**Possible Issues**

1. **Sequential API Calls**

The middleware might be calling APIs **one after another** instead of in parallel.

Fix: Use Promise.all() (Node.js) or multi-curl (PHP) to make concurrent requests.

1. **Blocking Middleware**

The rate-limiter could be **pausing** requests rather than rejecting/queuing them efficiently.

Fix: Switch to a **token bucket algorithm** or non-blocking queue.

1. **Redis Latency**

If Redis is hosted remotely or under heavy load, each check adds latency.

Fix: Use local in-memory cache for short-term throttling, fallback to Redis for persistence.