RateLimit Policy Proposal

HTTPAPI Working Group - IETF 117 San Francisco

July 27th, 2023

Darrel Miller

Draft-07

100 units within 60 seconds (protection)

10000 units within a day (quota)

RateLimit-Policy: 100;w=60, 10000;w=86400

{ 50 requests and 30 seconds}

RateLimit: limit=100, remaining=50, reset=30 // protection limit

{ 1 requests and 40 seconds}

RateLimit: limit=10000, remaining=90, reset=200 // quota limit

Assumption: quota units for different policies are the same for the same resource, however different requests may consume different quantity of units.

Proposal 1

RateLimit-Policy: protection;l=100;w=60, quota;l=10000;w=86400

{ 50 requests and 30 seconds}

RateLimit: policy=protection, remaining=50, reset=30

{ 1 request and 40 seconds}

RateLimit: policy=quota, remaining=90, reset=200

Policy key ties current usage state to declared policy

Quota units can be different for requests or by policy.

Should policy key be required in RateLimit field?

Can RateLimit headers be returned for different policies?

Should the window/reset be optional for "free trial" scenario?

Cannot return limits for multiple policies.

Proposal 2

RateLimit-Policy: protection;l=100;w=60, quota;l=10000;w=86400

{ 50 requests and 30 seconds}

RateLimit: protection;r=50; t=30

{ 1 request and 40 seconds}

RateLimit: quota;r=90; t=200

Using sfItem to allow reporting on multiple policies

Don't need to declare that key values don't allow parameters

Could enable extension parameters in the RateLimit field.