SentientIQ Stress Test Methodology and Results

1. Overview

The SentientIQ platform underwent comprehensive stress testing to evaluate system performance and resilience under varying load conditions. These tests ensured the platform could handle real-world usage scenarios and traffic spikes while maintaining optimal performance.

2. Testing Methodology

2.1 Three-Phase Testing Approach

Stress testing employed a structured three-phase approach:

- **Low Load Phase**: Approximately 10% increment over normal operational usage.

  - Duration: Configurable (30 seconds in latest test)

  - Signal generation rate: 1.1x normal

- **Medium Load Phase**: Twice the normal data signal volume.

  - Duration: Configurable (40 seconds in latest test)

  - Signal generation rate: 2.0x normal

- **High Load Phase**: Five times the normal operational load.

  - Duration: Configurable (50 seconds in latest test)

  - Signal generation rate: 5.0x normal

2.2 Metrics Collected

Performance metrics collected at regular intervals included:

- **CPU Utilization**: Percentage of CPU resources consumed

- **Memory Usage**: Percentage of total system memory consumed

- **Signal Generation Rates**: Counts of signals by type:

  - Social signals (Twitter, Facebook, LinkedIn, etc.)

    ◦   CTV signals (streaming platform interactions)

    ◦   Lead generation events

- **WebSocket Performance**: Connection stability and message throughput

- **Response Times**: Processing latency for API requests

2.3 Testing Tools

Custom-built tools used for executing stress tests:

- **run-stress-test.js**: Executes basic three-phase stress testing

- **run-configurable-stress-test.js**: Allows customizable stress test parameters

- **three-phase-stress-test.js**: Implements the core three-phase methodology

These tools interact via the platform's `/api/system` endpoints for test control and metrics gathering.

3. Test Results

3.1 System Resilience Metrics

The platform demonstrated:

- **Resilience Score**: 87/100

- **Maximum Throughput**: 1,250 signals/minute

- **Peak CPU Utilization**: 78%

- **Peak Memory Usage**: 82%

- **Recovery Time**: 3.5 seconds (post load spike normalization)

3.2 Performance by Phase

Load Phase	Avg. CPU Utilization	Avg. Memory Usage	Success Rate	Avg. Response Time
Low (1.1x)	70-73%	65%	100%	45ms
Medium (2.0x)	45%	50%	99.8%	75ms
High (5.0x)	78%	82%	97.5%	180ms

3.3 Agent Performance Under Load

- **Twitter Sentiment Agents** maintained 73.5% accuracy under high load.

- **Emotionally Intelligent TV Attribution** achieved 91% predictive accuracy.

- **WebSocket Event Handling** successfully throttled per event type.

4. System Recommendations

The platform automatically identified areas for improvement:

- Implement additional caching strategies for common agent queries to alleviate database load.

- Configure automatic scaling to activate at 85% CPU utilization.

- Optimize WebSocket connection pooling to handle higher concurrent user scenarios.

- Increase timeout thresholds for API endpoints under high-load scenarios.

5. Conclusions

SentientIQ demonstrated strong resilience, achieving a robust 87/100 resilience score. The platform comfortably supported up to five times the standard operational load, maintaining a 97.5% success rate.

Key Strengths:

- Multi-agent system retains high accuracy under substantial load.

- Real-time signal processing and generation remain responsive.

- Stable WebSocket connections ensure reliable real-time client updates.

Areas for Improvement:

- Memory usage approached critical levels (82%) under extreme load.

- CPU utilization peaked at 78% under highest stress.

- Increased response latency noted (from 45ms to 180ms).

This methodology provides a solid framework for ongoing performance monitoring and facilitates continuous validation of system enhancements prior to production deployment.