Network Behavior

Matthew Cariseo, Sam Howard, Matthew Szpak, Evan White
Computer Science, University of New Hampshire, Durham, NH 03824

Introduction

UNH's Palo Alto firewalls generate tens of thousands of log entries daily, capturing every allowed, denied, and dropped network session with metadata including source IP, country, application, and matched rule. Without automated tooling, identifying meaningful deviations, such as a country suddenly accounting for unexpected traffic volume or a rare application spiking which requires hours of manual analysis that security staff cannot sustain at scale.

This project delivers an automated firewall analytics platform: a daily log ingestion and anomaly detection pipeline paired with a web dashboard that surfaces results without requiring SQL knowledge or direct log file access.

Requirements

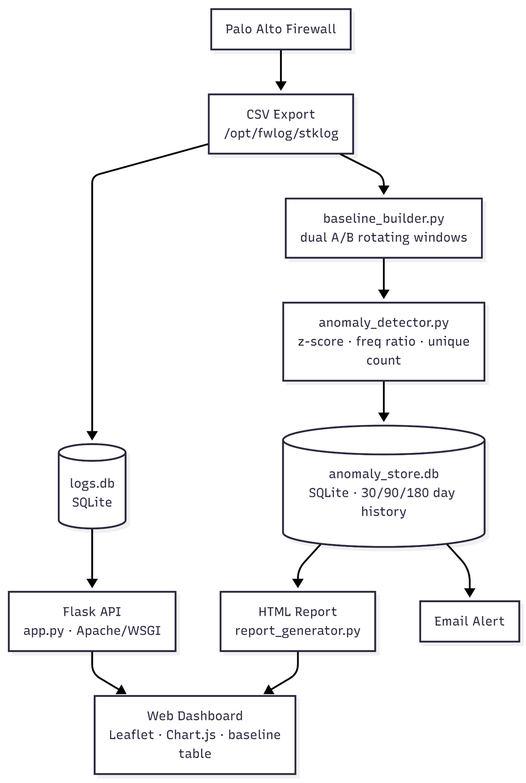
Functional Requirements:

- Ingest daily Palo Alto firewall CSV logs (5,000,000+ rows/day) automatically via cron
- Store logs in a queryable database supporting filtering by action, source country, application, rule, and IP address
- Compute rolling statistical baselines (30-day, 90-day, 180-day)
- Detect three anomaly types:
  - Categorical frequency shifts
  - Numeric z-score deviations (σ ≥ 3.0)
  - Unique-value-count changes (ratio ≥ 1.5×)
- Score anomalies (LOW / MED/ HIGH / CRITICAL) with bonuses for:
  - Security-relevant fields
  - Persistent trends
  - High-volume days
- Generate daily HTML anomaly reports and send email alerts above a configurable threshold
- Serve a web dashboard with geographic, timeline, and baseline comparison views
- Support log search with column-specific filters and custom SQL WHERE clauses

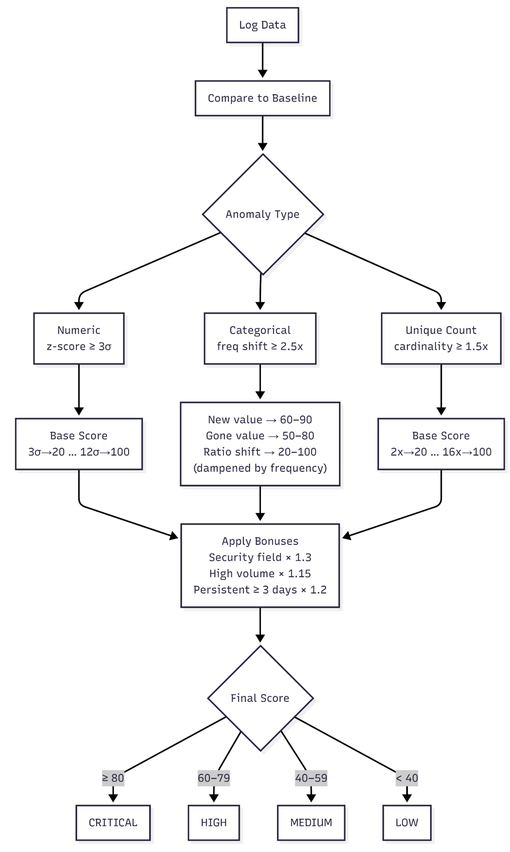
Non-functional Requirements:

- Process a full day's logs in under 5 minutes
- Dashboard initial load time under 15 seconds

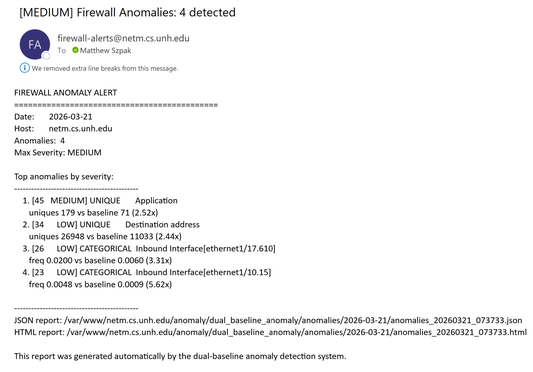
System Architecture



Scoring Chart



Anomaly Email



[MEDIUM] Firewall Anomalies: 4 detected

FA firewall-alerts@netm.cs.unh.edu
To Matthew Szpak

We removed extra line breaks from this message.

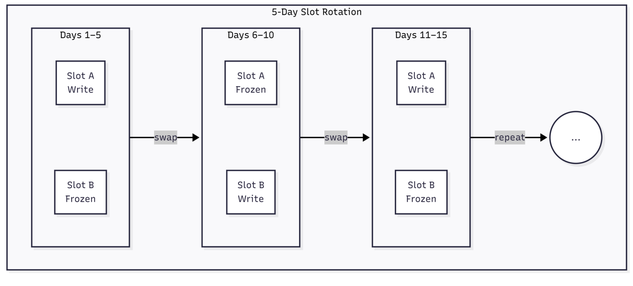
FIREWALL ANOMALY ALERT
==========
Date: 2026-03-21
Host: netm.cs.unh.edu
Anomalies: 4
Max Severity: MEDIUM

Top anomalies by severity:
---
1. [45 MEDIUM] UNIQUE Application
uniques 179 vs baseline 71 (2.52x)
2. [34 LOW] UNIQUE Destination address
uniques 26948 vs baseline 11033 (2.44x)
3. [26 LOW] CATEGORICAL Inbound Interface[ethernet1/17.610]
freq 0.0200 vs baseline 0.0060 (3.31x)
4. [23 LOW] CATEGORICAL Inbound Interface[ethernet1/10.15]
freq 0.0048 vs baseline 0.0009 (5.62x)

JSON report: /var/www/netm.cs.unh.edu/anomaly/dual_baseline_anomaly/anomalies/2026-03-21/anomalies_20260321_073733.json
HTML report: /var/www/netm.cs.unh.edu/anomaly/dual_baseline_anomaly/anomalies/2026-03-21/anomalies_20260321_073733.html

This report was generated automatically by the dual-baseline anomaly detection system.

A/B baseline swap



Implementation

The project uses:

- Python 3 with pandas (chunked CSV processing)
- PyYAML (centralized configuration)
- sqlite3 (log and anomaly storage)

Flask serves the web layer via Apache/WSGI on a UNH CS network server.

- Chart.js renders the traffic timeline with server-side caching
- Leaflet.js with CartoDB dark tiles renders the geographic choropleth map

All detection thresholds, scoring breakpoints, email recipients, and pipeline paths are defined in a single config yaml file, making the system tunable without code changes.

The full pipeline of ingest, baseline, detect, report, email is orchestrated by a shell script invoked by cron each morning.

Evaluation and Conclusion

The system satisfies all primary goals:

- Automated daily reports delivered before business hours
- Severity model produces actionable HIGH/CRITICAL alerts
- LOW-severity noise retained in reports (not emails)
- Dampening system suppresses false positives from sparse fields
- Dashboard provides full traffic context without SQL or log access

Limitations:

- Dependent on quality of Palo Alto CSV exports
- Sparse/noisy fields may still produce occasional false positives
- A/B baseline rotation can partially absorb sustained anomalies
- Purely statistical approach; lacks semantic threat understanding

Acknowledgements

Project Sponsor: Scott Kitterman
Project Advisor: Lisa Henry