

This is not the original Google Slides file. This is a copy.

# Perfectly (Load) Balanced, as all requests should be

Christine Kim

Nim Jayawardena



#IstioCon



Christine Kim (@christineskim\_)

Developer Programs Engineer, Google Cloud



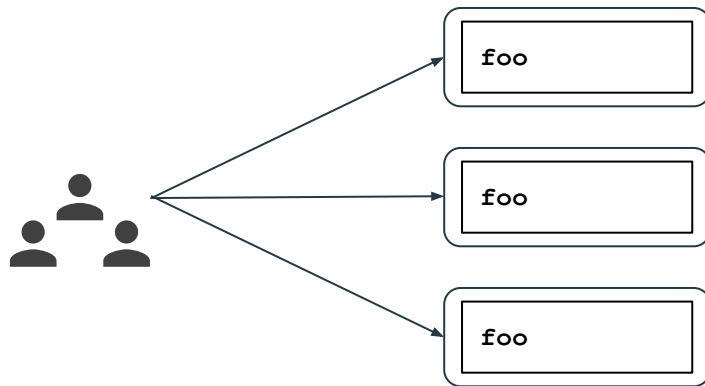
Nim (@NimTheHuman)

Developer Programs Engineer, Google Cloud



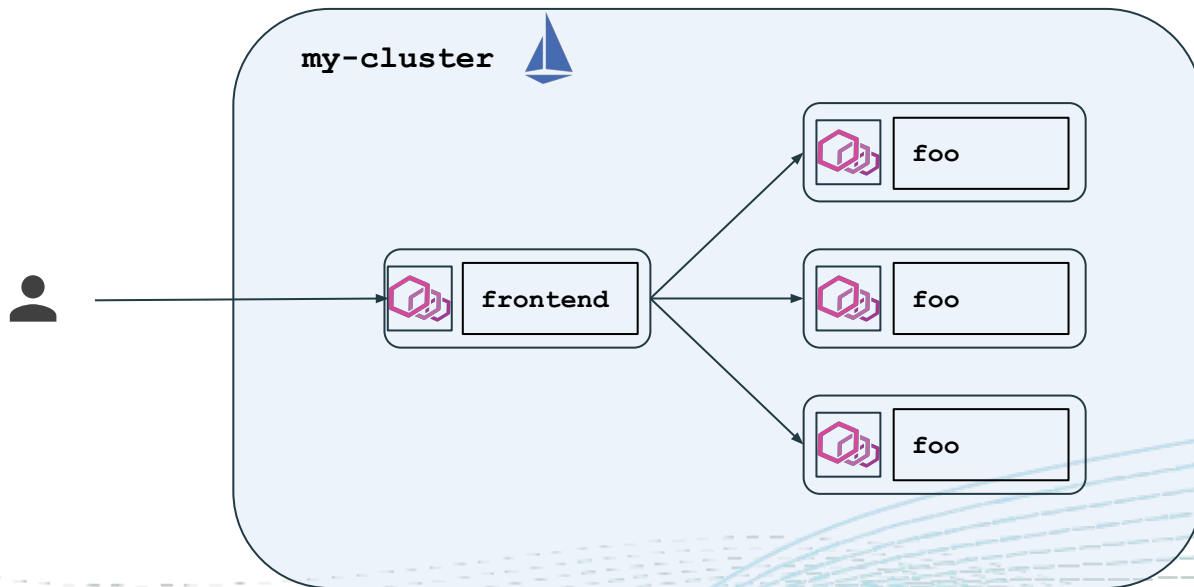
# What is load balancing?

- Distributes incoming network traffic to across to backend services
- A simple example:



# What is load balancing?

- Distributes incoming network traffic to across to backend services
- Picture a single cluster:



# LB with Istio out of the box

Leveraging Envoy, Istio offers:

- Round Robin (*default for now*) [ROUND\_ROBIN]
- Random [RANDOM]
- Least requests [LEAST\_CONN]

```
apiVersion:
networking.istio.io/v1alpha3
kind: DestinationRule
metadata:
  name: bookinfo-ratings
spec:
  host: ratings.prod.svc.cluster.local
  trafficPolicy:
    loadBalancer:
      simple: ROUND_ROBIN
  subsets:
    - name: v1
      labels:
        version: v1
    - name: v2
      labels:
        version: v2
      trafficPolicy:
        loadBalancer:
          simple: RANDOM
```

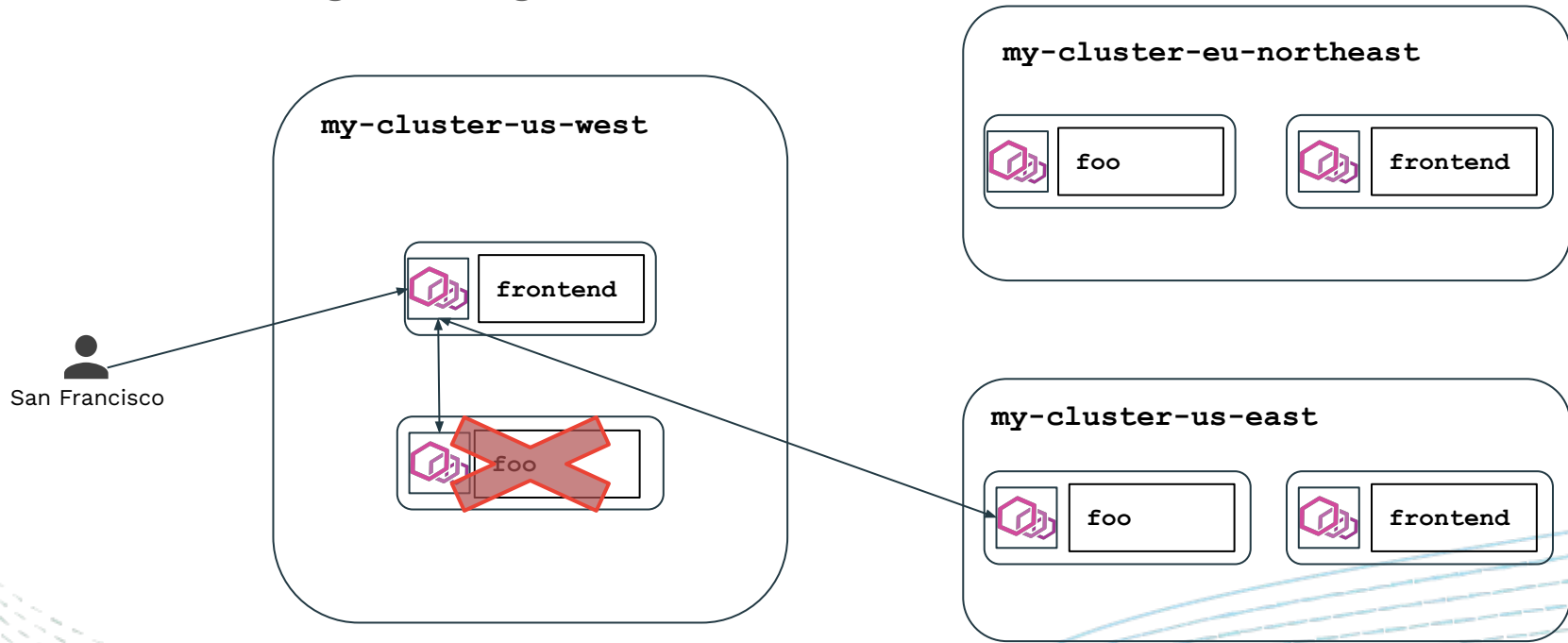
<https://istio.io/latest/docs/concepts/traffic-management/#load-balancing-options>

#IstioCon



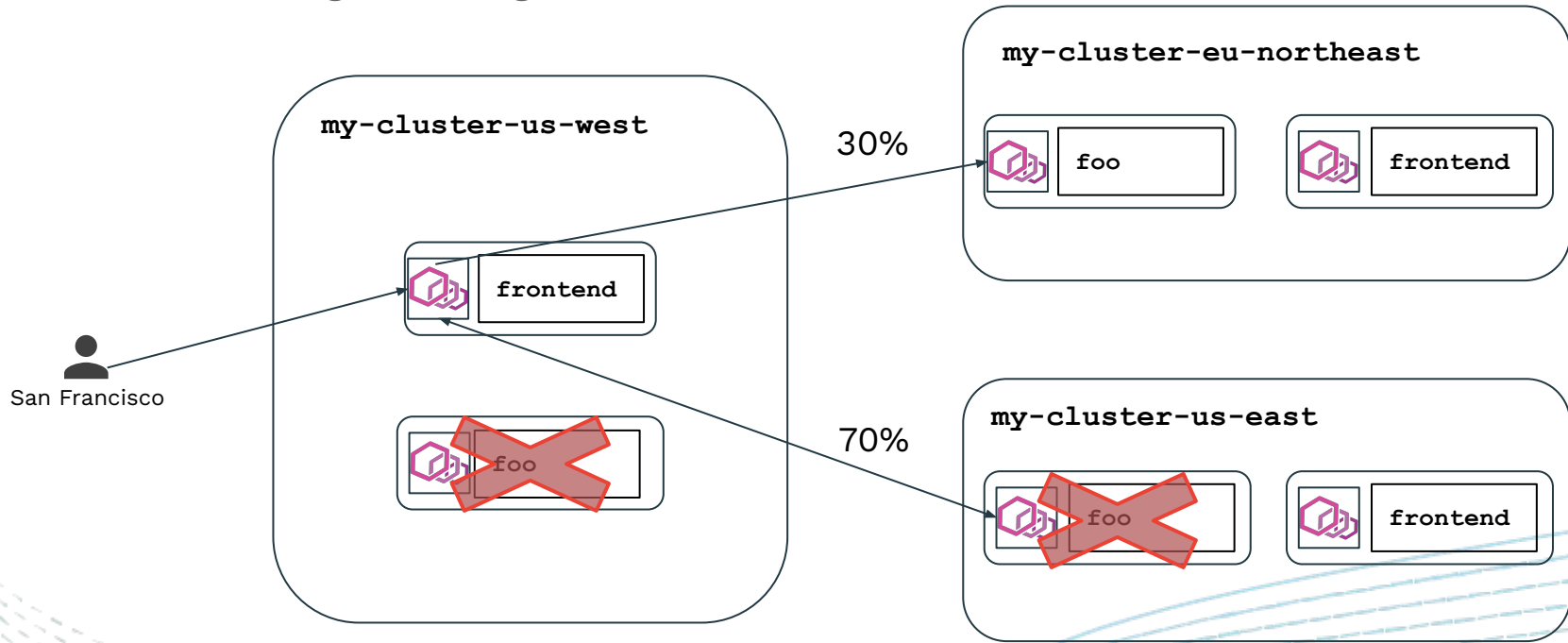
# Let's scale up

- Picture a high scale, global cluster:



# But wait... there's more

- Picture a high scale, global cluster:



## 1. Weighted Least Request (WLR)

- Distributes traffic away from stressed endpoints while respecting weights
- This will be the new default (renamed to **LEAST\_REQUEST**)

## 2. Modified Locality Failover

- Improves distribution of traffic when there are only few endpoints/locality

### Better Load Balancing by Default

Shared with Istio Community



Owners: [nathanmittler@google.com](mailto:nathanmittler@google.com)  
[howardjohn@google.com](mailto:howardjohn@google.com)  
Working Group: networking

Status: WIP | In Review | Approved | Obsolete  
Created: 2021-12-01  
Approvers: WG-lead-X [], WG-lead-Y [], ...

#### TL;DR

Address issues with locality failover load balancing so that it is safe to enable by default, providing a better customer experience for multi-cluster out-of-the-box.

#### Overview

The current default load balancing behavior in Istio is round-robin across all endpoints in the mesh, regardless of locality.





# Demo of Locality Failover

#IstioCon



[Pull requests](#)
[Issues](#)
[Marketplace](#)
[Explore](#)

[GoogleCloudPlatform / microservices-demo](#)
Public
Unwatch 237
Fork 3.6k
Star 12k

[Code](#)
[Issues 35](#)
[Pull requests 8](#)
[Discussions](#)
[Actions](#)
[Projects](#)
[Wiki](#)
[Security](#)
[Insights](#)
[Settings](#)

[main](#)
18 branches
19 tags

Go to file
Add file
Code

[mathieu-benoit](#)
[GRPC\\_HEALTH\\_PROBE\\_VERSION=v0.4.11 \(#803\)](#)
✓ 5686ca8 1 hour ago
682 commits

.github	Upgrade Golang microservices to Go 1.18.0 (#790)	21 days ago
docs	NetworkPolicy resources (#778)	last month
hack	Move YAML document separator to top of release manifests (#799)	7 days ago
istio-manifests	Replace "whitelist" with "allow" in Istio ServiceEntry (#489)	15 months ago
kubernetes-manifests	Turn off GCP integrations by default (#665)	4 months ago
pb	Update Product List to Match Cymbal Branding/Story (#600)	7 months ago
release	Release v0.3.6 (#705)	3 months ago
src	GRPC_HEALTH_PROBE_VERSION=v0.4.11 (#803)	1 hour ago
.gitignore	Add instructions for workload identity-enabled GKE clusters (#423)	2 years ago
CODEOWNERS	Update CODEOWNERS (#573)	10 months ago
CODE_OF_CONDUCT.md	chore: add a Code of Conduct (#677)	3 months ago
CONTRIBUTING.md	Add development principles (#56)	4 years ago
LICENSE	add LICENSE, CONTRIBUTING.md	4 years ago
README.md	NetworkPolicy resources (#778)	last month
SECURITY.md	chore: add SECURITY.md (#678)	3 months ago
cloudbuild.yaml	cleanup: add region tags to yaml and sh (#684)	3 months ago
renovate.json	Constrain python and added config for renovate bot (#780)	14 days ago
scaffold.yaml	Use scaffold modules & update docs (#609)	6 months ago

About

Sample cloud-native application with 10 microservices showcasing Kubernetes, Istio, gRPC and OpenCensus.

[onlineboutique.dev](#)

[kubernetes](#)
[google-cloud](#)
[grpc](#)
[samples](#)
[gke](#)
[istio](#)
[sample-application](#)
[opencensus](#)
[scaffold](#)

Readme

Apache-2.0 License

Code of conduct

12k stars

237 watching

3.6k forks

Releases 11

v0.3.6 (Latest) on Feb 2

+ 10 releases

Packages

No packages published

[Publish your first package](#)

Used by 20

12

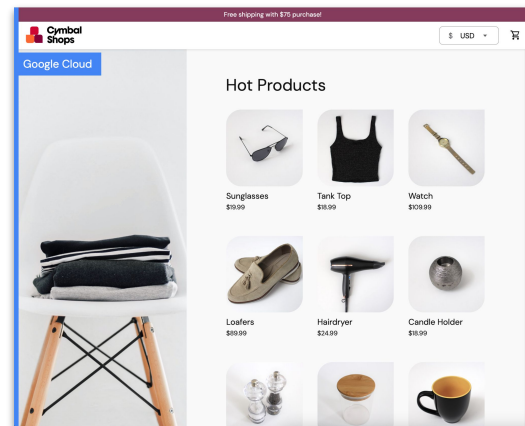
Contributors 96

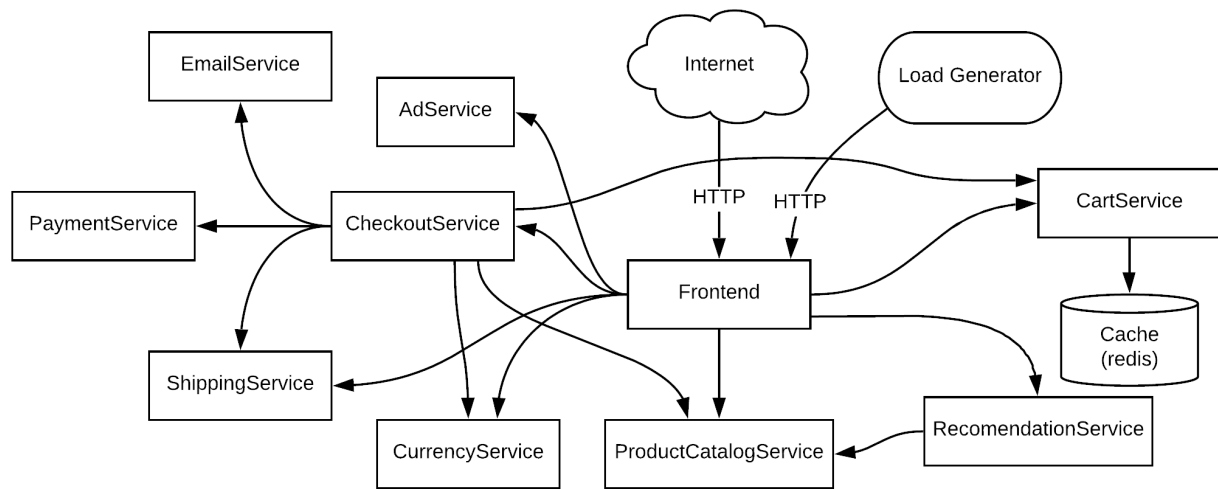
+ 85 contributors

Languages

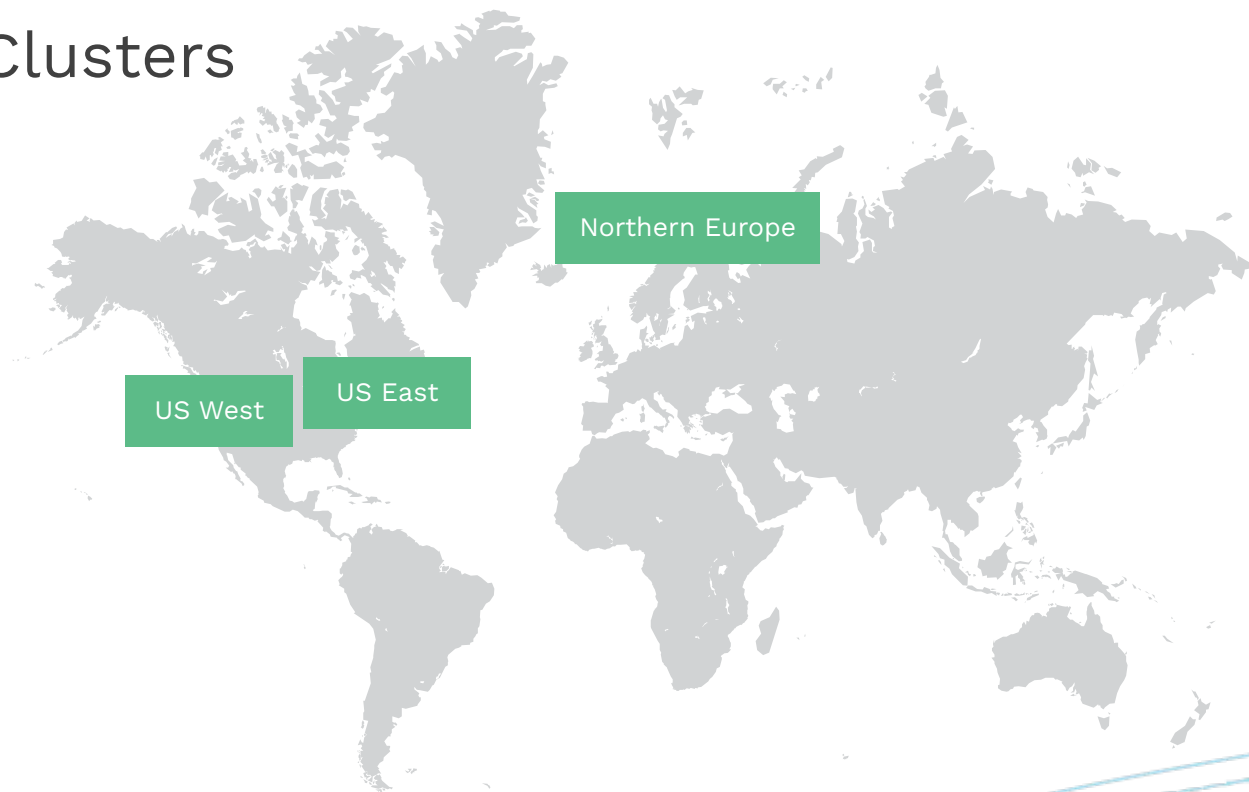
README.md

Continuous Integration - Main/Release passing





# 3 Clusters



#IstioCon



```
apiVersion: networking.istio.io/v1alpha3
kind: DestinationRule
metadata:
  name: frontend-service
spec:
  host: frontend.online-boutique.svc.cluster.local
  trafficPolicy:
    connectionPool:
      http:
        maxRequestsPerConnection: 1
    loadBalancer:
      simple: LEAST_CONN
    localityLbSetting:
      enabled: true
      failover:
        - from: us-east1
          to: us-west1
    outlierDetection:
      consecutive5xxErrors: 1
      interval: 1s
      baseEjectionTime: 1m
```



```
apiVersion: networking.istio.io/v1alpha3
kind: DestinationRule
metadata:
  name: frontend-service
spec:
  host: frontend.online-boutique.svc.cluster.local
  trafficPolicy:
    connectionPool:
      http:
        maxRequestsPerConnection: 1
    loadBalancer:
      simple: LEAST_CONN
      localityLbSetting:
        enabled: true
        failover:
          - from: us-east1
            to: us-west1
    outlierDetection:
      consecutive5xxErrors: 1
      interval: 1s
      baseEjectionTime: 1m
```

## Connection Pool

Force each HTTP request to create a new TCP connection (for demo purposes)



```
apiVersion: networking.istio.io/v1alpha3
kind: DestinationRule
metadata:
  name: frontend-service
spec:
  host: frontend.online-boutique.svc.cluster.local
  trafficPolicy:
    connectionPool:
      http:
        maxRequestsPerConnection: 1
    loadBalancer:
      simple: LEAST_CONN
      localityLbSetting:
        enabled: true
        failover:
          - from: us-east1
            to: us-west1
    outlierDetection:
      consecutive5xxErrors: 1
      interval: 1s
      baseEjectionTime: 1m
```

## Outlier detection

Defines what it means for an endpoint to be “unhealthy”



```
apiVersion: networking.istio.io/v1alpha3
kind: DestinationRule
metadata:
  name: frontend-service
spec:
  host: frontend.online-boutique.svc.cluster.local
  trafficPolicy:
    connectionPool:
      http:
        maxRequestsPerConnection: 1
    loadBalancer:
      simple: LEAST_CONN
      localityLbSetting:
        enabled: true
        failover:
          - from: us-east1
            to: us-west1
    outlierDetection:
      consecutive5xxErrors: 1
      interval: 1s
      baseEjectionTime: 1m
```





```
apiVersion: networking.istio.io/v1beta1
kind: DestinationRule
metadata:
  name: frontend-service
spec:
  host: frontend.online-boutique.svc.cluster.local
  trafficPolicy:
    connectionPool:
      http:
        maxRequestsPerConnection: 1
    trafficPolicy:
      loadBalancer:
        localityLbSetting:
          enabled: true
          distribute:
            - from: us-east1/*
              to:
                "us-west1/*": 70
                "europe-north1/*": 30
      outlierDetection:
        consecutive5xxErrors: 100
        interval: 1s
        baseEjectionTime: 1m
```



# Thank you!

Christine Kim (@christineskim\_)  
Nim (@NimTheHuman)

#IstioCon

