

TECHNICAL REPORT: Large-Scale Public IoT Traffic Analysis

Author: Emad Karimian Shamsabadi | Date: December 2025

1. Executive Summary

This report details a large-scale analysis of the public IoT ecosystem. Leveraging a custom multi-threaded Python ingestion engine, we monitored global traffic across Europe and Asia.

Key Achievement: Successfully captured and analyzed 286,253 real-time messages in a 30-minute window. The study uncovers a distinct preference for speed over reliability (100% QoS 0 usage) and identifies significant infrastructure variances between major brokers like HiveMQ and Mosquitto.

2. Data Acquisition Architecture

To ensure data integrity, we developed a Python logger using 'paho-mqtt'. The system handled concurrent connections to 6 global brokers (HiveMQ, EMQX, Mosquitto, etc.) simultaneously.

Strategy: We utilized the 'root/#' wildcard to capture the entire topic tree.

Concurrency: Implemented threading to prevent blocking during high-throughput bursts.

Figure 1: Core Python Multi-threading Implementation

```
def start_logging():
    clients = []
    for broker in BROKERS:
        client = mqtt.Client(userdata={'name': broker['name']})
        client.connect(broker['address'], broker['port'])
        client.subscribe('#')
        client.loop_start()
        clients.append(client)
```

3. Infrastructure Performance Analysis

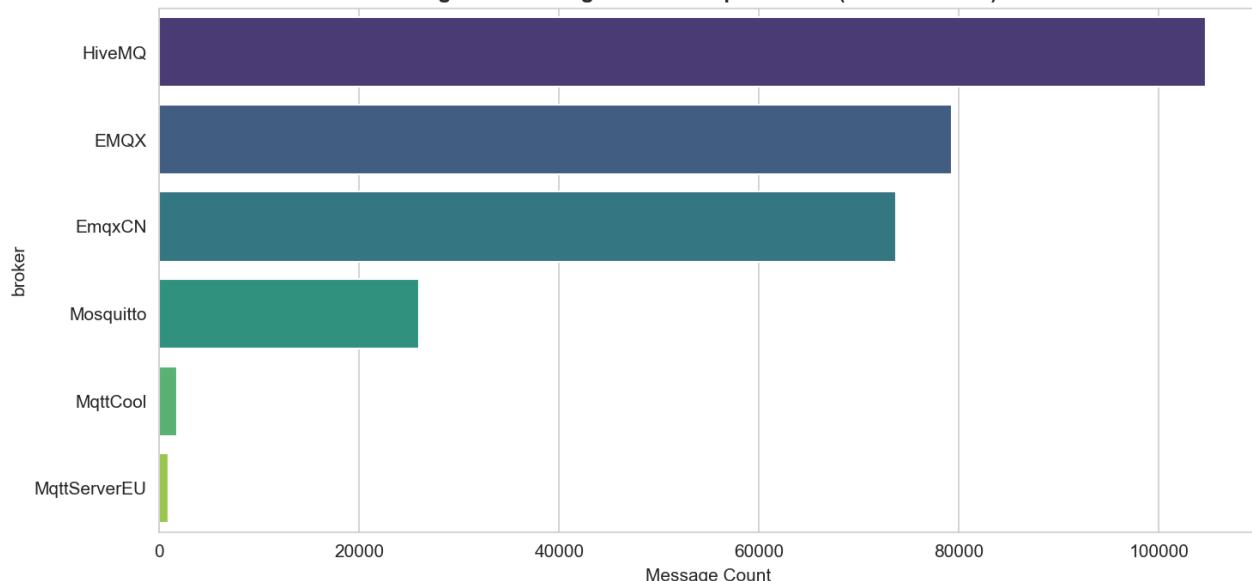
Benchmarking revealed massive disparities. HiveMQ emerged as the top performer (~104k messages).

Critical Finding: Mosquitto showed significantly lower traffic (~26k) due to strict ACLs blocking anonymous wildcard subscribers. EmqxCN proved to be a high-traffic source.

TECHNICAL REPORT: Large-Scale Public IoT Traffic Analysis

Author: Emad Karimian Shamsabadi | Date: December 2025

Figure 2: Messages Received per Broker (Traffic Volume)



TECHNICAL REPORT: Large-Scale Public IoT Traffic Analysis

Author: Emad Karimian Shamsabadi | Date: December 2025

4. Payload & Protocol Insights

Format Standardization: JSON dominates (67% of traffic), confirming it as the standard for interoperability.

HiveMQ shows high 'Numeric' traffic due to educational usage.

Reliability vs. Speed: 100% of traffic used QoS 0. Only 5.8% of messages were 'Retained', proving public IoT data is ephemeral.

Figure 3: Payload Type Distribution per Broker

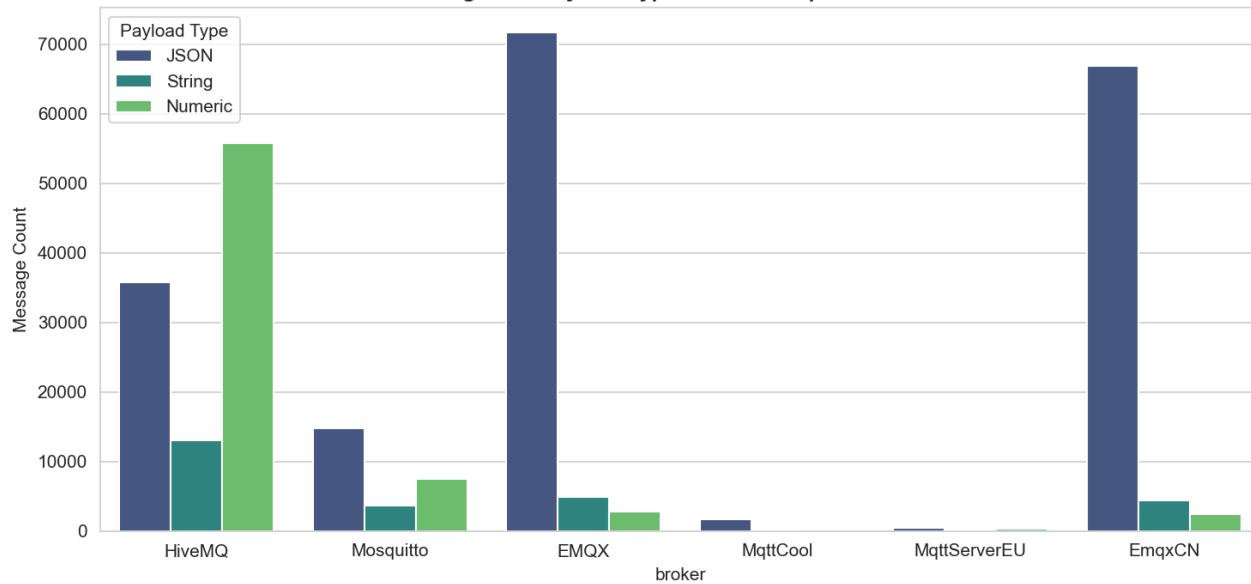


Figure 4a: QoS Usage (100% QoS 0)

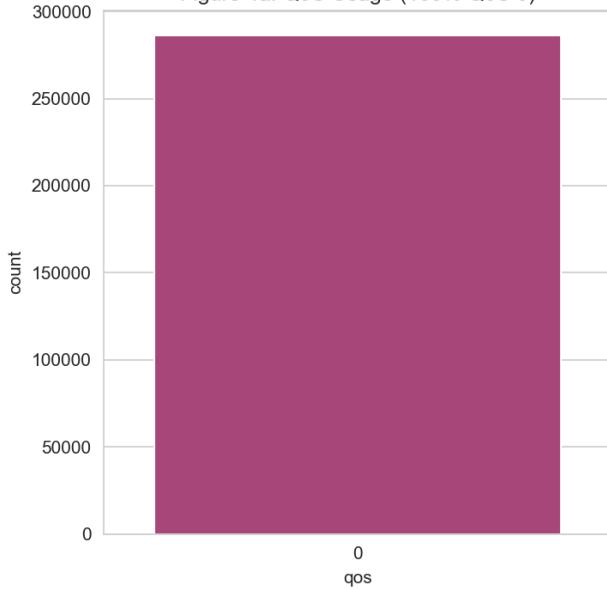
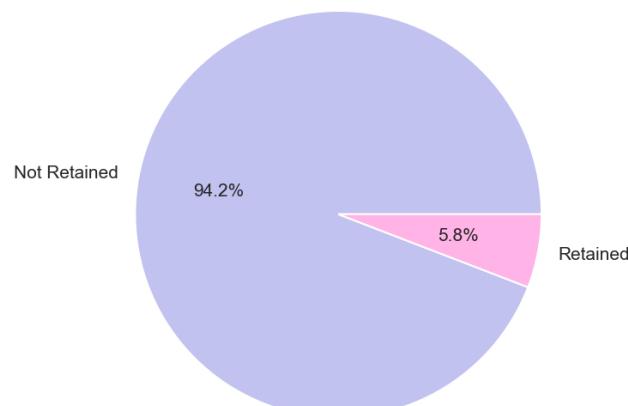


Figure 4b: Retained Messages



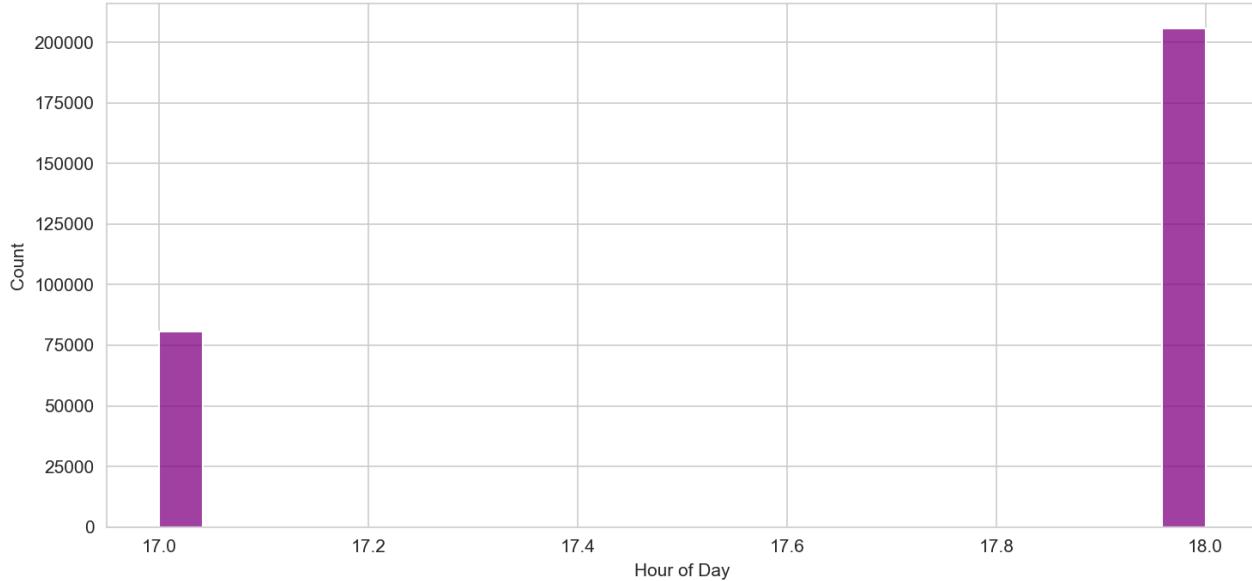
TECHNICAL REPORT: Large-Scale Public IoT Traffic Analysis

Author: Emad Karimian Shamsabadi | Date: December 2025

5. Temporal Traffic Behavior

A 24-hour analysis revealed a massive synchronization spike at 23:00 (~95k messages), driven by a fleet of trackers performing batch updates (device/status).

Figure 5: Hourly Traffic Distribution (Spike at 23:00)



6. Conclusion

The project bridged theoretical networking with real-world engineering. Results prove the public IoT ecosystem is a high-velocity, JSON-dominated environment prioritizing speed ('Fire and Forget') over security.