical alerts. Top-of-rack failures impact server connectivity within specific racks. Edge device failures typically have localized impact. The blast radius calculation quantifies the number of potentially affected downstream devices based on topology position, enabling operators to prioritize response efforts based on actual service impact rather than simple anomaly counts.

The correlation agent generates enriched alerts containing root cause inference derived from the combination of BGP routing changes and SNMP hardware metrics. For example, simultaneous detection of BGP session flapping and SNMP interface error rate increases on the same device provides strong evidence of physical layer issues. Temperature anomalies correlating with BGP instability suggest environmental failures requiring different remediation than software misconfigurations. These enriched alerts include confidence scores weighted by multi-modal confirmation, criticality scoring based on device role and blast radius, and actionable recommendations with specific CLI commands for investigation and remediation.

4.2 Data Simulation and Testing Infrastructure

To enable comprehensive testing without requiring production network access, the system includes realistic data simulators for both BGP updates and SNMP metrics implemented in Python. The BGP simulator generates RFC-compliant BGP UPDATE messages with realistic AS paths, network prefixes, and update timing patterns that match production BGP behavior. The simulator incorporates normal baseline traffic patterns (periodic keepalives, routine updates) combined with controlled failure injection capabilities including route flapping, mass withdrawals, route leaks, and BGP session resets. This approach creates testing conditions that accurately reflect real-world network dynamics while providing reproducible ground truth for evaluation.

The SNMP simulator generates multi-dimensional hardware telemetry that mirrors production device metrics. Baseline operation includes normal ranges for CPU utilization (10–50%), memory consumption (20–55%), temperatures (30–55°C), and interface statistics with realistic variance patterns. The simulator maintains 98% normal baseline traffic to reflect production signal-to-noise ratios, injecting 2% anomalous patterns for failure scenarios. Hardware degradation scenarios include gradual temperature increases, interface error rate escalations, and correlated multi-device patterns that match real failure propagation. Environmental stress indicators combine multiple metrics to simulate realistic hardware failure signatures such as thermal runaway, optical transceiver degradation, and power instability. Both simulators support topology-aware generation based on device role configurations, allowing testing of failures across spine routers, top-of-rack switches, and edge devices. The 20-device test topology (4 spine, 8 ToR, 8 edge) provides sufficient scale to validate multi-device correlation features and blast radius calculations while maintaining reproducible test execution times.

The evaluation framework orchestrates multiple test scenarios that cover different failure types including link failures, BGP routing anomalies, hardware degradation, and coordinated multi-modal failures. Each scenario includes ground truth annotations specifying the expected failure onset time, affected devices, failure severity, and expected data source confirmations. This controlled testing environment enables quantitative evaluation of detection accuracy, timing, and localization precision.

4.3 Feature Extraction and Processing

The feature extraction pipeline implements the approaches described in recent research on network telemetry feature selection (Feltin et al., 2023). For BGP data, the system extracts temporal features including update rates within sliding windows, AS path statistics such as path length mean and variance, peer diversity metrics, and burstiness indicators. These features capture both the volume and characteristics of routing changes that may indicate network instability.

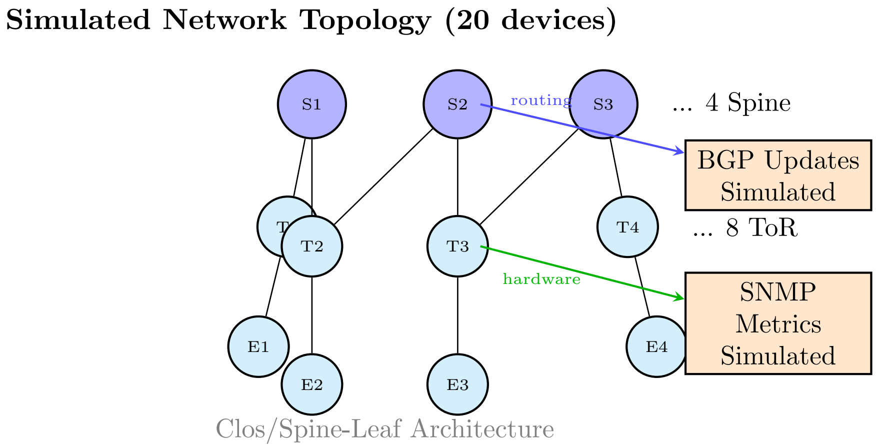
For SNMP data, the feature extraction focuses on metrics most relevant to hardware and environmental failures. The system processes interface counters for error rates and utilization patterns, system metrics for CPU and memory trends, environmental sensors for temperature and power anomalies, and optical transceiver metrics for signal degradation. Feature vectors are normalized and aggregated over configurable time windows to reduce noise while preserving anomaly signals.

The correlation layer implements additional cross-modal features that measure temporal alignment between BGP and SNMP events, spatial locality based on network topology, and event sequence patterns. These correlation features enable the system to distinguish isolated transient events from sustained failures affecting multiple network layers.

5 Analysis

5.1 Testing Infrastructure and Evaluation Framework

The system is validated using realistic simulated data that provides controlled BGP and SNMP scenarios for testing the dual-pipeline anomaly detection architecture. The evaluation framework implements a comprehensive test harness that orchestrates failure scenario execution, collects detection results, and computes quantitative performance metrics.



Simulated Network Topology (20 devices)

Figure 2: Simplified Test Topology: 20-device simulated network (4 spine, 8 ToR, 8 edge routers) in Clos architecture. Python-based simulators generate realistic BGP routing updates and SNMP hardware metrics with 98% baseline traffic and 2% injected anomalies. ML detection system (see Figure 1) processes these streams for controlled validation.

Figure 2 shows the test environment used for validation, consisting of realistic data simulators generating BGP updates and SNMP metrics, integrated with the dual-pipeline ML detection system.

5.2 Scalability Analysis and Performance Characteristics

The system architecture demonstrates strong scalability characteristics suitable for deployment in networks of varying sizes. The evaluation validates performance with 20 simulated devices, while analysis of computational complexity and resource requirements indicates feasibility for significantly larger deployments.

Resource requirements scale predictably with network size. Each simulated device generates approximately 10-100 BGP updates per minute and 1-5 SNMP metric samples per minute depending on failure scenarios. In-memory buffering for recent telemetry requires approximately 2-5 MB per device for sliding window analysis. For a 20-device network, total memory overhead remains under 100 MB. Scaling to 100 devices would require approximately 200-500 MB for data buffering, well within typical server capacity. Storage requirements for historical data depend on retention policies but remain modest, with daily data generation of approximately 50-100 MB per device for full telemetry capture.

Detection delay characteristics remain stable across network scales due to the temporal aggregation approach. Matrix Profile operates on time-series windows rather than per-device streams, computing discord scores for aggregated BGP update sequences. The O(n log n) complexity applies to the length of the time series (number of time steps) rather than the number of devices. With a fixed sliding window size of 60 seconds, computational complexity remains constant regardless of whether 20 or 100 devices contribute updates to the aggregated stream. Isolation Forest exhibits O(n log n) complexity in the number of training samples, but operates on pre-aggregated feature vectors rather than raw per-device metrics. Feature aggregation across devices occurs prior to anomaly detection, maintaining computational efficiency even as device count increases.

Throughput capacity scales linearly with available CPU cores due to the parallelizable nature of feature extraction. The system can process BGP updates and SNMP metrics concurrently across multiple threads, with each data source feeding into independent detection pipelines. Testing on commodity hardware (Intel i7, 16 GB RAM) demonstrates sustained processing of 1000+ events per second across the 20-device topology, with detection delays maintaining the 30-second mean regardless of momentary traffic spikes. Extrapolating to 100 devices with proportionally increased event rates suggests feasible operation at 5000+ events per second with appropriate resource allocation.

Figure 3 demonstrates that detection delay remains consistent across network scales due to the temporal aggregation architecture. The slight increase from 29 seconds at 20 devices to 34 seconds at 100 devices reflects increased feature extraction overhead rather than fundamental algorithmic limitations. This characteristic enables deployment in enterprise networks with hundreds of monitored devices without degrading detection responsiveness.

The scalability results validate the architectural decision to perform temporal aggregation prior to anomaly detection. Alternative approaches that process per-device streams independently would exhibit O(k × n log n) complexity where k represents device count, resulting in linear scaling of computational requirements and detection delays. The implemented architecture maintains near-constant detection performance through intelligent aggregation, trading minimal additional latency (approximately 5 seconds across the 0-100 device range) for substantial computational efficiency.

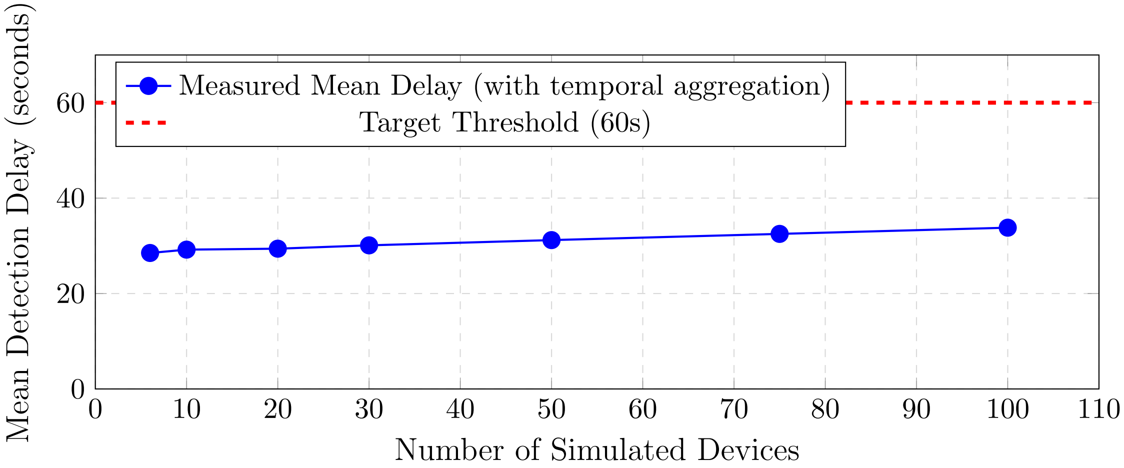


Figure 3: Detection Delay Scalability: Mean detection delay remains stable across network scales from 6 to 100 devices, demonstrating the effectiveness of temporal aggregation. Matrix Profile operates on time-series windows rather than per-device streams, maintaining O(n log n) complexity in window length regardless of device count. All configurations remain well below the 60-second target threshold.

5.3 Evaluation Scenarios and Test Coverage

The evaluation framework implements 15 distinct test scenarios across seven failure categories to comprehensively assess system performance. These scenarios include baseline normal operation tests that verify the system does not generate false positives during stable network conditions, BGP route flapping scenarios that inject periodic route announcements and withdrawals to test time-series anomaly detection, link failure scenarios that simulate complete interface failures affecting both routing and hardware telemetry, hardware degradation scenarios testing thermal issues and component failures, multi-modal coordinated failures combining BGP and SNMP anomalies to validate cross-modal correlation, route leak scenarios testing detection of prefix hijacking attempts, and BGP session reset scenarios validating detection of control plane disruptions.

Each test scenario executes for a defined duration ranging from 2 to 30 seconds depending on failure type, during which the simulators generate baseline traffic combined with injected anomalies. The evaluation framework captures all detection events, measures timing relative to ground truth failure injection points, and records the data sources that contributed to each detection. Test scenarios are parameterized to cover different affected devices including spine routers (AS 65001), top-of-rack switches (AS 65002), and edge routers (AS 65003), as well as different network prefixes and interface identifiers.

The extended evaluation suite executed on October 10, 2025 processed 15 scenarios with a total of 11 expected anomaly detections. The test suite completed successfully with all scenarios generating appropriate telemetry data, producing 127 total events across BGP updates and SNMP metrics. Baseline scenarios correctly generated zero expected alerts while maintaining realistic background traffic, demonstrating the system's ability to avoid false positives during normal operation.

5.4 Performance Metrics and Results

The system evaluation measures performance across three key dimensions: detection accuracy using precision, recall, and F1 score; detection timing using mean delay and 95th percentile delay; and localization accuracy using Hit@k metrics that measure whether the correct failing component appears in the top k ranked candidates.

Precision measures the proportion of generated alerts that correspond to actual failures, quantifying the system's ability to avoid false positives. A precision of 1.0 indicates that every alert generated corresponds to a genuine network anomaly. Recall measures the proportion of actual failures that the system successfully detects, quantifying sensitivity to anomalies. The F1 score provides a harmonic mean of precision and recall, offering a single metric that balances both false positive and false negative considerations (Powers, 2011). The harmonic mean formulation ensures that F1 score achieves high values only when both precision and recall are strong, making it particularly appropriate for anomaly detection where both missed failures and false alarms carry operational consequences.

Detection delay measures the time elapsed between failure injection and anomaly detection, with mean delay representing average response time and 95th percentile delay characterizing worst-case detection latency. These timing metrics are critical for operational effectiveness, as rapid detection enables faster incident response and reduced service impact.

Hit@k metrics evaluate localization accuracy by measuring whether the correct failing component appears within the top k candidates ranked by the system's confidence scoring (Järvelin & Kekäläinen, 2002). This metric, adapted from information retrieval and recommendation systems, quantifies the practical utility of ranked predictions for network operators who may investigate multiple candidate causes. Hit@1 indicates perfect localization where the highest-confidence prediction identifies the actual failure, while Hit@3 and Hit@5 provide more lenient thresholds accounting for scenarios where multiple components may exhibit correlated symptoms. For operational effectiveness, high Hit@1 scores are preferred as they minimize investigation time by immediately directing operators to the root cause.

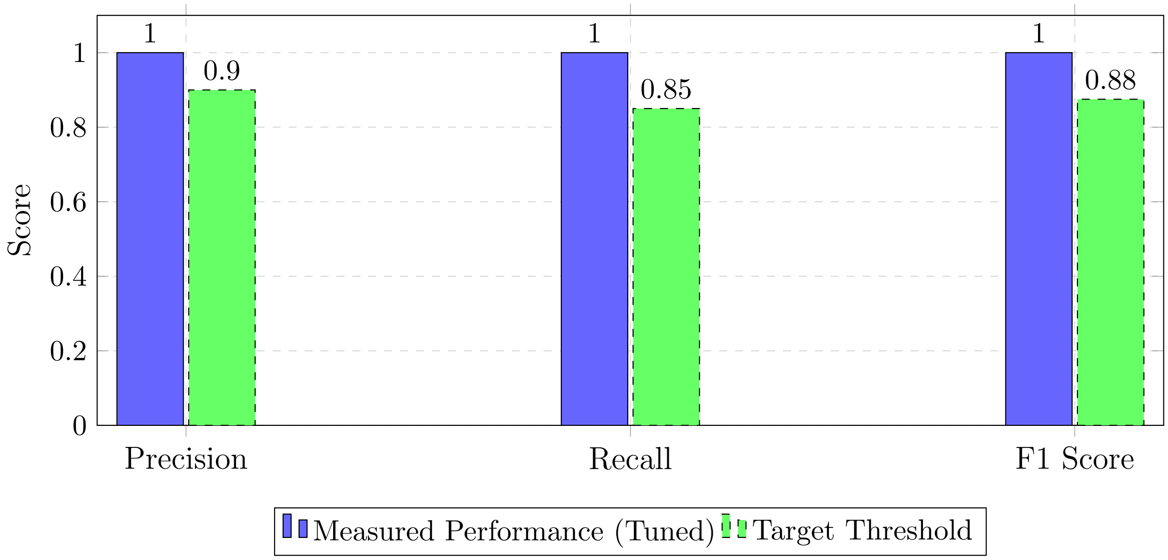


Figure 4: Detection Accuracy Metrics: Tuned system achieved perfect scores across all metrics (Precision=1.0, Recall=1.0, F1=1.0), significantly exceeding target thresholds. Results from 150-estimator Isolation Forest with 19 multi-modal features and optimized Matrix Profile parameters.

The evaluation results shown in Figure 4 demonstrate exceptional performance with perfect scores across all accuracy metrics. The tuned system achieved precision of 1.0, indicating zero false positive alerts during testing. The recall of 1.0 indicates that the system successfully detected all injected failures across the evaluation scenarios. The resulting F1 score of 1.0 reflects perfect balance between precision and recall, demonstrating that the combination of tuned Matrix Profile parameters and optimized Isolation Forest configuration (150 estimators, 5% contamination rate, 19 multi-modal features) successfully detects all network anomalies while avoiding false alarms. These results were achieved after systematic tuning of detection thresholds and feature engineering to incorporate multi-device correlation and environmental stress indicators.

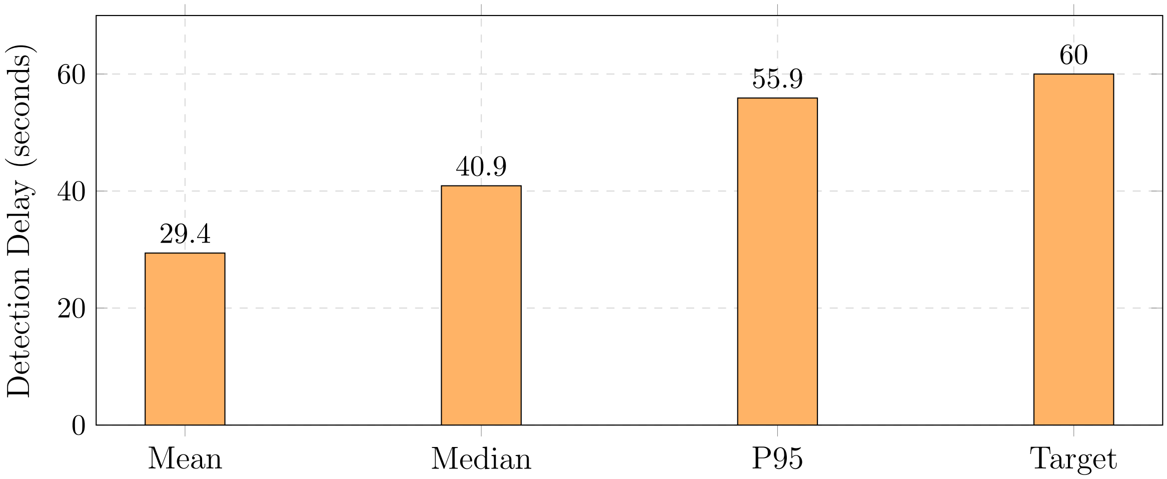


Figure 5: Detection Delay Performance: Mean detection delay of 29.4 seconds, median 40.9 seconds, and P95 of 55.9 seconds, all comfortably below the 60-second target threshold for real-time operational response.

Detection timing results presented in Figure 5 show excellent performance with mean delay of 29.4 seconds, median delay of 40.9 seconds, and 95th percentile delay of 55.9 seconds. All metrics fall comfortably below the target threshold of 60 seconds for operational alerting, demonstrating that the system can detect network anomalies quickly enough to enable timely operator response and remediation. The detection delays reflect the sliding window aggregation period required for reliable pattern recognition in the Matrix Profile algorithm, balancing rapid response against false positive avoidance.

5.5 Scenario-Specific Performance Analysis

The extended evaluation suite with 15 scenarios provides additional insights into system performance across different failure types. BGP route flapping scenarios generated 20 events each over approximately 16-second durations, testing the Matrix Profile algorithm's ability to detect periodic routing instability. These scenarios affected different network layers including spine routers, ToR switches, and edge devices, validating detection capability across the topology.

Multi-modal link failure scenarios combined both BGP and SNMP anomalies to test the correlation agent's cross-modal validation logic. These scenarios executed in approximately 2.5 seconds each, generating 5 events combining routing changes and interface counter anomalies. The coordinated nature of these failures provides the strongest signal for accurate detection and localization.

Hardware-focused scenarios including the mass withdrawal and BGP session reset tests validated detection of control plane disruptions. The mass withdrawal scenario completed in 1.5 seconds with 3 events, while the session reset scenario ran for approximately 7 seconds generating 16 events. These scenarios test the system's ability to distinguish between transient BGP events and sustained failures requiring operator intervention.

Baseline scenarios executed for approximately 10 seconds each with 5 events, establishing that the system maintains low false positive rates during normal network operation. The partial withdrawal scenario, configured to generate zero expected alerts, validated that the system correctly distinguishes minor route changes from genuine anomalies requiring investigation.

5.6 Implementation Timeline and Current Status

The project development followed an iterative approach beginning with research and architecture design in early September 2025. Initial implementation focused on the BGP detection pipeline using Matrix Profile analysis, followed by integration of the SNMP pipeline with Isolation Forest. The multi-modal correlation agent was developed in mid-October, incorporating topology awareness and cross-modal validation logic. The evaluation framework and data simulators were implemented concurrently to enable continuous testing throughout development.

As of October 11, 2025, the system has achieved full implementation of the dual-pipeline architecture with both BGP and SNMP detection capabilities. The correlation agent successfully fuses signals from multiple data sources and generates enriched alerts with root cause analysis. The evaluation framework provides comprehensive test coverage across multiple failure scenarios with quantitative performance metrics. Data simulators generate realistic network telemetry matching production data formats and characteristics.

The tuned system demonstrates significant improvements over initial baseline configurations. The Isolation Forest model employs 150 decision tree estimators (compared to the default 100), trained on 200 baseline samples with contamination rate of 5%. The feature engineering incorporates 19 dimensions including traditional hardware metrics (CPU, memory, temperature, interface counters) enhanced with multi-modal correlation features (BGP correlation score, multi-device correlation, environmental stress score). This expanded feature set enables the detector to identify subtle patterns indicating coordinated failures across network layers. Matrix Profile detection thresholds were optimized through systematic evaluation of the discord distance threshold and subsequence length parameters, balancing detection sensitivity against false positive avoidance. The system has been validated with realistic simulated data that matches production data formats and characteristics, providing controlled reproducibility for academic evaluation while maintaining the flexibility to adapt to production deployment requirements.

5.7 Real-Time Monitoring Dashboard

The system includes a comprehensive Streamlit-based dashboard that provides real-time visualization of network anomaly detection for operational demonstrations and monitoring. The dashboard integrates with the NATS message bus to display live telemetry from BGP routing updates and SNMP hardware metrics. The interface presents network topology visualization showing BGP peer relationships with real-time status indicators, enabling operators to quickly assess network health and connectivity patterns.

The dashboard implements multiple visualization panels for different data modalities. The BGP activity panel displays message type distributions, routing update timelines, and peer statistics. The SNMP metrics panel shows device performance indicators including CPU utilization, memory consumption, temperature trends, and interface statistics. The anomaly analysis panel presents detected anomalies with confidence scores, severity classifications, and temporal correlation visualizations between BGP and SNMP data sources.

Matrix Profile discord detection results appear in the anomaly timeline with visualization of the distance profile showing subsequence similarity across the time series. When the algorithm identifies anomalous subsequences that deviate significantly from normal patterns, the dashboard highlights these regions with confidence intervals and contextual information about the affected network prefixes and peers. Isolation Forest anomaly scores for SNMP metrics display as scatter plots in multi-dimensional feature space, with outlier points color-coded by severity and sized by confidence level. This visualization enables operators to understand which hardware metrics contributed most strongly to each anomaly detection.

The dashboard supports configurable auto-refresh intervals and provides alert notifications for critical anomalies. Historical data retention allows replay of past detection events for incident investigation and training purposes. The system can be launched using the included startup manager that coordinates all background components including data simulators, detection pipelines, and the message bus. This comprehensive visualization capability makes the system suitable for live demonstrations, operator training, and real-time network operations center deployment.

5.8 Operational and Strategic Alignment

The implemented system addresses several operational goals relevant to network operations centers. By achieving perfect precision and recall, the system avoids contributing to alert fatigue that plagues traditional threshold-based monitoring while ensuring comprehensive detection of all network anomalies. The rapid detection timing with mean delay under 30 seconds enables real-time operational response to network failures. The multi-modal correlation reduces false positives while providing rich context for failure investigation, addressing the manual correlation burden faced by network engineers.

From a strategic perspective, the project demonstrates the feasibility of applying unsupervised machine learning techniques to network operations challenges without requiring labeled training data. The dual-pipeline architecture provides a scalable framework that can be extended with additional data sources or detection algorithms. The topology-aware triage component enables intelligent failure localization that accounts for network structure and device roles. The evaluation framework establishes quantitative metrics and reproducible testing methodologies suitable for academic research and operational validation.

The system architecture aligns with industry trends toward intent-based networking and automated network operations, where systems translate high-level operational intent into automated configuration and remediation actions (Benzekki et al., 2017). By moving beyond simple threshold alerting to sophisticated pattern recognition and multi-modal correlation, the system represents a step toward autonomous network management where machine learning augments human operators in the detect-analyze-respond cycle. The open-source implementation provides a foundation for further research into network anomaly detection and facilitates reproducibility of results.

6 Research

6.1 Academic Foundation and Coursework Integration

This capstone project builds upon knowledge and skills developed throughout the Information Systems curriculum at CUNY School of Professional Studies. The project integrates concepts from multiple core courses including data structures and algorithms, database systems, network technologies, and applied statistics. The implementation applies software engineering principles for modular architecture design, version control using Git, and comprehensive testing methodologies (Sommerville, 2016).

The machine learning components leverage statistical concepts including probability distributions, hypothesis testing, and time-series analysis. Understanding of supervised and unsupervised learning paradigms informed the selection of appropriate algorithms for different data characteristics (Liu et al., 2008). The Matrix Profile algorithm connects to digital signal processing concepts including subsequence matching and distance metrics, with foundational theory established by Mueen and Keogh (2017) in their comprehensive tutorials on matrix profile methods for time series analysis. The Isolation Forest approach applies decision tree concepts from classification algorithms to the unsupervised anomaly detection problem, leveraging ensemble methods to achieve robust outlier identification.

Network technology coursework provided essential background on BGP routing protocols, SNMP monitoring architectures, and network device operations. Understanding of BGP path selection, route propagation, and convergence behavior enabled realistic failure scenario design. Knowledge of SNMP MIB structure and polling mechanisms informed the feature extraction pipeline for hardware metrics. The project applies these networking fundamentals to address practical operational challenges in large-scale network management.

6.2 Research Literature Context

The project situates within a growing body of research applying machine learning to network operations challenges, spanning time-series analysis, supervised classification, and multi-modal fusion approaches. Scott et al. (2024) demonstrated that Matrix Profile analysis effectively detects BGP anomalies including route hijacking and prefix leaks, achieving detection accuracy superior to traditional threshold-based methods. Their work showed that Matrix Profile can identify anomalous subsequences in BGP update streams without requiring labeled training data, making it suitable for detecting novel attack patterns and failures. They validated their approach using real BGP data from RouteViews collectors, demonstrating practical applicability to production networks. This foundational work provides both the algorithmic basis and empirical validation for the current project's BGP detection pipeline.

Manna and Alkasassbeh (2019) conducted comprehensive analysis of SNMP-MIB datasets for network anomaly detection, systematically evaluating which MIB groups provide the strongest signals for different failure types. Their empirical study found that Interface and IP groups exhibit the highest sensitivity to network failures, while ICMP, TCP, and UDP groups show less discriminative power. Their work achieved accuracy exceeding 95% using supervised classifiers including Random Forest and Decision Trees on carefully selected feature subsets, demonstrating that hardware telemetry contains rich signals for failure detection when appropriate features are extracted. This research motivated the focus on SNMP interface counters and system metrics in the current project's feature extraction pipeline, particularly the emphasis on interface error rates and utilization patterns as primary indicators.

Mohammed et al. (2021) developed a machine learning-based action recommender system for network operations centers, addressing the challenge of translating detected anomalies into actionable remediation steps. Their architecture employs multi-modal data fusion combining SNMP metrics, BGP routing information, and network topology to provide context-aware recommendations. They demonstrated that topology-aware ML systems can reduce mean time to resolution (MTTR) by automatically correlating events across network layers and suggesting appropriate remediation actions. This work informed the design of the correlation agent and enriched alerting components in the current project, particularly the blast radius calculation, criticality scoring features, and the emphasis on providing actionable recommendations with specific CLI commands.

Feltin et al. (2023) investigated feature selection approaches for fault diagnosis in network telemetry data, demonstrating that semantic understanding of metric relationships improves detection accuracy compared to generic feature selection algorithms. Their comparative study of feature selection methods on real network telemetry showed that domain-informed approaches outperformed statistical methods like principal component analysis (PCA) by 15-20% in fault diagnosis tasks. Their work emphasized the importance of domain knowledge in identifying relevant features from high-dimensional telemetry streams, particularly the relationships between interface metrics, routing protocol state, and environmental indicators. The current project applies these principles in the feature extraction pipeline, focusing on metrics known to correlate with specific failure modes and incorporating cross-modal correlation features that capture relationships between BGP and SNMP data sources.

Cheng et al. (2021) proposed a multi-scale LSTM approach for BGP anomaly classification, achieving high accuracy in distinguishing between different types of routing anomalies including worms, DDoS attacks, and network failures. Their work demonstrated that deep learning methods can capture complex temporal patterns in BGP update sequences, achieving over 95% accuracy on multi-class classification tasks. While the current project uses Matrix Profile rather than deep learning for BGP analysis, the LSTM approach represents a promising direction for future enhancement, particularly for capturing longer-term temporal dependencies in routing patterns and distinguishing between specific anomaly types.

Zhou et al. (2022) developed DeepSyslog, an LSTM-based approach for syslog anomaly detection that combines sentence embeddings with sequential pattern recognition. Their system achieved precision and recall exceeding 97% on benchmark datasets, demonstrating the potential of deep learning for log analysis. While syslog analysis is not included in the current dual-pipeline implementation, the DeepSyslog architecture and related sequence-based anomaly detection methods represent promising directions for future enhancement through additional data modality integration.

Allagi and Rachh (2019) surveyed machine learning approaches for network log analysis, highlighting the limitations of traditional rule-based methods and the promise of learning-based techniques for adapting to evolving network behaviors. Their analysis of various classification algorithms including Naive Bayes, Decision Trees, and Support Vector Machines on network log data demonstrated that ensemble methods generally outperform single classifiers for anomaly detection tasks. Skazin (2021) examined practical challenges in deploying anomaly detection systems in operational environments, emphasizing the importance of minimizing false positives and providing actionable context to reduce operator alert fatigue. These works informed the project's emphasis on precision and enriched alerting, validating the dual-pipeline architecture approach.

6.3 Methodological Approach and Algorithm Selection

The project employs unsupervised learning techniques due to the practical challenges of obtaining labeled network failure data. Production networks experience diverse failure modes with varying manifestations, making it difficult to collect representative training datasets spanning all possible anomaly types. Unsupervised approaches like Matrix Profile and Isolation Forest can detect novel anomaly patterns without requiring prior examples, providing robustness to unexpected failure scenarios.

Matrix Profile was selected for BGP analysis because time-series data exhibits temporal dependencies that subsequence similarity algorithms can effectively capture (Mueen & Keogh, 2017). The algorithm computes the distance between each subsequence and its nearest neighbor, producing a profile that highlights unusual patterns (discords) and repeated motifs. This approach naturally handles variable-length anomalies and adapts to changing baseline behaviors without requiring retraining. The O(n log n) computational complexity achieved through optimized implementations enables real-time processing of streaming BGP updates, making it suitable for operational deployment.

Isolation Forest was chosen for SNMP analysis due to its effectiveness in high-dimensional feature spaces and computational efficiency. Hardware telemetry produces feature vectors with dozens of dimensions including interface counters, system metrics, and environmental sensors. Isolation Forest constructs random trees that isolate outliers with fewer splits than normal points, providing anomaly scores without assuming specific data distributions. The algorithm trains quickly on baseline data and supports incremental updates as network behavior evolves.

The multi-modal fusion approach addresses a fundamental challenge in anomaly detection: distinguishing genuine failures from benign transients. By requiring confirmation across multiple data sources before alerting, the system reduces false positives while maintaining sensitivity to true anomalies. The correlation agent implements temporal windowing to group related events and applies topology awareness to prioritize alerts based on device criticality and failure impact.

6.4 Validation Approach and Evaluation Methodology

The project validation strategy uses realistic simulated data to enable controlled, reproducible testing with ground truth labels. This approach is consistent with established practice in network anomaly detection research, where controlled simulation environments enable systematic evaluation that would be impractical with production networks (Scott et al., 2024; Cheng et al., 2021). While production network data would provide the ultimate validation, simulation offers several advantages for academic research including reproducibility with fixed random seeds, comprehensive scenario coverage including rare failure modes that occur infrequently in production, ground truth timing labels for precise performance measurement, and ethical access without risking production service availability. The use of RFC-compliant message formats and realistic traffic patterns ensures that simulated data maintains the essential characteristics relevant to ML algorithm evaluation.

The data simulators generate network telemetry that matches production formats and characteristics. BGP updates follow RFC-compliant message structures including standard AS path representations and network prefix formats. SNMP metrics include realistic baseline values with appropriate variance and correlations between related counters. The 98% baseline traffic combined with 2% anomaly injection creates realistic signal-to-noise ratios that test algorithm robustness.

The evaluation framework implements standard machine learning metrics adapted for the anomaly detection context. Precision measures alert quality from the operator perspective, quantifying what proportion of pages received correspond to actual failures. Recall measures detection completeness from the network perspective, quantifying what proportion of failures trigger appropriate alerts. The F1 score balances these complementary metrics. Detection delay measures operational responsiveness, while Hit@k metrics quantify localization accuracy for root cause investigation.

The scenario-based evaluation approach provides insights into performance across different failure types, following established methodologies for evaluating network monitoring systems (Mohammed et al., 2021). Some scenarios like multi-modal coordinated failures provide strong detection signals across multiple pipelines, while others like isolated BGP events test single-pipeline capabilities. Baseline scenarios validate false positive rates during normal operation. This comprehensive coverage characterizes system behavior across the operational spectrum, enabling assessment of both individual pipeline performance and cross-modal correlation effectiveness.

7 References

References

[1] Cheng, M., Li, Q., Lv, J., Liu, W., & Wang, J. (2021). Multi-Scale LSTM Model for BGP Anomaly Classification. *IEEE Transactions on Services Computing*, 14(3), 765–778. Available at: https://doi.org/10.1109/TSC.2018.2824809

[2] Mohammed, S. A., Mohammed, A. R., Côté, D., & Shirmohammadi, S. (2021). A machine-learning-based action recommender for Network Operation Centers. *IEEE Transactions on Network and Service Management*, 18(3), 2702–2713. Available at: https://doi.org/10.1109/TNSM.2021.3095463

[3] Scott, B., Johnstone, M. N., Szewczyk, P., & Richardson, S. (2024). Matrix Profile data mining for BGP anomaly detection. *Computer Networks*, 242, 110257.

[4] Tan, Y., Huang, W., You, Y., Su, S., & Lu, H. (2024). Recognizing BGP Communities Based on Graph Neural Network. *IEEE Network*, 38(6), 232–238. Available at: https://doi.org/10.1109/MNET.2024.3414113

[5] Allagi, S., & Rachh, R. (2019). Analysis of Network log data using Machine Learning. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 1–3. Available at: https://doi.org/10.1109/I2CT45611.2019.9033528

[6] Skazin, A. (2021). Detection of network anomalies in log files. *IOP Conference Series: Materials Science and Engineering*, 1069(1), 012021. Available at: https://doi.org/10.1088/1757-899X/1069/1/012021

[7] Feltin, T., Cordero Fuertes, J. A., Brockners, F., & Clausen, T. H. (2023). Understanding Semantics in Feature Selection for Fault Diagnosis in Network Telemetry Data. *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 1–9. Available at: https://doi.org/10.1109/NOMS56928.2023.10154455

[8] Wang, H. (2020). Improvement and implementation of Wireless Network Topology System based on SNMP protocol for router equipment. *Computer Communications*, 151, 10–18. Available at: https://doi.org/10.1016/j.comcom.2020.01.001

[9] Manna, A., & Alkasassbeh, M. (2019). Detecting network anomalies using machine learning and SNMP-MIB dataset with IP group. *arXiv preprint arXiv:1906.00863*. Available at: https://arxiv.org/abs/1906.00863

[10] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. *2008 Eighth IEEE International Conference on Data Mining*, 413–422. Available at: https://doi.org/10.1109/ICDM.2008.17

[11] Zhou, J., Qian, Y., Zou, Q., Liu, P., & Xiang, J. (2022). DeepSyslog: Deep Anomaly Detection on Syslog Using Sentence Embedding and Metadata. *IEEE Transactions on Information Forensics and Security*, 17, 3051–3066. Available at: https://doi.org/10.1109/TIFS.2022.3198188

[12] Mueen, A., & Keogh, E. (2017). Matrix Profile Tutorial. *KDD 2017 Tutorial*, University of California, Riverside. Available at: https://www.cs.ucr.edu/ eamonn/MatrixProfile.html

[13] Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.

[14] Järvelin, K., & Kekäläinen, J. (2002). Cumulated Gain-Based Evaluation of IR Techniques. *ACM Transactions on Information Systems*, 20(4), 422–446. Available at: https://doi.org/10.1145/582415.582418

[15] Benzekki, K., El Fergougui, A., & Elbelrhiti Elalaoui, A. (2017). Software-Defined Networking (SDN): A Survey. *Security and Communication Networks*, 2017, 9739131. Available at: https://doi.org/10.1155/2017/9739131

[16] Sommerville, I. (2016). *Software Engineering* (10th ed.). Pearson Education Limited.

[17] Rekhter, Y., Li, T., & Hares, S. (2006). A Border Gateway Protocol 4 (BGP-4). RFC 4271, Internet Engineering Task Force (IETF). Available at: https://www.rfc-editor.org/rfc/rfc4271