

## Lab 14 - Rate Limiting

The Gloo Platform Rate Limiting API allows for very fine grained configuration to meet the needs of the end user. First a user must decide the type of rate limiting they want to enforce. Below is some common types of rate limiting.

Types of rate limiting

- Global inbound - rate limit x number of requests per second
- User based - User can make x number of requests per second
- Per backend - Prevent too many requests from overwhelming a given endpoint

Links:

- [Rate Limit Docs](#)
- [RateLimitServerConfig API](#)
- [RateLimitServerSettings API](#)
- [RateLimit API](#)
- [RateLimitClientConfig API](#)
- [RateLimitPolicy API](#)
- To better view the ratelimiting, we are going to disable circuit breaking if its enabled

```
kubectl delete outlierdetectionpolicy outlier-detection --context management -n app-team
```

- Apply the RateLimitPolicy to enable rate limiting

```
kubectl apply --context management -f data/rate-limit-policy.yaml
```

- Refresh UI a few times should see `x-envoy-ratelimited: [true]`

```
Trailers:
  content-type: [application/grpc]
  date: [Wed, 08 Mar 2023 14:42:05 GMT]
  server: [envoy]
  x-envoy-ratelimited: [true]
  x-envoy-upstream-service-time: [2]
```

- Remove the RateLimitPolicy

```
kubectl delete --context management -f data/rate-limit-policy.yaml
```