Global Network Traffic Analysis Report

Investigation of Suspicious Web Traffic Events from WAF and VPC Flow Logs

Project Metadata

Project Prepared By: Gurram Sree Pavani

Date of Preparation: 2025-12-03

Contact Information: sreepavani8112001@gmail.com

1. Problem Statement

An internet-facing production HTTPS application hosted within an AWS VPC and protected by a Web Application Firewall (WAF) is generating repeated alerts categorized as “Suspicious Web Traffic.” These alerts correspond to anomalous HTTP(S) request patterns that may indicate probing, scanning, or malicious attempts to interact with the application’s endpoints. Ensuring visibility into these events is critical for maintaining the integrity and security of the application stack.

The dataset examined in this analysis consists of 282 WAF and VPC flow log entries. Each event includes detailed network telemetry and enriched security attributes such as bytes_in, bytes_out, timestamps (creation_time and end_time), source IP and destination IP values, country codes, protocol information (HTTPS), destination port (443), detection types (waf_rule), rule names, metadata fields, and response codes. These fields collectively allow correlation of network behavior with security rule triggers and potential adversarial actions.

Key business and security-driven questions arise from this data: identifying which countries and IP addresses generate suspicious traffic, understanding temporal fluctuations in event volume, and detecting extreme spikes in inbound or outbound bytes that may signal reconnaissance or data exfiltration attempts. These questions form the foundation for improving firewall tuning, initiating targeted investigations, and enhancing the organization’s threat detection posture.

The overarching objective of the analysis is to characterize global patterns of suspicious web traffic, rank the highest-risk source entities, quantify traffic volume, and provide actionable recommendations supporting cybersecurity operations and incident response workflows.

2. Techniques and Tools Used

2.1 Data Collection & Understanding

- Imported WAF and VPC flow logs into Pandas DataFrames
- Reviewed network, timing, security, and metadata fields
- Understood structure: 282 events, 16 relevant fields

2.2 Data Cleaning & Preprocessing

- Checked for missing values across bytes, timestamps, IPs, protocols
- Standardized timestamps and numeric data types
- Created time-based derived fields (date, hour, Timeonly)
- Validated response codes and port consistency
- Ensured country codes follow ISO standards

2.3 Exploratory Data Analysis (EDA)

- Generated descriptive stats for bytes_in/out
- Grouped events by source IP and source country
- Computed total bytes and distinct IP/country counts
- Analyzed event trends over time

2.4 Data Quality & Outlier Analysis

- Identified extreme spikes in bytes_in/out
- Flagged anomalous flows for deeper investigation
- Verified consistency of waf_rule detection type

2.5 Visualization & Dashboard (Power BI)

- Developed KPIs for event count, total bytes, distinct IPs and countries
- Built time-series, bar charts, and detailed event tables
- Filtered exclusively for suspicious HTTPS traffic

3. Tools & Technologies Summary

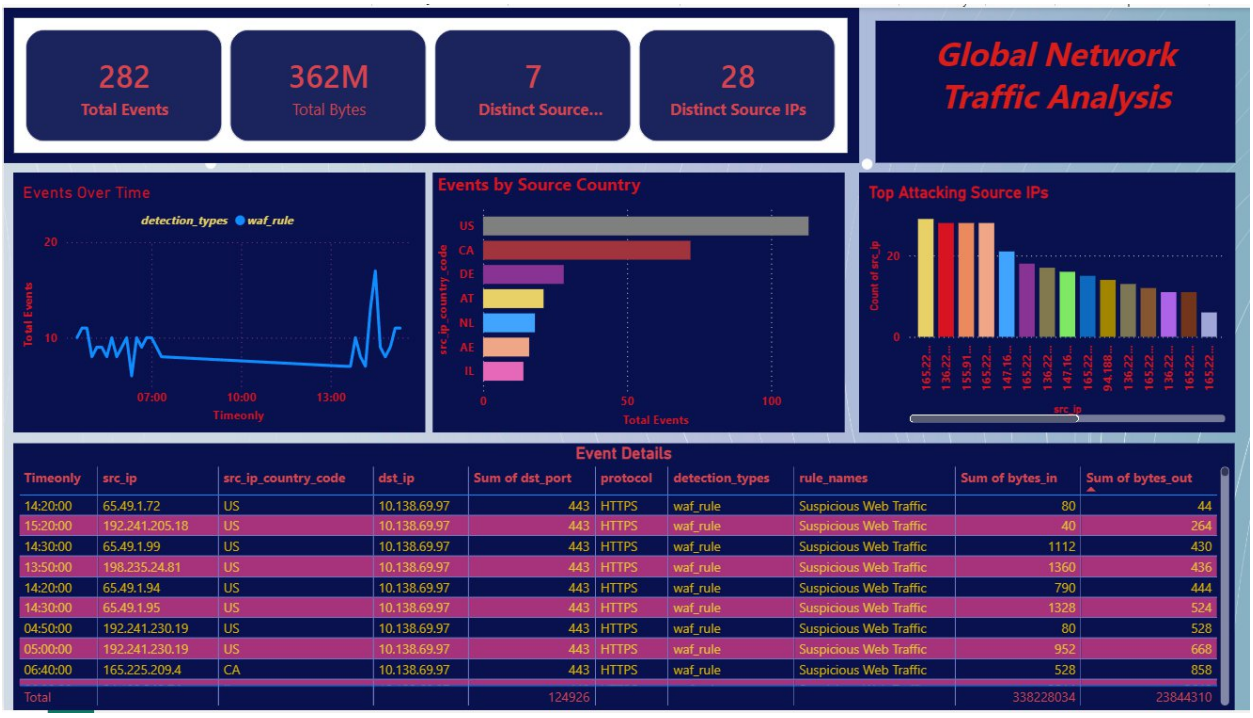
Category	Tools / Technologies	Purpose
Data Processing	Python, Pandas, NumPy	Data loading, cleaning, aggregation
Notebook Env	Jupyter Notebook	Interactive EDA and documentation
Cloud Platform	AWS VPC Flow Logs, WAF logs	Source of traffic data
Visualization	Microsoft Power BI	Interactive dashboards and KPIs
Storage / Format	CSV or JSON	Raw and cleaned data formats

4. Key Findings & Security Insights

- Total of 282 suspicious HTTPS events targeting 10.138.69.97 on port 443
- Majority of events originate from a small number of countries (US, CA leading)
- 28 distinct source IPs, several producing high inbound byte counts (>25M)
- Repeated successful sessions with response.code 200 despite WAF alerts
- Recommendations: tighten WAF rules, geo-block high-risk regions, monitor large flows

5. Dashboard Screenshots & Interpretation

Figure 1: Global Network Traffic Analysis Dashboard



Timeonly	src_ip	src_ip_country_code	dst_ip	Sum of dst_port	protocol	detection_types	rule_names	Sum of bytes_in	Sum of bytes_out
14:20:00	65.49.1.72	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	80	44
15:20:00	192.241.205.18	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	40	264
14:30:00	65.49.1.99	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	1112	430
13:50:00	198.235.24.81	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	1360	436
14:20:00	65.49.1.94	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	790	444
14:30:00	65.49.1.95	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	1328	524
04:50:00	192.241.230.19	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	80	528
05:00:00	192.241.230.19	US	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	952	668
06:40:00	165.225.209.4	CA	10.138.69.97	443	HTTPS	waf_rule	Suspicious Web Traffic	528	858
Total				124926				338228034	23844310

6. Conclusion & Recommendations

The analysis demonstrates that suspicious web traffic is concentrated among a limited set of international sources, with several IPs exhibiting potentially malicious behavior such as high-volume data transfers or repeated probing attempts. Strengthening WAF configurations, applying rate limits, and monitoring anomalous connections are recommended next steps. Integration with SIEM platforms and machine learning-based anomaly detection could significantly enhance long-term threat visibility.