

API Rate Limiting and Throttling Policy Guide

Overview

This guide explains the policies and recommendations for API rate limiting and throttling, ensuring reliable service and fair usage among all clients.

Definitions

- **Rate Limiting:** Restricts the number of API requests a client can make within a defined period.
- **Throttling:** Controls API traffic to prevent overload, usually by temporarily rejecting or slowing down requests.

Standard Rate Limits

API Plan	Requests per Minute	Burst Limit
Free	60	10
Standard	600	60
Enterprise	6000	600

Headers

The API returns the following HTTP headers to help you manage your rate limit:

- `X-RateLimit-Limit` : Maximum number of requests allowed.
- `X-RateLimit-Remaining` : Requests remaining in the current window.
- `X-RateLimit-Reset` : Time when the rate limit window resets (in UTC epoch seconds).

```
HTTP/1.1 429 Too Many Requests
X-RateLimit-Limit: 60
X-RateLimit-Remaining: 0
X-RateLimit-Reset: 1717995600
Retry-After: 60
```

Best Practices

- Monitor rate limit headers and adjust usage accordingly.
- Implement exponential backoff retries for 429 status codes.
- Avoid making unnecessary API calls.
- Cache responses when possible.

Error Response

```
{
  "error": "rate_limit_exceeded",
  "message": "API rate limit exceeded. Please wait and retry your request."
}
```

Contact & Support

For questions or increased limits, contact

