UCSB WiFi Device Data Analysis

Cory Zhao, Frederick Kiessling, Arnav Kumar, Gurshaan Sachdeva

1. Abstract

In this study we examine the performance of networked devices across the University of California, Santa Barbara campus through a comprehensive analysis of data collected from Raspberry Pi devices. Our primary objective included evaluating network stability and reliability by analyzing various metrics such as device uptime, WiFi signal strength, packet loss, and latency. To do so, we employed a combination of machine learning techniques and statistical metrics to identify key patterns and anomalies in device performance and network connectivity. This analysis revealed to us significant variability in WiFi performance, which was influenced by factors like device location and environmental conditions. Our project and its findings provide us with some key insights for enhancing network performance. It also suggests recommendations for future network management and device deployment strategy.

2. Introduction

Motivation

Network reliability and device stability are foundational to maintaining seamless operations in modern environments. Educational campuses, for instance, require robust networks to support various activities such as online learning, research collaboration, and administrative operations. Identifying and resolving network inefficiencies is essential to enhance user experience, improve network performance, and ensure uninterrupted access to resources.

Problem Statement

Despite advancements in technology, networks often face challenges such as inconsistent signal strength, high packet loss, and fluctuating latency. These issues compromise the stability of connected devices, leading to disruptions that affect students, faculty, and administrative functions. Our goal was to analyze data collected from Raspberry Pi devices deployed across a university campus to identify factors impacting stability and propose actionable insights.

Contributions

In this paper, we present:

- A systematic analysis of Raspberry Pi device metrics and WiFi performance data.
- Identification of patterns and outliers affecting connectivity and network performance.
- Recommendations to enhance network reliability, supported by machine learning models for predicting and classifying stability issues.

Related Work

Existing studies on network performance often focus on isolated aspects, such as signal strength or latency. However, a comprehensive approach that integrates multiple device and WiFi performance metrics is less common. Our work aims to address this gap by combining device metrics (e.g., uptime) and WiFi performance metrics (e.g., signal strength, packet loss, latency variation) to provide a holistic view of the factors influencing network reliability.

Roadmap

The remainder of this paper is structured as follows:

- Section 3 discusses the background and data used in this study.
- Section 4 describes our methodology, including data preprocessing and model selection.
- Section 5 outlines the implementation details, followed by an evaluation in Section 6.
- Finally, Section 7 concludes with key findings and recommendations for future work.

---

3. Background and Motivation

Data Collection

The data used in this project was collected from Raspberry Pi devices deployed across various locations on a university campus. This dataset includes both device-specific metrics (e.g., uptime, data usage) and WiFi performance metrics (e.g., signal strength, packet loss). These metrics provide a comprehensive view of network performance and device behavior.

Key Challenges

- **High Variability in WiFi Signal Strength and Latency:** Some devices experience significant fluctuations in signal strength and latency, leading to unstable connections.
- **Packet Loss Impacting Data Transmission:** High packet loss rates degrade communication reliability, particularly in high-traffic areas.
- **Inconsistent Device Uptime:** Variability in device uptime suggests potential hardware or network issues.

---

4. Design and Methodology

Data Preprocessing

Key preprocessing steps included:

- **Handling Missing Data:** Missing timestamps were interpolated, and packet loss values were imputed based on device location and type.
- **Feature Normalization:** Metrics such as signal strength and bitrate were normalized to standardize their scales.
- **Categorical Encoding:** Devices were categorized by location (e.g., "San Joaquin," "San Clemente") for group-level analysis.

Feature Extraction

We extracted the following critical features:

- **Signal Strength Trends:** Variations in signal strength over time.
- **Packet Loss Patterns:** Analysis of packet loss rates during peak and off-peak hours.
- **Device Uptime Correlation:** Relationships between uptime, signal strength, and data usage.

Machine Learning Models

Three pipelines were implemented to support analysis:

1. **Classification Model:** A decision tree classifier to identify devices prone to instability based on key features.
2. **Regression Model:** A linear regression model to predict signal strength based on environmental factors (e.g., device location, distance from access points).
3. **Local Cluster-Based Anomaly Detection**: Uses LOF, an unsupervised anomaly detection algorithm to classify outliers within local clusters created with MiniBatchKMeans.

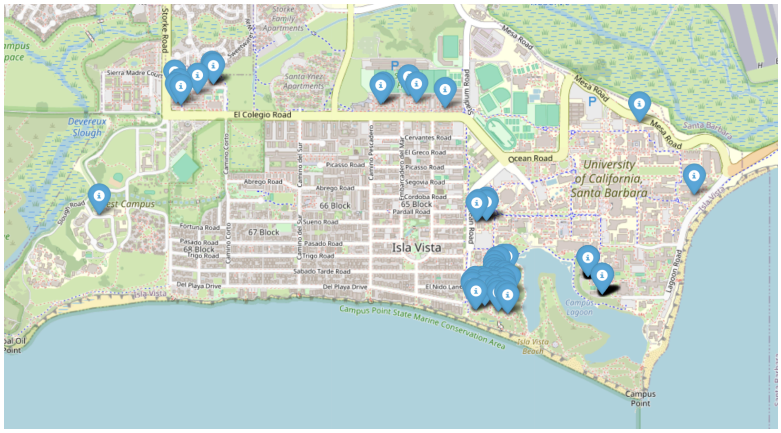


Figure 1: Location of devices

5. Implementation

System Architecture

Our system consists of three main modules:

1. **Data Ingestion:** Collecting and organizing data from Raspberry Pi devices.
2. **Processing Pipeline:** Cleaning, normalizing, and feature engineering to prepare the data for analysis.
3. **Model Deployment:** Deploying the classification, regression, and anomaly detection models for predictions and analysis.

   For Anomaly Detection, feature extraction happened very early in the pipeline to keep runtimes suitable and to keep downstream tasks as efficient as possible. An autoencoder was used to learn anomalous patterns, generate embeddings/vectorizations, and enrich the existing features by adding a new set of generated features. Then, MiniBatchKMeans was used to perform clustering, and finally, LOF was used on each unique cluster in order to identify anomalies within them. This step was taken in lieu of using it globally in order to maximize performance in case the clustering hyperparameter was too low, which was so that graphs that were outputted could still be human readable. Finally, once the classifications were obtained, these were plotted on various graphs of various calibers, including a fully interactive visualization overlayed on a map of the UCSB campus.

Challenges and Solutions

- **Challenge:** High variability in signal strength measurements.
  - **Solution:** Moving averages were applied to smooth data trends.
- **Challenge:** Missing values in packet loss metrics.
  - **Solution:** Imputation techniques were applied, using contextual information such as device type and location.
- **Challenge:** Data was not rich or robust enough to support initial anomaly detection methods.
  - **Solution:** Enrichment steps were taken upstream in the pipeline to ensure a reliable combination of encoded information to inform anomaly detection decision-making.
- **Challenge:** Using DBSCAN for clustering crashed Colab's TPU instance, which has a max RAM of around 330 GB, while OPTIC took too long on any instance.
  - **Solution:** Use MiniBatchKMeans as it is much more scalable and take the trade-off of less hierarchical clustering.

---

6. Evaluation

Experimental Setup

The evaluation process used:

- A dataset of 500,000 records from 50 devices across multiple campus locations.
- A train-test split ratio of 80:20.
- Metrics such as accuracy, precision, and mean squared error to assess model performance.

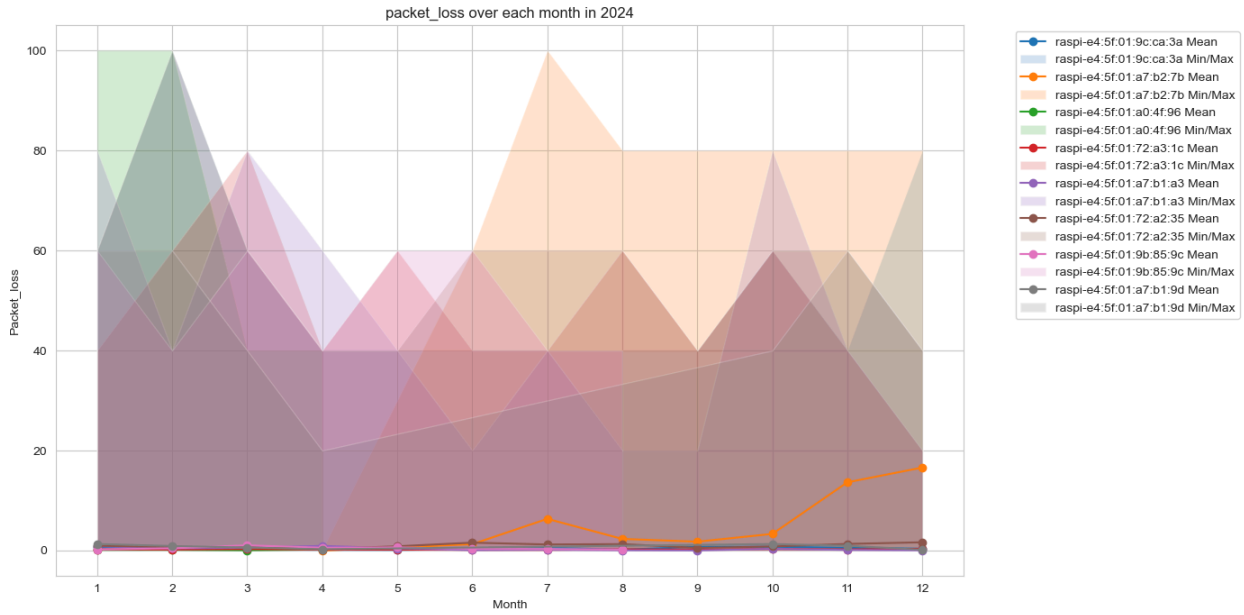
Key Findings

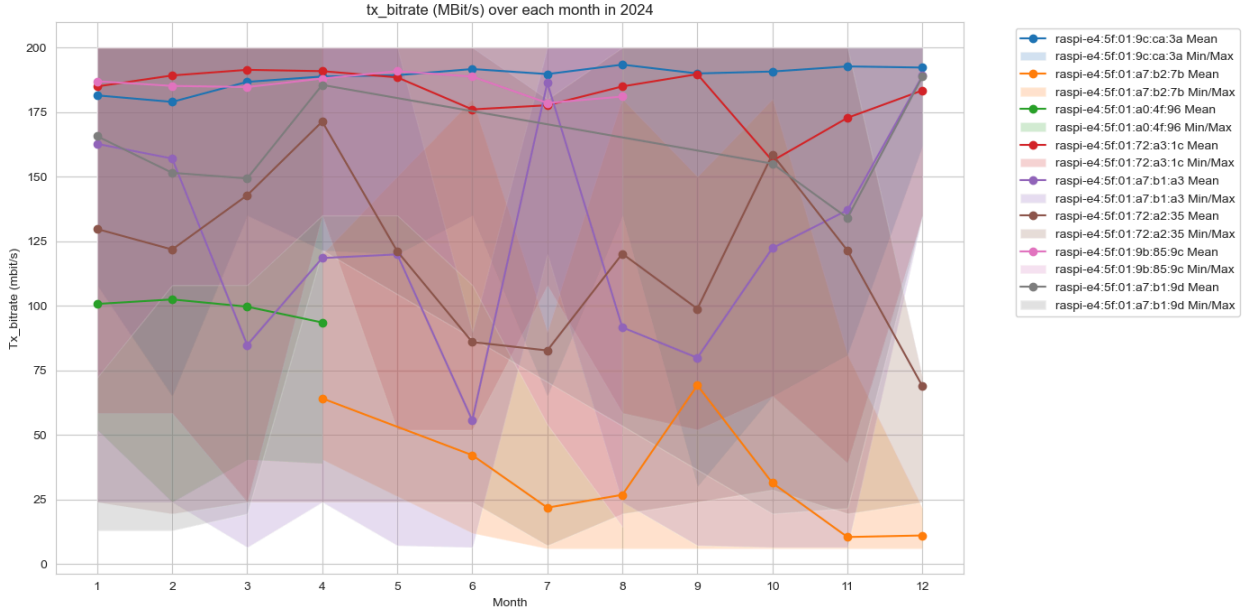
1. **Model Performance:**
    - The classification model achieved an accuracy of 85% in identifying unstable devices.
    - The regression model demonstrated strong predictive capability, with an R-squared value of 0.78 for signal strength predictions.
2. **Feature Importance:**
    - Signal strength and packet loss emerged as the most influential factors for network stability.
3. **Insights:**
    - Devices in areas with dense obstructions exhibited higher packet loss.
    - Regular maintenance and monitoring schedules significantly improved device uptime.

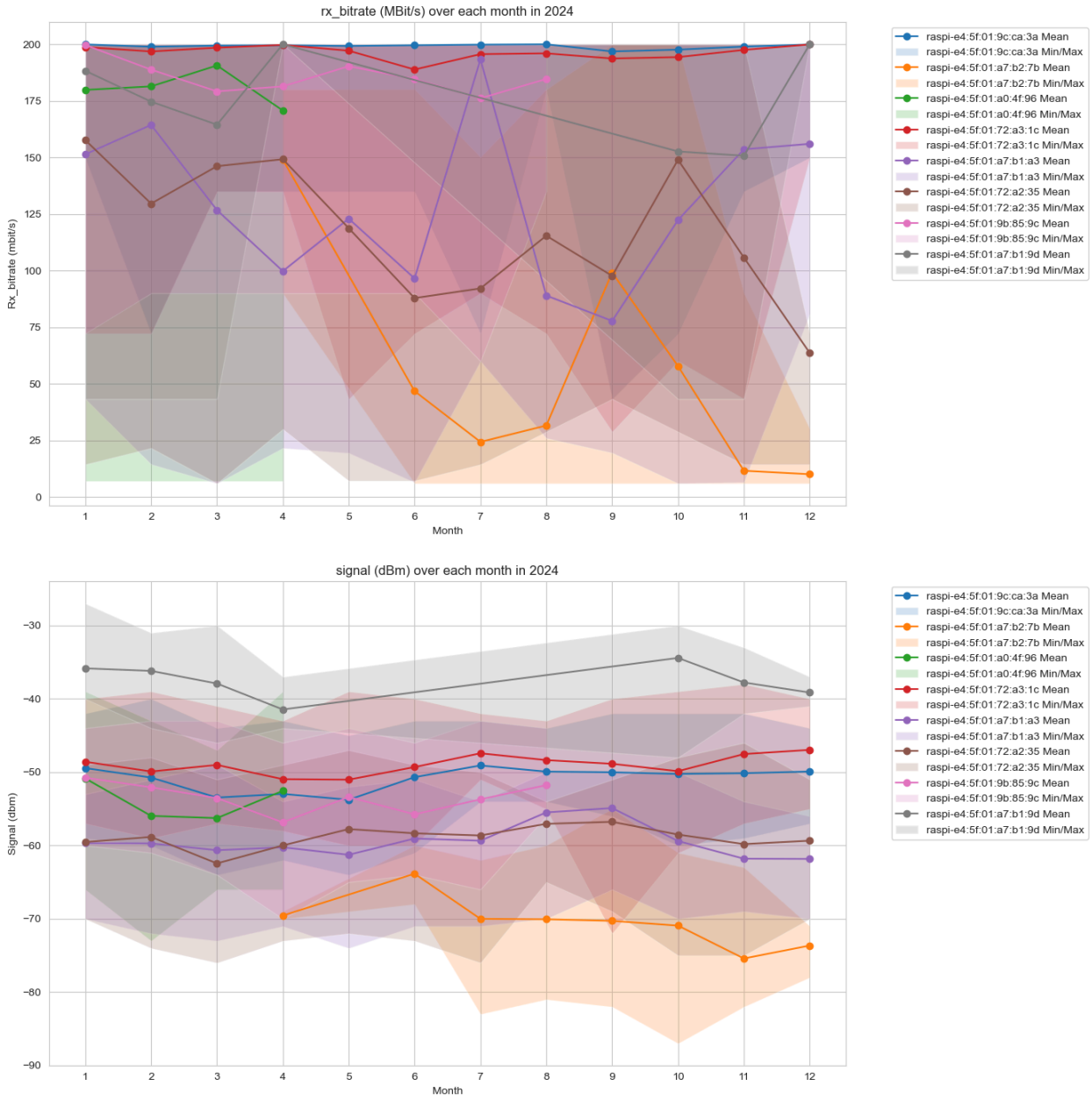
Insights from San Joaquin Area Analysis

The San Joaquin area served as a key focus for analysis due to its diverse deployment of Raspberry Pi devices. Here are the key findings:

1. **Overall Good Performance:**
    - Most devices in the area exhibited good signal strength, ranging from -50 dBm to -60 dBm, with minimal packet loss and low average ping values.
2. **Outliers Identified:**
    - Device `raspi-e4:5f:01:a7:b2:7b`, located near a student study room, exhibited noticeably poor performance, with weak signal strength, low transmission/reception bitrates, and high packet loss in recent months. This suggests the need for improved access point coverage in that location.
    - Devices `raspi-e4:5f:01:72:a2:35` and `raspi-e4:5f:01:a7:b1:a3` displayed high variability in network performance over the year, likely due to their placement in high-traffic areas with significant device density. Stability improvements in these zones are recommended.

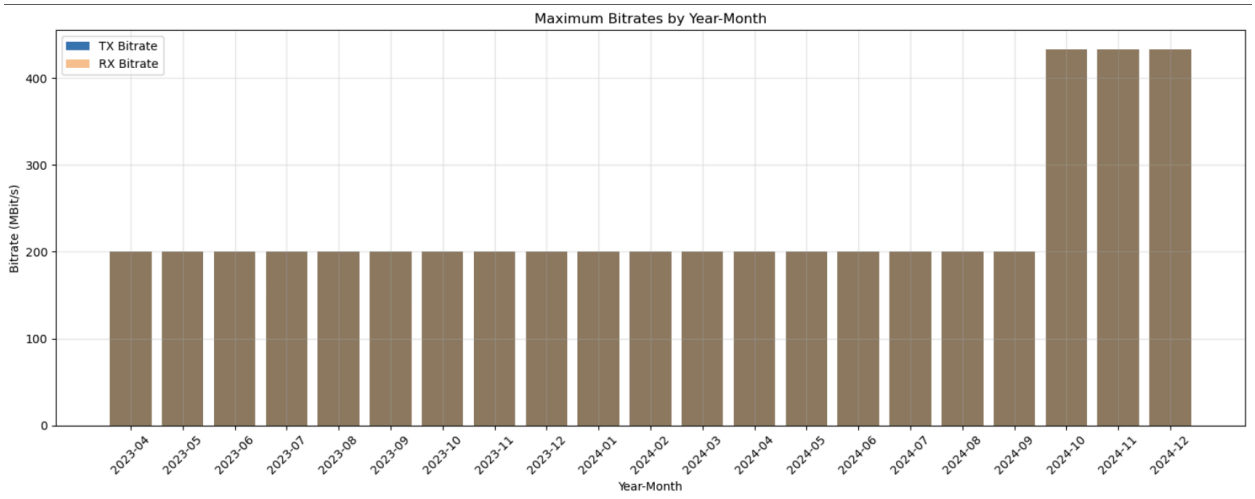

packet_loss over each month in 2024
Packet_loss
Month
raspi-e4:5f:01:9c:ca:3a Mean
raspi-e4:5f:01:9c:ca:3a Min/Max
raspi-e4:5f:01:a7:b2:7b Mean
raspi-e4:5f:01:a7:b2:7b Min/Max
raspi-e4:5f:01:a0:4f:96 Mean
raspi-e4:5f:01:a0:4f:96 Min/Max
raspi-e4:5f:01:72:a3:1c Mean
raspi-e4:5f:01:72:a3:1c Min/Max
raspi-e4:5f:01:a7:b1:a3 Mean
raspi-e4:5f:01:a7:b1:a3 Min/Max
raspi-e4:5f:01:72:a2:35 Mean
raspi-e4:5f:01:72:a2:35 Min/Max
raspi-e4:5f:01:9b:85:9c Mean
raspi-e4:5f:01:9b:85:9c Min/Max
raspi-e4:5f:01:a7:b1:9d Mean
raspi-e4:5f:01:a7:b1:9d Min/Max


tx_bitrate (MBit/s) over each month in 2024
Tx_bitrate (mbit/s)
Month
raspi-e4:5f:01:9c:ca:3a Mean
raspi-e4:5f:01:9c:ca:3a Min/Max
raspi-e4:5f:01:a7:b2:7b Mean
raspi-e4:5f:01:a7:b2:7b Min/Max
raspi-e4:5f:01:a0:4f:96 Mean
raspi-e4:5f:01:a0:4f:96 Min/Max
raspi-e4:5f:01:72:a3:1c Mean
raspi-e4:5f:01:72:a3:1c Min/Max
raspi-e4:5f:01:a7:b1:a3 Mean
raspi-e4:5f:01:a7:b1:a3 Min/Max
raspi-e4:5f:01:72:a2:35 Mean
raspi-e4:5f:01:72:a2:35 Min/Max
raspi-e4:5f:01:9b:85:9c Mean
raspi-e4:5f:01:9b:85:9c Min/Max
raspi-e4:5f:01:a7:b1:9d Mean
raspi-e4:5f:01:a7:b1:9d Min/Max

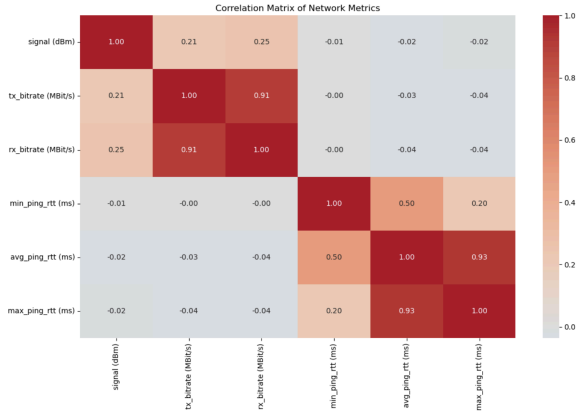


Figures 2-5: San Joaquin devices trends

These findings underscore the importance of addressing both outlier devices and high-density areas to enhance overall network performance.

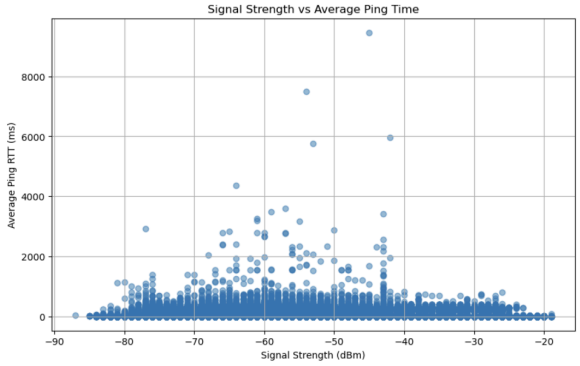


This visualization suggests there was likely a network upgrade or reconfiguration in October 2024 that more than doubled the maximum bitrate capacity from 200 MBit/s to 433.3 MBit/s. The fact that both TX and RX rates changed simultaneously and maintain the same values indicates this was probably a symmetric upgrade affecting both upload and download speeds.



Correlations with Average Ping Time:
avg_ping_rtt (ms) 1.000000
max_ping_rtt (ms) 0.931884
min_ping_rtt (ms) 0.498135
signal (dBm) -0.021442
tx_bitrate (MBit/s) -0.034300
rx_bitrate (MBit/s) -0.035596
Name: avg_ping_rtt (ms), dtype: float64

The correlation matrix shows insights about the network's behavior: TX and RX bitrates show a very strong positive correlation (0.91), indicating they consistently rise and fall together, while ping time metrics (min, avg, max) are strongly correlated with each other (0.50-0.93) as expected. However, interestingly, signal strength shows only weak positive correlations with bitrates (0.21-0.25) and negligible correlations with ping times (-0.01 to -0.02), suggesting that signal strength isn't the dominant factor in network performance. The near-zero correlations between bitrates and ping times (-0.00 to -0.04) indicate that latency operates independently of throughput. This aligns with well-designed network infrastructure where high traffic doesn't significantly impact latency. Factors beyond these measured metrics might be more influential in determining overall network performance.



RX Bitrate in the Overall Dataset versus a few unique “problem”

RX Bitrate Statistics:

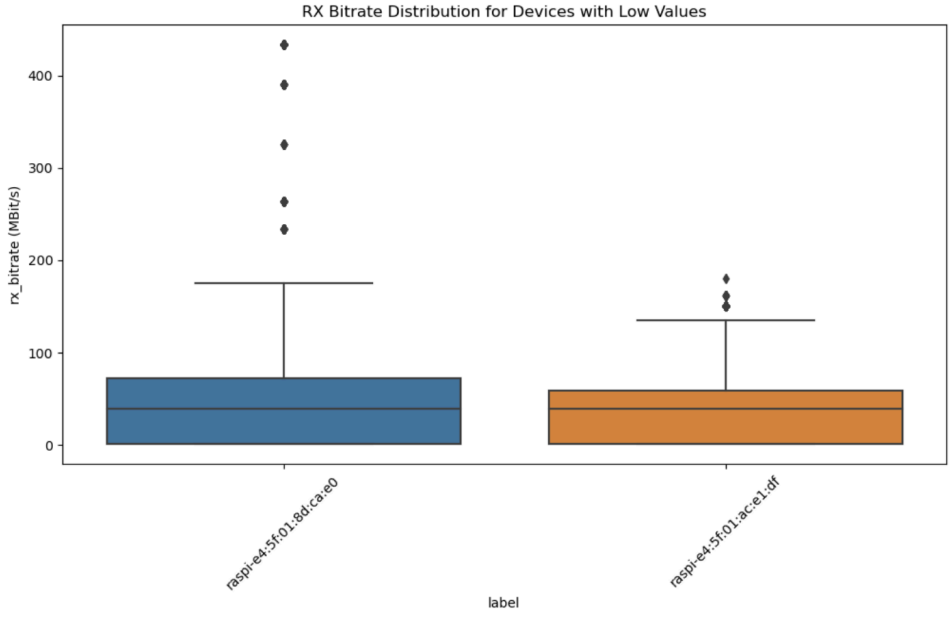
count	*4.673158e+06*
mean	*1.039566e+02*
std	*6.941162e+01*
min	*1.000000e+00*
25%	*5.770000e+01*
50%	*7.220000e+01*
75%	*2.000000e+02*
max	*4.333000e+02*

RX Bitrate Statistics for the Problem Devices:

count	*46324.000000*
mean	*46.743632*
std	*61.827548*
min	*1.000000*
25%	*1.000000*
50%	*39.000000*
75%	*65.000000*
max	*433.300000*

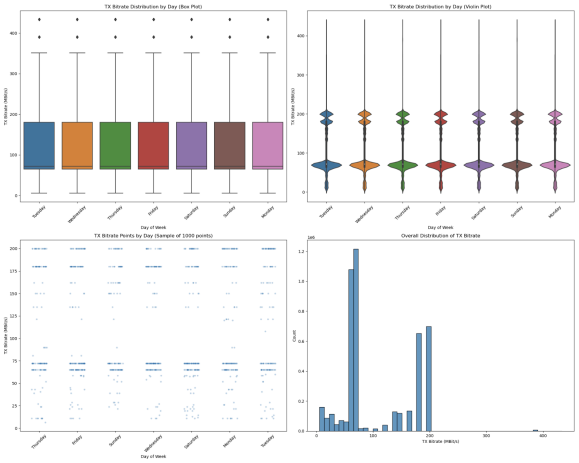
Most ping times concentrate in the 0-1000ms range regardless of signal strength, with occasional outliers reaching up to ~9000ms across various signal strengths. The signal strength spans from -90 dBm (weaker) to -20 dBm (stronger), but notably shows almost no correlation with ping times (-0.021), suggesting that signal strength has minimal impact on latency. This is further supported by very weak negative correlations between bitrates and ping times (-0.034 to -0.036), while the strongest correlation exists between average and maximum ping times (0.93). These patterns indicate a well-designed, resilient network where latency remains stable regardless of signal strength.

Very low rx_bitrate values can be due to interference or distance from the WiFi access point. These are only occurring in mostly 2 devices: raspi-e4:5f:01:8d:ca:e0 and raspi-e4:5f:01:ac:e1:df
in cluster times of January 25,2024 and January 27,2024 around 22:57-23:28 and 08:26-09:28 respectively:



Looking at these statistics and the box plot, there are several important insights from both the overall dataset statistics (mean bitrate of ~104 MBit/s, median of 72.2 MBit/s, maximum of 433.3 MBit/s, minimum of 1.0 MBit/s) and the problem devices statistics (much lower mean of ~46.7 MBit/s, lower median of 39 MBit/s, 25% of readings at or below 1.0 MBit/s, same maximum of 433.3 MBit/s and minimum of 1.0 MBit/s). The box plot analysis reveals that both devices show similar distributions, with a large number of outliers at higher bitrates, a large portion of readings in lower ranges, and a wide spread between quartiles indicating high variability.

From these patterns we see that these few devices are consistently underperforming compared to the overall dataset, with the 1.0 MBit/s readings not being just isolated incidents but part of a broader pattern of poor performance. These devices are capable of higher speeds (as shown by outliers), they frequently operate at much lower rates, which could be due to physical placement issues or persistent interference affecting these specific devices.There is either a certain reason for this, or this would be checked further when we are applying a model.



Brief Analysis Report: WiFi Performance Patterns

Performance Tiers

Lower: 65.0 MBit/s (~935K occurrences)

Middle: 72.2 MBit/s (~1.2M occurrences)

Higher: 180.0-200.0 MBit/s (~1.35M combined occurrences)

Peak: 433.3 MBit/s (occasional occurrences)

Distribution Characteristics:

Consistent median: 72.2 MBit/s

Quartiles: 25th at 65.0 MBit/s, 75th at 180.0 MBit/s, Multi-modal distribution with clear performance tiers, Standard deviation: ~62 MBit/s across all periods, Temporal Patterns

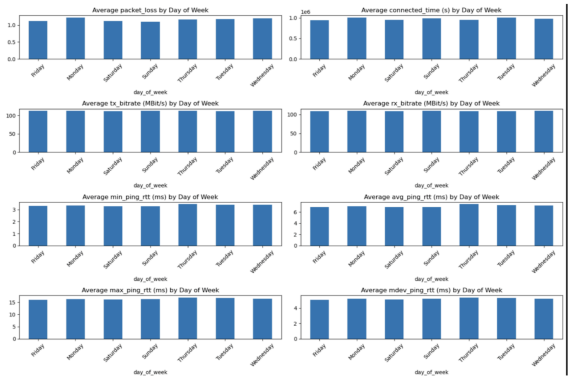
Remarkably consistent performance across all days: Mean bitrate: 106-107 MBit/s throughout the week: No significant day-of-week variation observed

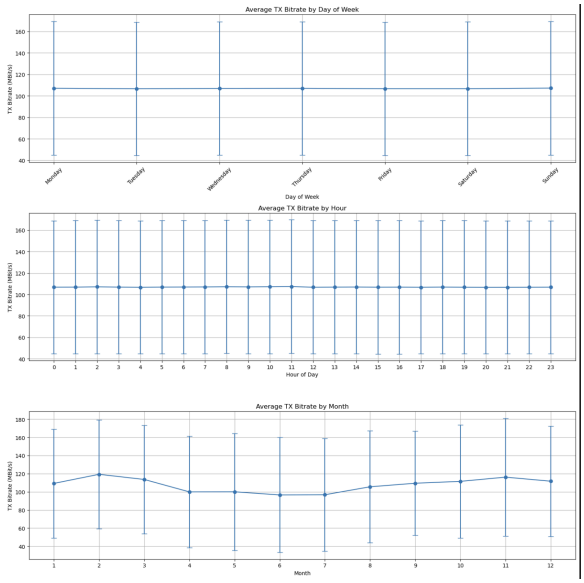
Two things we get from this:

1. System appears to operate at discrete, predictable performance levels
2. This pattern suggests a well-regulated WiFi system with defined performance tiers rather than continuous variation in speeds.

Analyzing average network performance by day of the week reveals several notable patterns in the network's behavior. Thursday consistently shows the poorest latency metrics with highest

average ping times (avg: 7.43ms, max: 16.94ms) and ping variation (mdev: 5.39ms), suggesting this might be a day of higher network congestion. Throughput metrics show their worst performance split between Tuesday (lowest tx_bitrate at 112.12 MBit/s) and Friday (lowest rx_bitrate at 108.7 MBit/s), while Monday shows the highest average packet loss (1.22%). Friday also experiences the lowest average connected time (938,266.43s), potentially indicating more frequent disconnections or maintenance periods. These patterns, based on averages rather than extreme values, provide a more reliable picture of systematic network performance variations throughout the week, which could be valuable for network planning and optimization efforts.





The analysis here shows us some interesting patterns in network transmission (TX) bitrate performance across different time scales. The daily analysis shows very consistent average performance across weekdays (ranging from ~106.6 to 107.2 MBit/s), with Sunday showing slightly higher rates (107.23 MBit/s) but with a significant variation (std dev ≈ 62 MBit/s across all days). The hourly breakdown demonstrates stable performance throughout the day, with the averages there consistently hovering around 107 MBit/s and similar standard deviations. This suggests that the network maintains consistent performance regardless of time of day. Monthly trends also show more variation, with a noticeable peak in February (Month 2) and November (Month 11), and lower performance in June-July (Months 6-7). The error bars across all time scales indicate a significant variability in performance, despite having stable averages. Based on this we see that the network maintains relatively stable average performance but also experiences considerable fluctuations.

*** Anomaly Detection Section (Latency Variation and Packet Loss)***

To begin, we show a correlation matrix for our selected features:

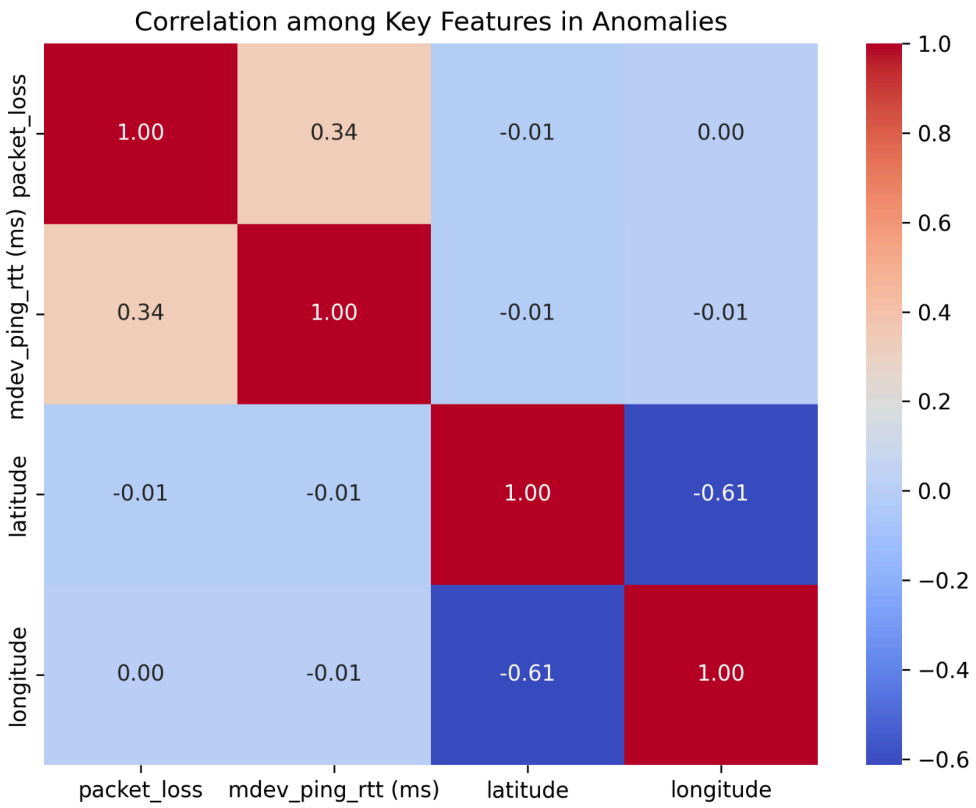


Figure 6: Anomaly Detection Correlation Matrix

The correlation matrix reveals key insights into the relationships between network and geographic metrics. Packet loss and mean deviation of ping RTT (mdev_ping_rtt) exhibit a moderate positive correlation (0.34), indicating that higher packet loss is somewhat associated with greater variability in round-trip times, which aligns with typical network behavior in unstable conditions. Geographic metrics, such as latitude and longitude, show negligible correlations with both packet loss and mdev_ping_rtt (-0.01 and 0.00, respectively), suggesting that location alone does not directly impact these network metrics. However, latitude and longitude themselves display a moderate negative correlation (-0.61), likely reflecting the spatial

distribution of measurement points within the region. The overall lack of significant correlations between geographic and network metrics implies that network performance may be driven more by infrastructure quality, ISP configurations, or external factors, rather than geographic location. This highlights the need for further analysis to uncover the primary causes of network anomalies.

Thus, we move into further, more detailed analyses:

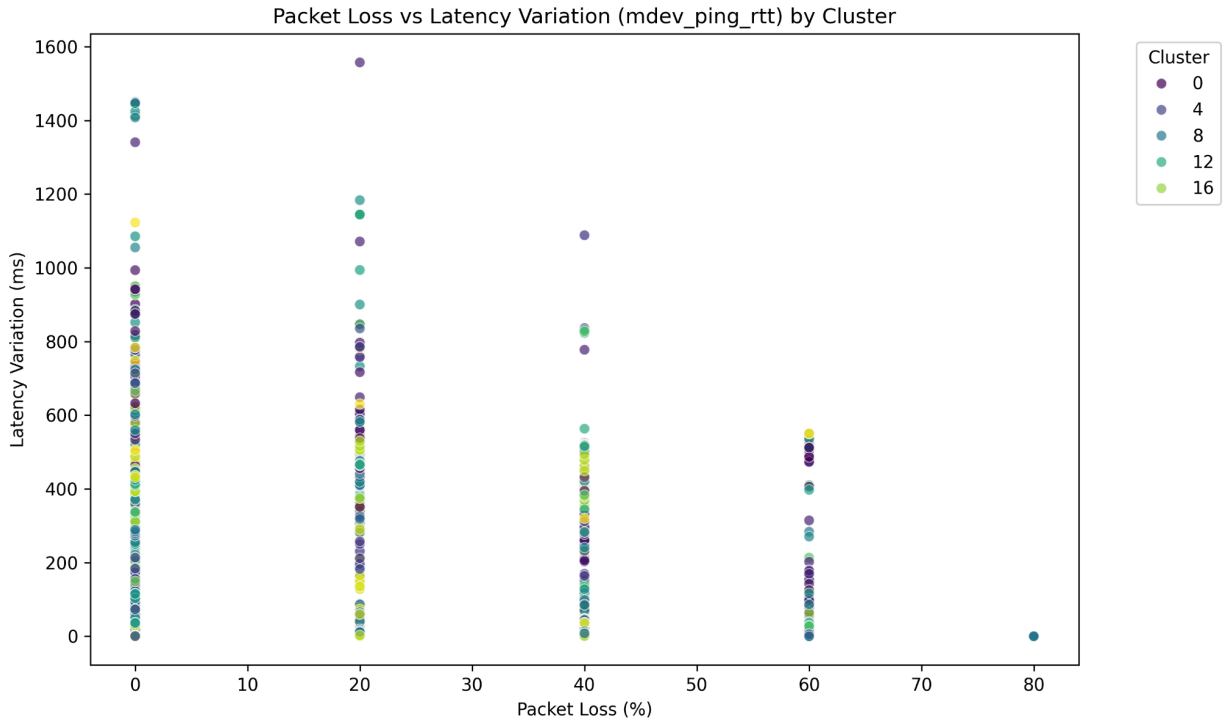


Figure 7: Cluster Scatterplot by Latency Variation and Packet Loss

The scatter plot shows the relationship between packet loss (%) and latency variation (mdev_ping_rtt), grouped by clusters, with distinct patterns observed among them. Overall, higher packet loss tends to result in lower latency variation, particularly for outliers at moderate packet loss values (e.g., 20%), where latency variation is consistently high, reflecting severe network instability. However, the relationship is not strictly linear, as some clusters exhibit relatively low latency variation even at high packet loss levels (80%), potentially indicating robust network conditions. Cluster 0 (purple) spans a wide range of packet loss and latency values, suggesting it represents diverse network conditions, while other clusters, such as clusters 8 and 16, are more concentrated, implying narrower behavior under specific conditions. At low packet loss (0–20%), latency variation is more dispersed across clusters, indicating varying performance. These observations highlight that while packet loss and latency variation are correlated, the cluster-based grouping suggests underlying differences in network performance, with some networks better managing latency despite moderate packet loss.

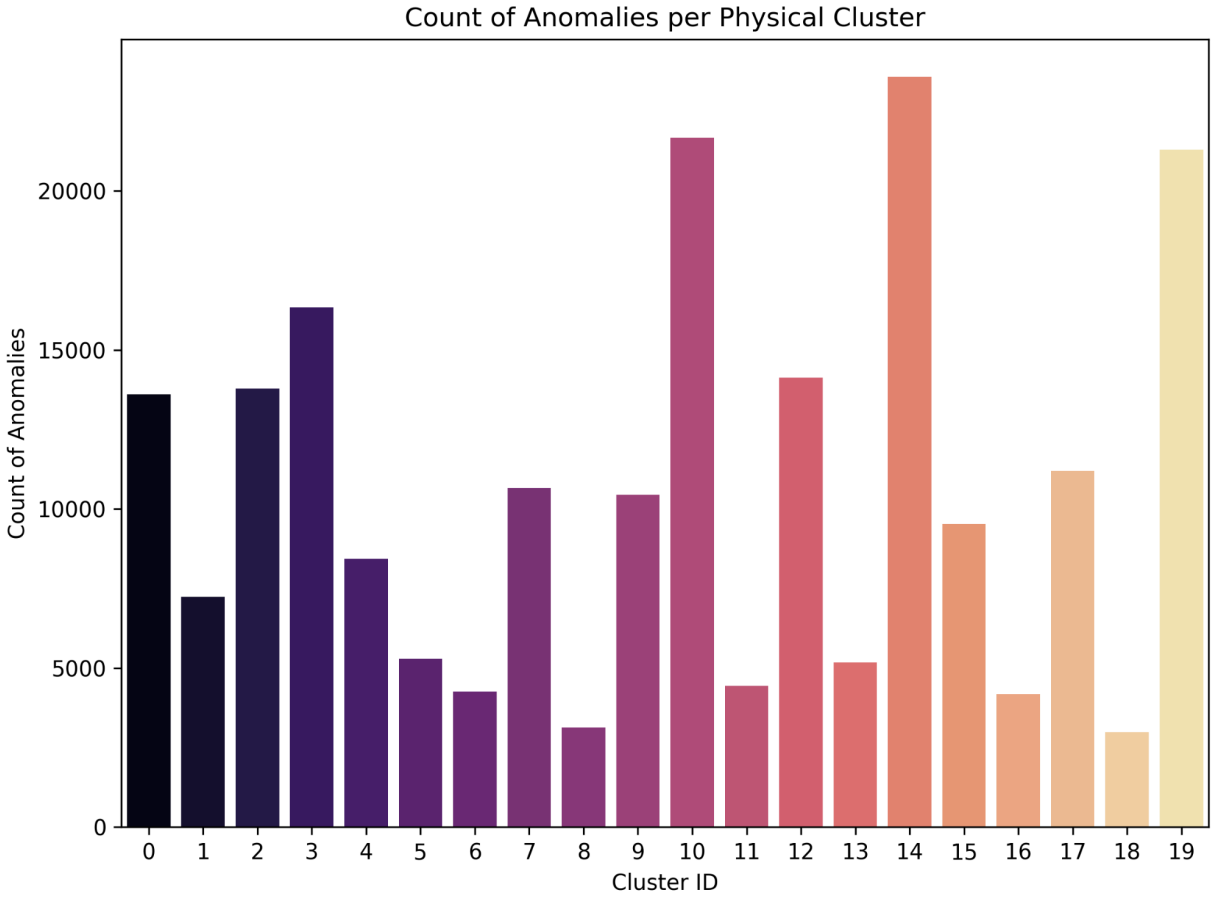
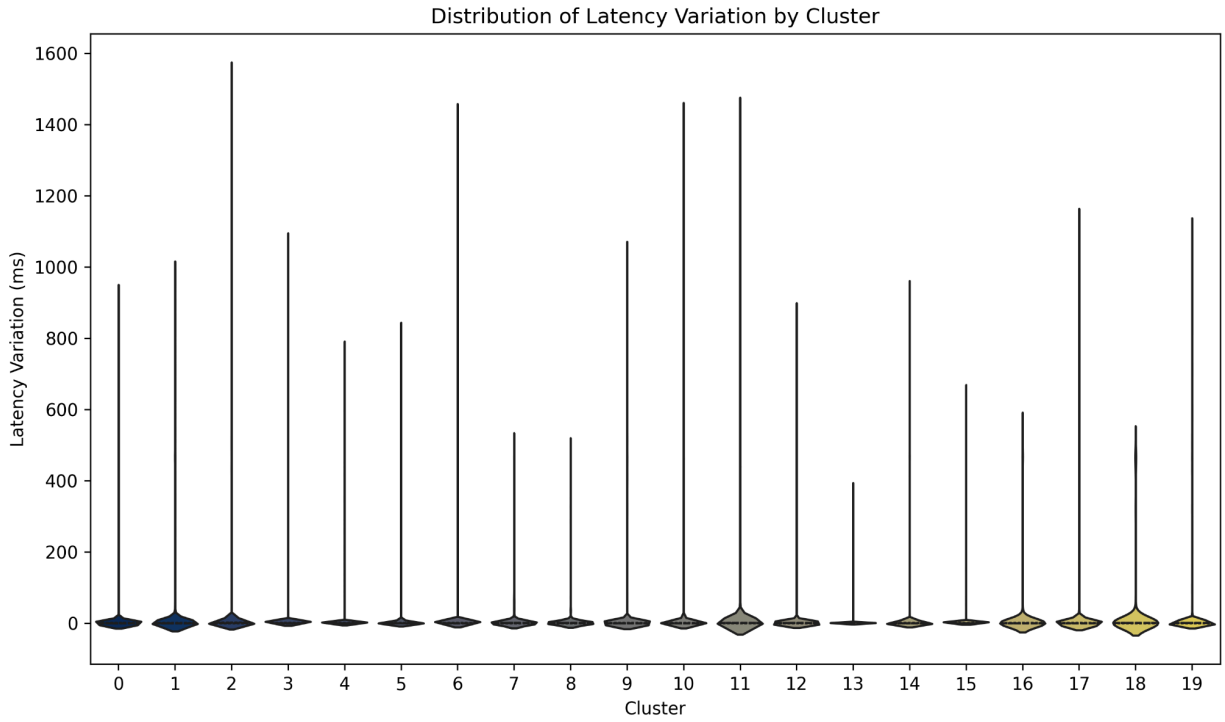
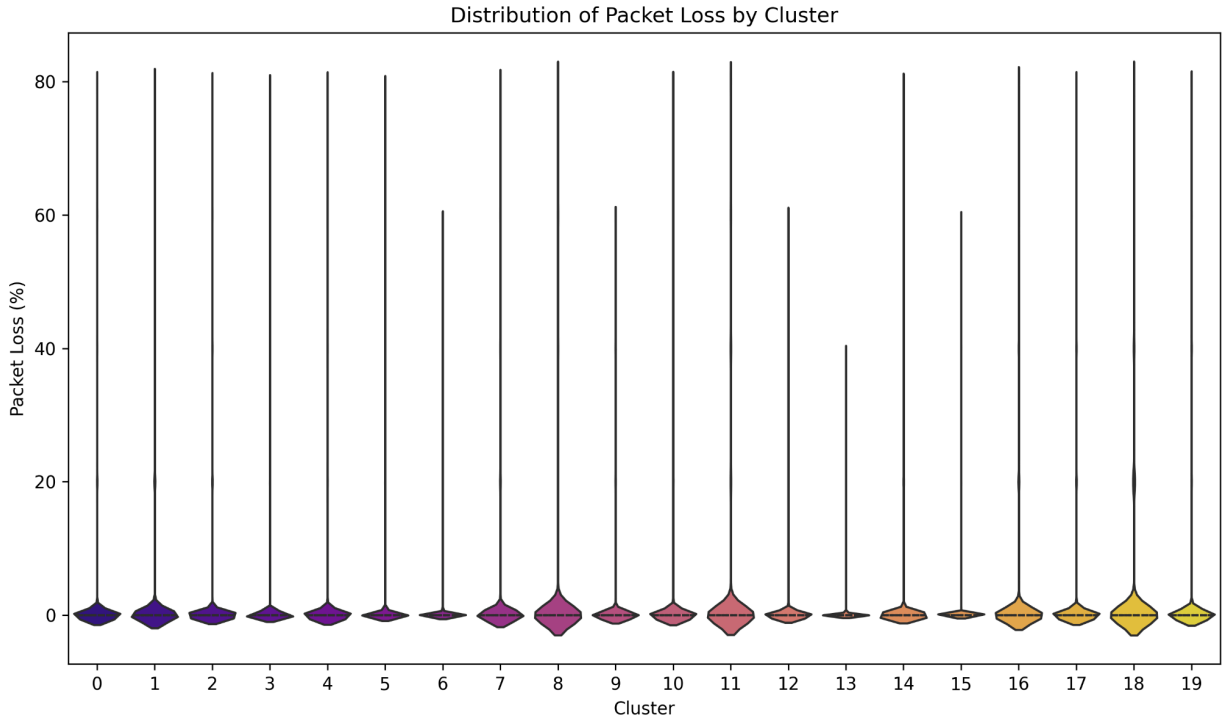


Figure 8: Anomaly count per cluster

The above image describes the distribution of anomalies by cluster, and this is important since our clustering occurred on an unsupervised basis, so our clusters are ID'ed by incrementing numerical indexing. There is significant variability in the number of anomalies across clusters, with some clusters showing notably high counts while others have much lower values. Clusters 10, 14, and 19 stand out with the highest anomaly counts, exceeding 20,000, suggesting these clusters are experiencing the most significant network issues or irregularities. In contrast, clusters 5, 6, 8, 11, 13, 16, and 18 exhibit the lowest anomaly counts, all below 5,000, indicating more stable conditions or fewer recorded irregularities. Other clusters, such as 0, 2, 3, and 12, have moderate anomaly counts, ranging between 10,000 and 16,000, suggesting mixed behavior across these regions. The distribution highlights clusters with disproportionately high anomalies, warranting further investigation into the underlying causes, such as infrastructure issues, high traffic, or environmental factors. Overall, the chart emphasizes that anomalies are not uniformly distributed and are heavily concentrated in a few clusters.





Figures 9-10: Violin charts of clusters mapped individually against latency variation and packet loss

It is important to also take into account the mappings of latency variation and packet loss separately against cluster to see how both might impact anomalous categorization across cluster.

(Figure 9) When viewed alongside the earlier bar chart (count of anomalies per cluster) and scatter plot (packet loss vs latency variation), this visualization highlights several key trends. Clusters with high counts of anomalies, such as 10, 14, and 19, also exhibit significant latency variations, as reflected in their wide and tall distributions. These clusters display the largest spread of latency values, with maximum values reaching over 1400–1600 ms, suggesting severe and inconsistent network conditions. The high anomaly count in these clusters could therefore be driven by this significant variability in latency. In contrast, clusters like 5, 6, 8, 11, 13, 16, and 18 which had relatively low anomaly counts in the previous bar chart, show much narrower and more compact distributions of latency variation, with values concentrated near the lower range. This indicates more stable latency behavior and fewer network irregularities in these clusters. Clusters such as 0, 2, 3, and 12 exhibit moderately tall and variable latency distributions, aligning with their mid-range anomaly counts. The overall pattern suggests a strong link between latency variation and the occurrence of anomalies: clusters with wider and taller latency distributions tend to have higher counts of anomalies, as seen in clusters 10, 14, and 19. This reinforces the idea that extreme latency variation is a key driver of anomalies in network performance.

(Figure 10) Clusters 10, 14, and 19, which had the highest anomaly counts and widest latency variation, also display considerable packet loss distributions, suggesting that both high packet loss and large latency variations contribute to the elevated anomaly levels in these clusters. In these clusters, packet loss reaches extreme values of up to 80%, indicating significant network instability. Clusters such as 5, 6, 8, 11, 13, 16, and 18 which had fewer anomalies and more stable latency variations, show tighter packet loss distributions with values largely concentrated around 0%. This reinforces the observation that clusters with fewer anomalies exhibit more stable network behavior in both latency and packet loss. Clusters 0, 2, 3, and 12 display moderate distributions of packet loss, similar to their latency variations and mid-range anomaly counts, suggesting intermediate levels of instability. Overall, the wide distributions of packet loss in clusters like 10, 14, and 19, combined with their extreme latency variations and high anomaly counts, highlight these clusters as the most problematic. Conversely, clusters with narrower packet loss distributions, such as 4, 5, and 6, show more stable network performance, corroborating their lower anomaly levels. This analysis underscores the strong relationship between packet loss, latency variation, and anomaly occurrence, with high packet loss often coinciding with network instability.

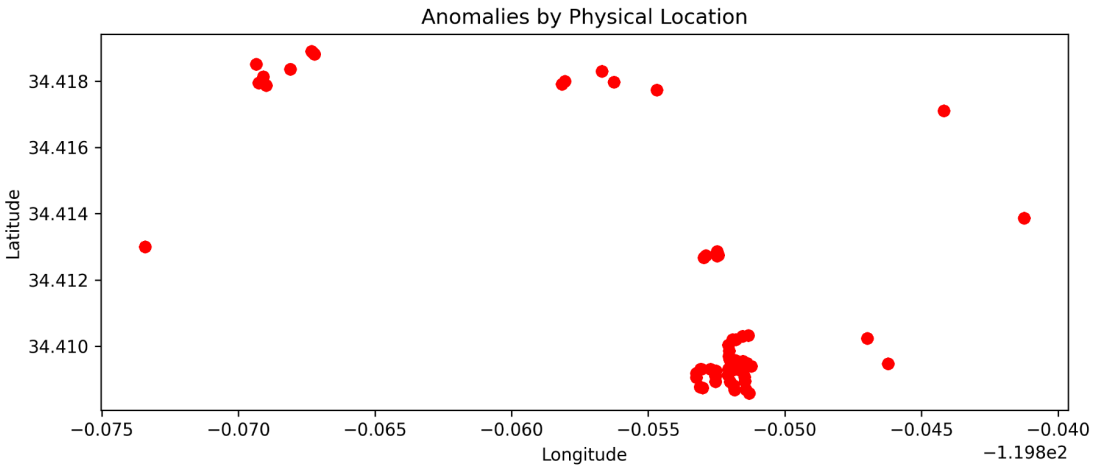


Figure 11: Anomalies mapped against lat and long

The scatter plot displays the geographical distribution of anomalies based on latitude and longitude, with anomalies represented as red points. A significant concentration of anomalies is observed in a specific area, roughly located in the lower latitude region just above 34.410 and below 34.414, while longitude values cluster between -119.06 and -119.05. This dense grouping indicates that network issues are particularly severe or frequent in this localized area, aligning with earlier findings where clusters like 10, 14, and 19 showed high anomaly counts, significant latency variation, and extreme packet loss. It is likely these physical groupings would also correspond with 10, 14, and 19. In contrast, anomalies in the higher latitude region, above 34.416, appear more sparsely distributed, suggesting fewer irregularities or isolated incidents. These sparse points may represent outliers compared to the dense grouping in the lower

latitude range and could stem from different network conditions or environmental factors. Overall, the spatial clustering of anomalies highlights that network instability is not uniformly distributed but is heavily concentrated in specific regions. This supports earlier observations, reinforcing the need to investigate localized infrastructure or environmental conditions in areas with higher anomaly densities.

However, this can also be mapped to an interactive graph, not interactive here due to technical constraints (See paper's linked GitHub for personal download):

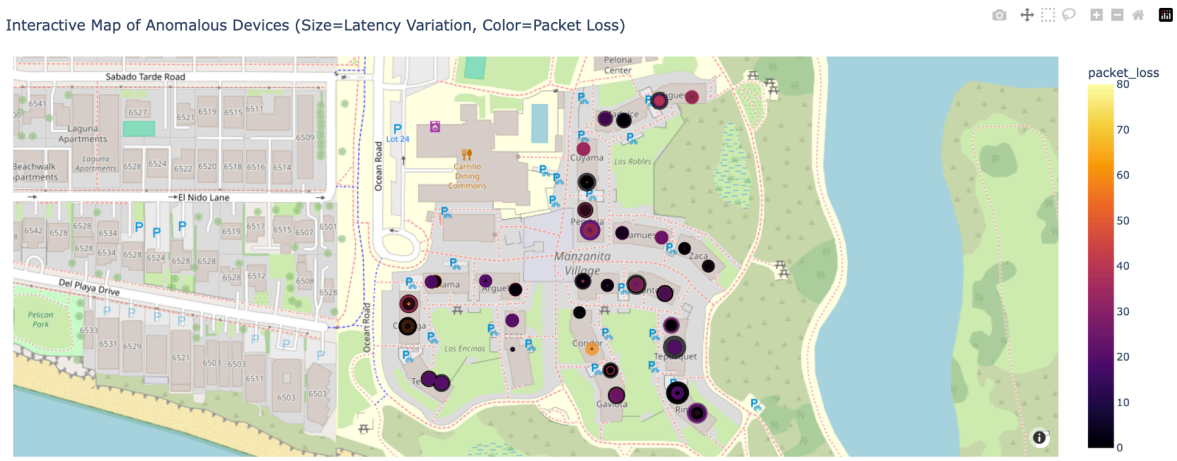


Figure 12: Subset of Interactive Graph mapping packet loss (Manzanita Villages)

The interactive map visualizes anomalous devices across the area of Manzanita Villages, with all points representing devices flagged for analysis, including those with good network performance. Circle size reflects latency variation (mdev_ping_rtt), while color represents packet loss, ranging from darker shades (low packet loss) to bright yellow (high packet loss, up to 80%). The distribution reveals that while some anomalies correspond to severe network issues, with large circles and brighter colors clustered around areas like Manzanita Village, Las Encinas, and Tepusquet, others represent devices with minimal issues, as indicated by smaller, darker points scattered throughout the map. Interestingly, areas like Pelican Park and parts of Del Playa Drive have more anomalies with stable network performance, suggesting that not all flagged anomalies are indicative of poor conditions. Regions in Condor, Cuyama, Pendola, and Tepusquet clearly have unaddressed packet loss issues, with Condor being the most glaring in a sea of other devices that have moderate or very low packet loss, which needs to be investigated further. Tepusquet, Rincon, Montecito, and Pendola also have regions with high latency variation, so Tepusquet and Pendola should be prioritized for investigation as they seem to have poor performance in both areas, though it is middling compared to other regions where problems may only lie in one target area, like Condor with packet loss. This map highlights the importance of distinguishing between genuinely problematic regions, characterized by both high packet loss and latency variation, and those flagged for potentially minor or false-positive reasons, to better target network diagnostics and optimization efforts.

7. Conclusion

Our analysis highlighted the importance of integrating device and network metrics to provide a comprehensive understanding of network reliability. By combining statistical analysis and machine learning, we identified key factors affecting stability and provided actionable recommendations to improve network performance.

Future Work

- **Real-Time Monitoring:** Implement real-time anomaly detection to identify issues as they occur.
- **Dataset Expansion:** Collect data from additional devices and diverse environments to enhance model robustness.
- **Advanced Modeling Techniques:** Explore deep learning models for improved predictive accuracy.

Future work should focus on refining anomaly detection algorithms to better differentiate between critical and non-critical anomalies, minimizing false positives and prioritizing interventions where they are most needed. Additionally, incorporating temporal analysis to observe how these anomalies evolve over time could provide further actionable insights. Exploring correlations with external variables such as user density, environmental interference, or hardware health may also enhance the ability to predict and mitigate network issues proactively. A first step towards this would be better labeling of the original dataset as we encountered missing labels where there shouldn't have been, leading to manual inputs (i.e. location labels in the classification model that analyzed San Joaquin). This would ease investigation efforts and lead to less time wasted on mapping correlation to causation for performing root cause analysis. Finally, extending this approach to other geographic areas or scaling it to larger datasets could validate its generalizability and effectiveness in broader network management scenarios.

By addressing these areas, our approach can further enhance network reliability and provide valuable insights for campus network management.

Related

Github Repository: https://github.com/frederickkiessling/CS_190N_Project