

Rate Limiting and Throttling Policy

Document Version: 1.0

Last Updated: [Date]

1. Purpose

This document outlines the rate limiting and throttling policies applied to the [API/Service Name] to ensure fair and stable access for all users.

2. Scope

The policy applies to all consumers of [API/Service Name] including public, partner, and internal users.

3. Definitions

- Rate Limiting:** Restricts the number of requests a user/application can make in a given time period.
- Throttling:** Temporarily reduces or blocks requests from users/applications that exceed limits.
- User:** An account or application accessing the API/service.

4. Policy Statement

To ensure system stability and availability, the following rate limits and throttling rules are enforced.

5. Rate Limits

User Type	Limit	Time Window	Scope
Public User	1000 requests	24 hours	Per API Key
Partner	100 requests	1 minute	Per IP
Internal	Unlimited*	N/A	Per Account

*Subject to monitoring and abuse prevention.

6. Throttling Actions

- Requests over the limit will receive a **429 Too Many Requests** response.
- Retry-After headers may be sent to indicate wait time before retrying.
- Repeated violations may result in temporary or permanent suspension of service.

7. Exemptions

Exemptions can be granted under special circumstances on a case-by-case basis with approval from [Contact/Team].

8. Monitoring & Enforcement

Usage patterns are monitored continuously. Violations are logged and may trigger automated or manual enforcement actions.

9. Contact

For questions or exemption requests, contact: [support@email.com].