PicNIC
Predictable Virtualized NIC

Praveen Kumar

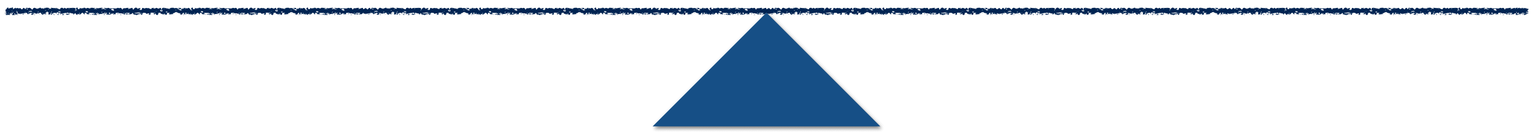
Nandita Dukkipati, Nathan Lewis, Yi Cui, Yaogong Wang, Chonggang Li,
Valas Valancius, Jake Adriaens, Steve Gribble, Nate Foster, Amin Vahdat

BEIJING
SIGCOMM
2019

Fundamental Tradeoff

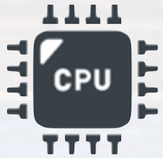
Performance Isolation

Efficiency

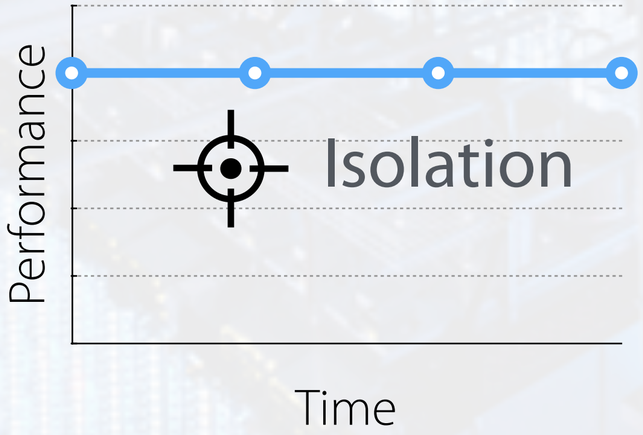


Strong Isolation vs. Efficiency

Tenant



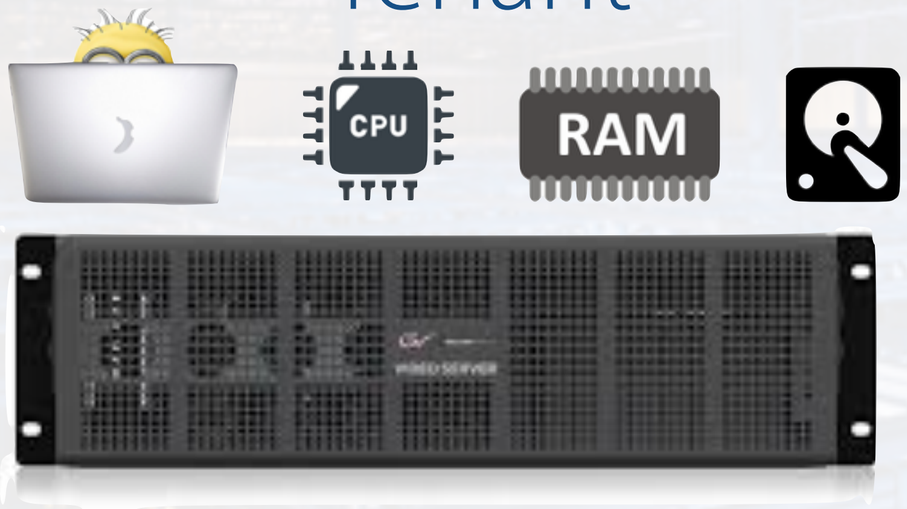
Performance



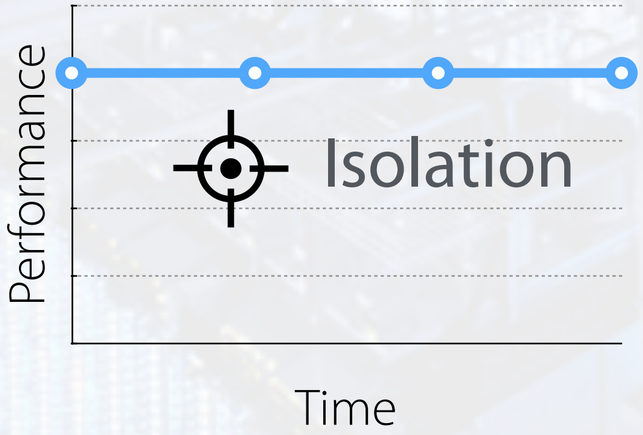
Provider

Strong Isolation vs. Efficiency

Tenant



Performance



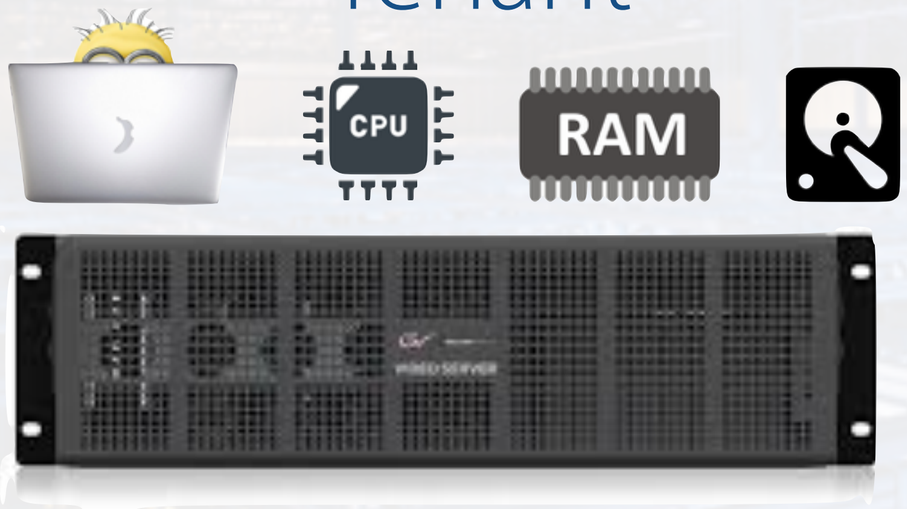
Provider

Efficient utilization

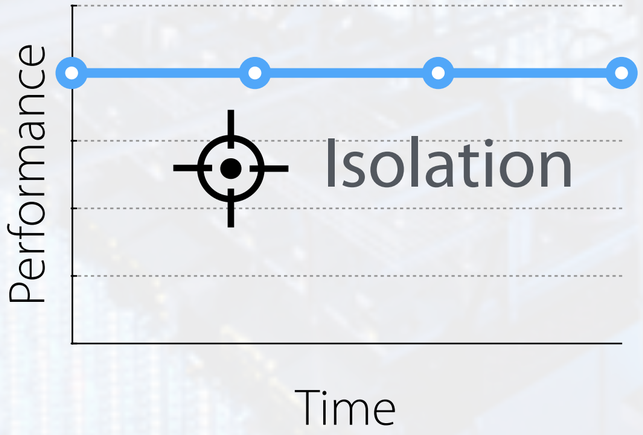
Provision for isolation or sharing?

Strong Isolation vs. Efficiency

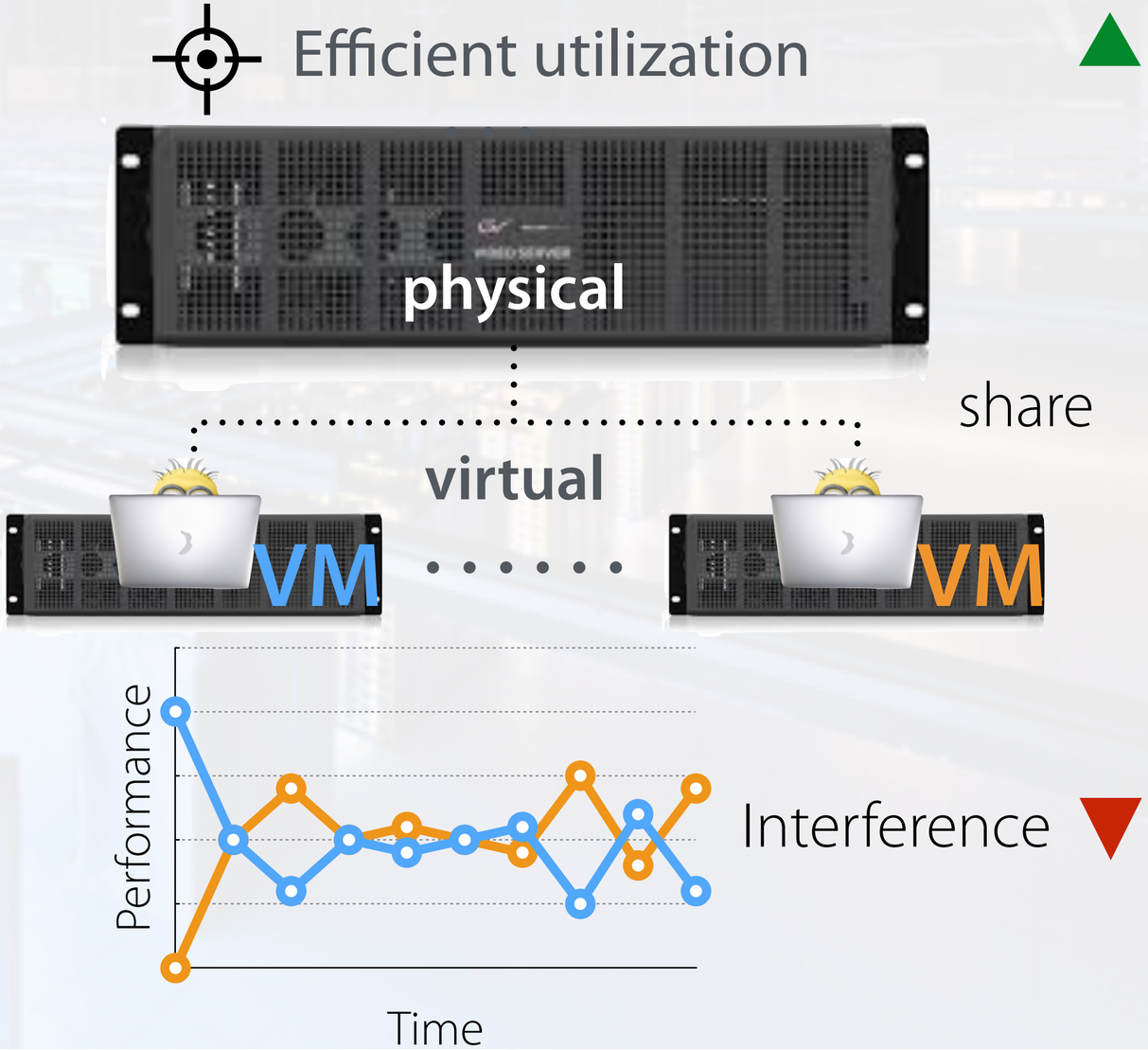
Tenant



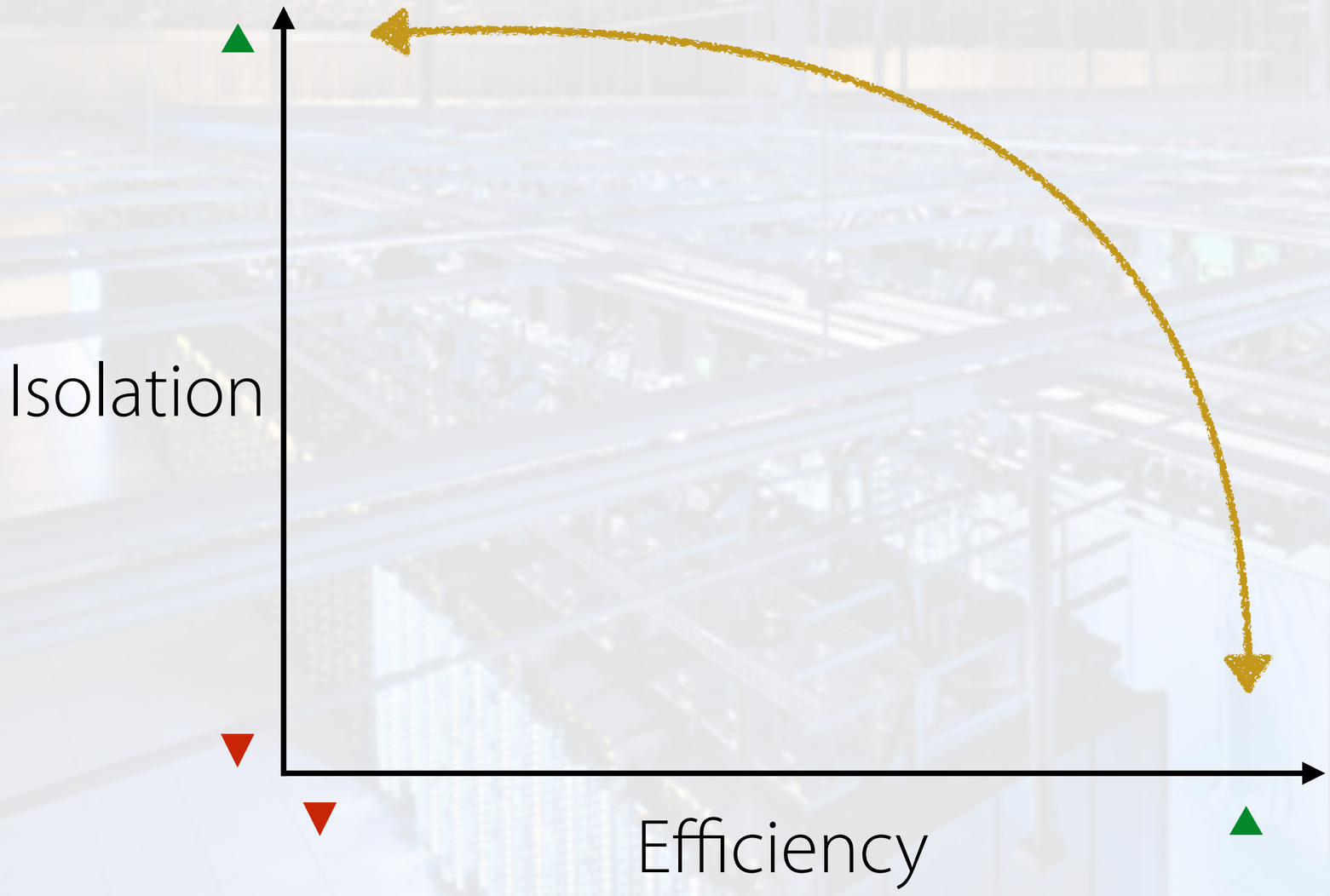
Performance



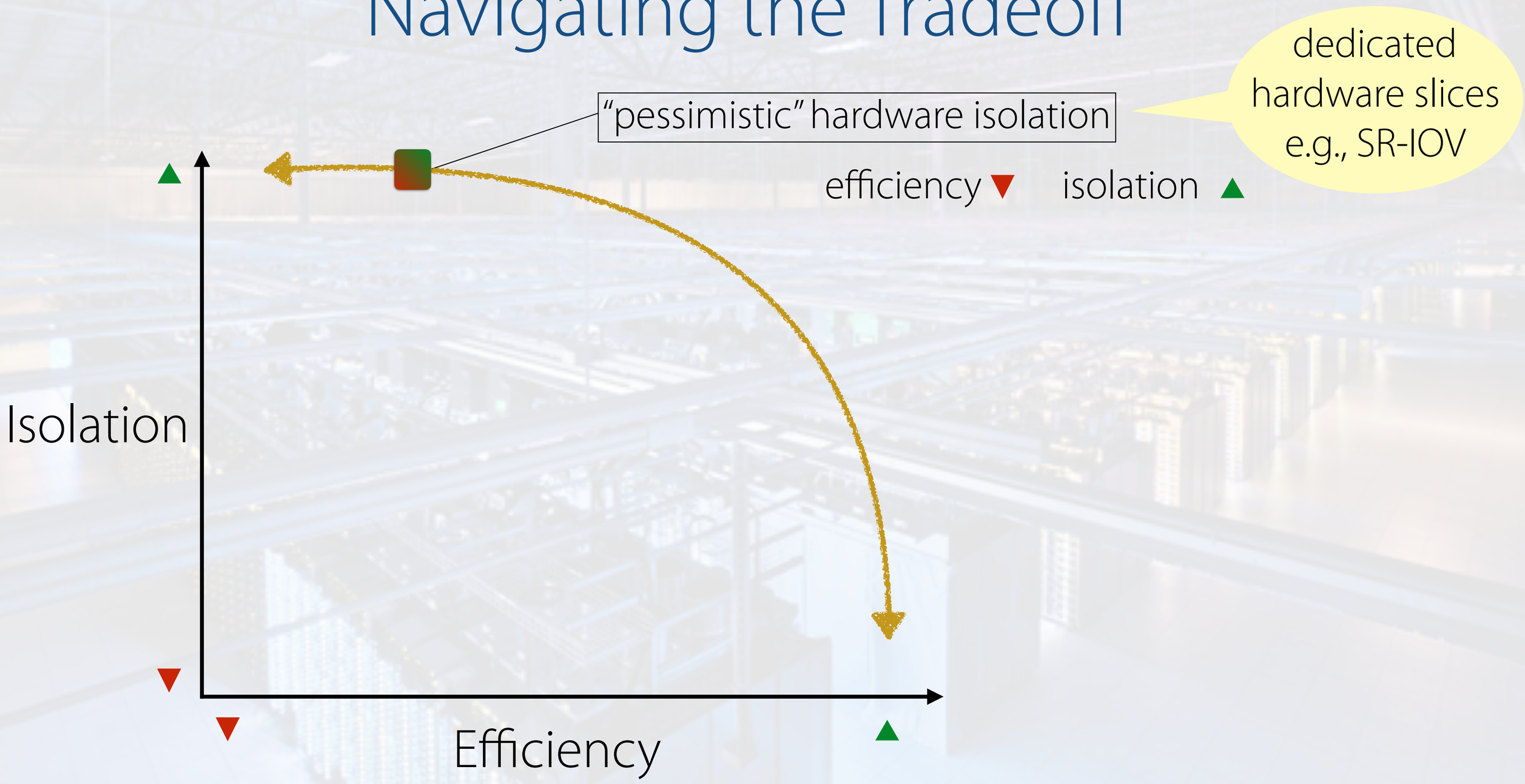
Provider



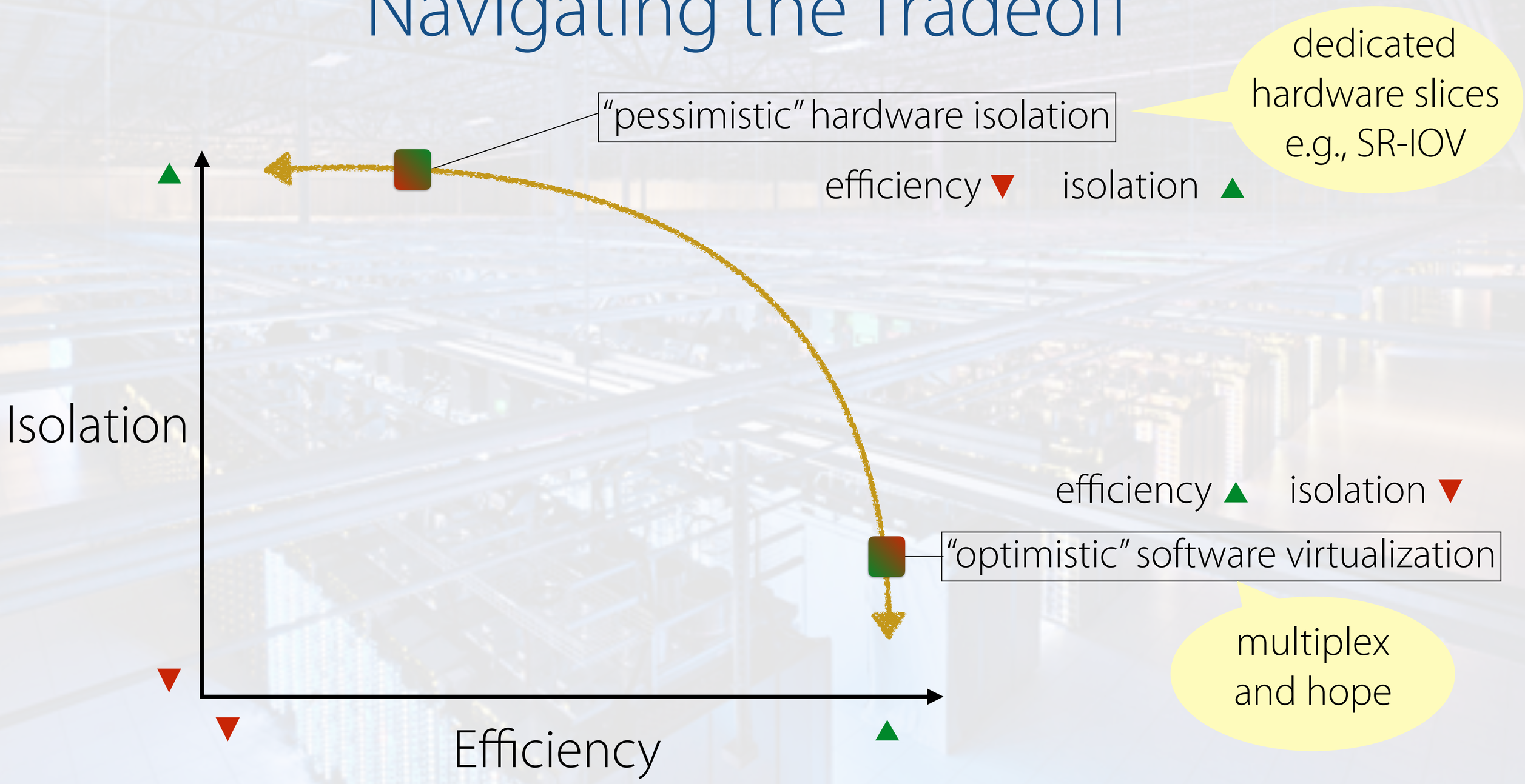
Navigating the Tradeoff



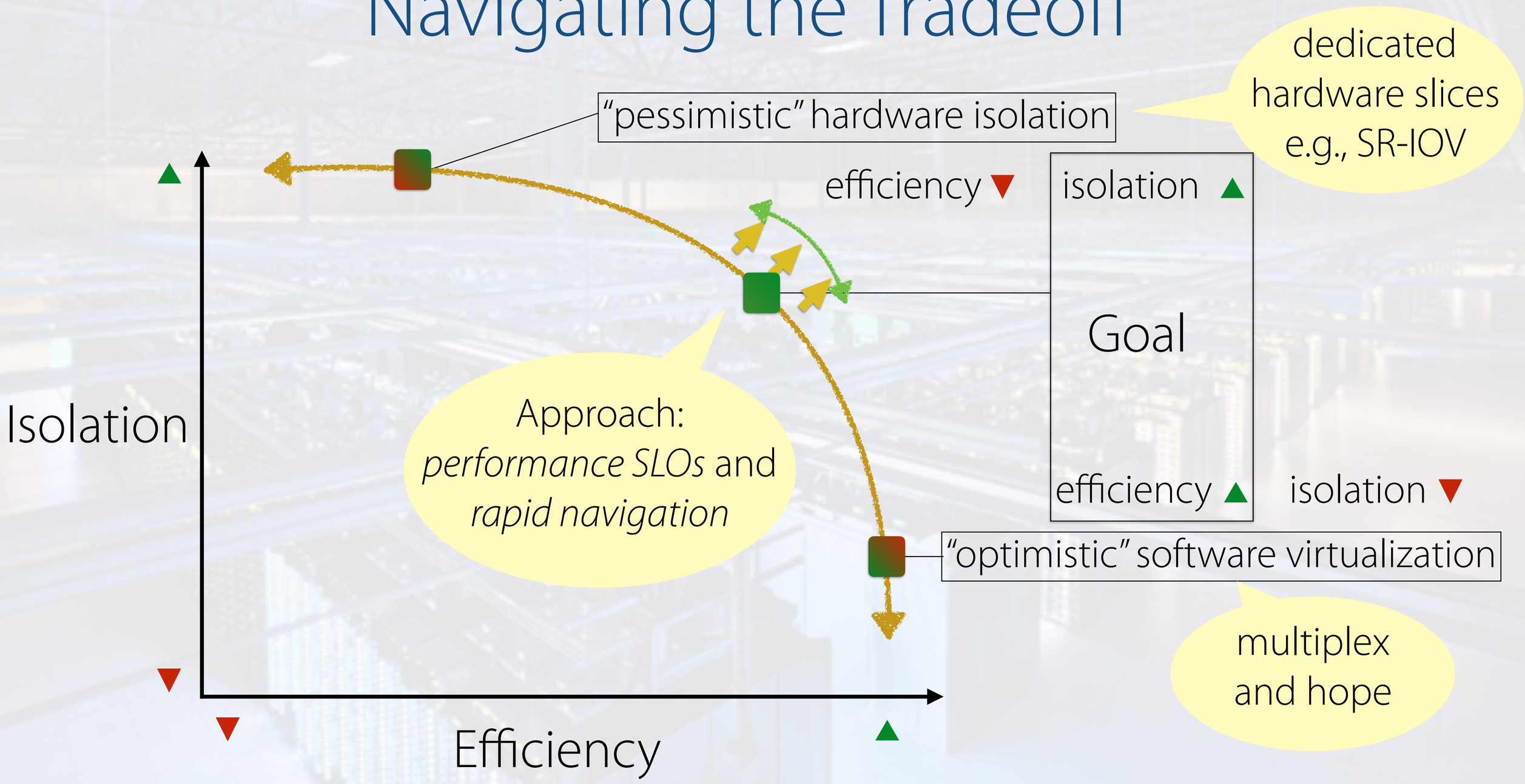
Navigating the Tradeoff



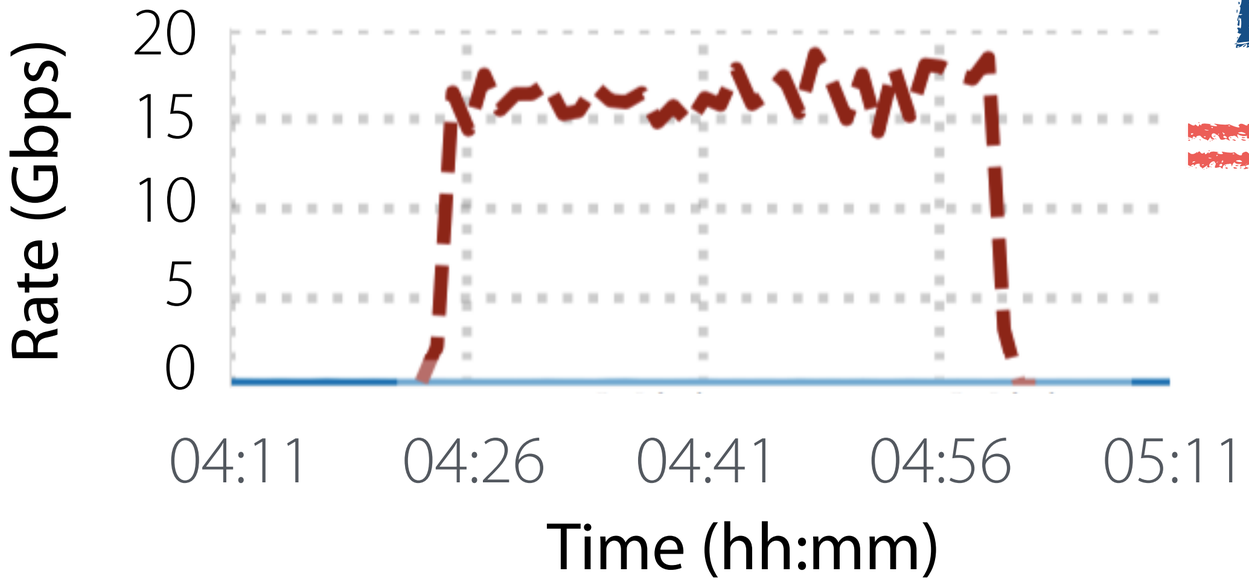
Navigating the Tradeoff

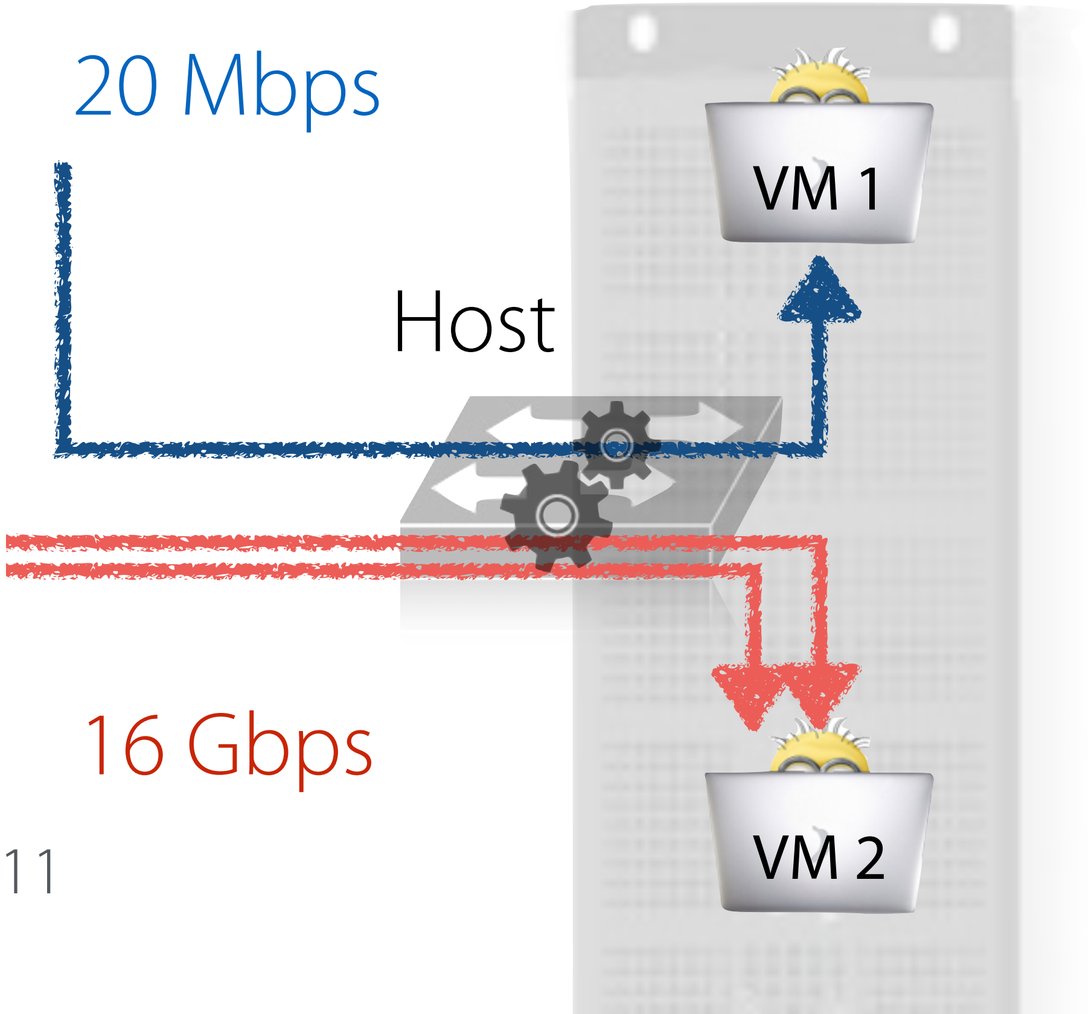


Navigating the Tradeoff

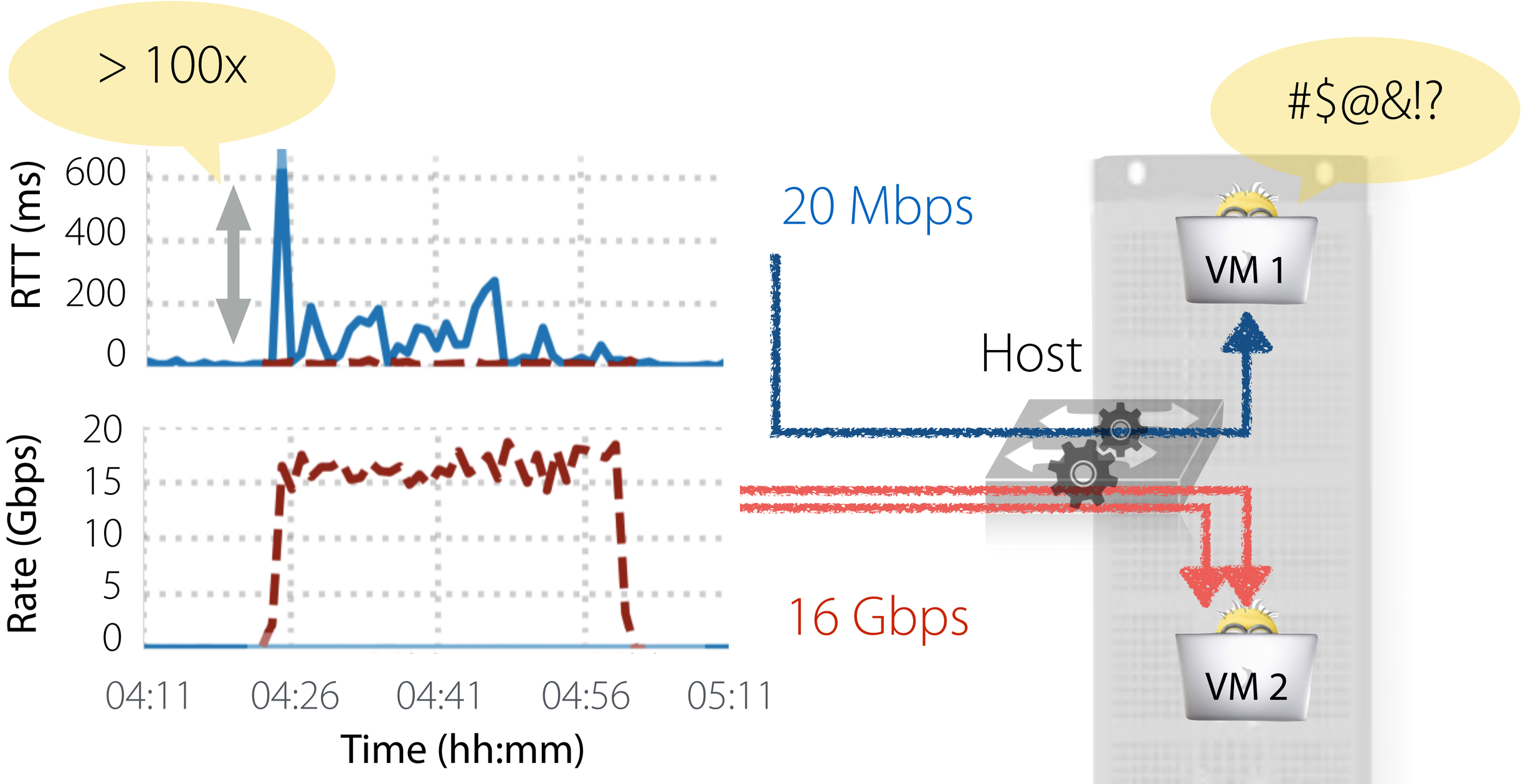


Example: Unpredictable Performance

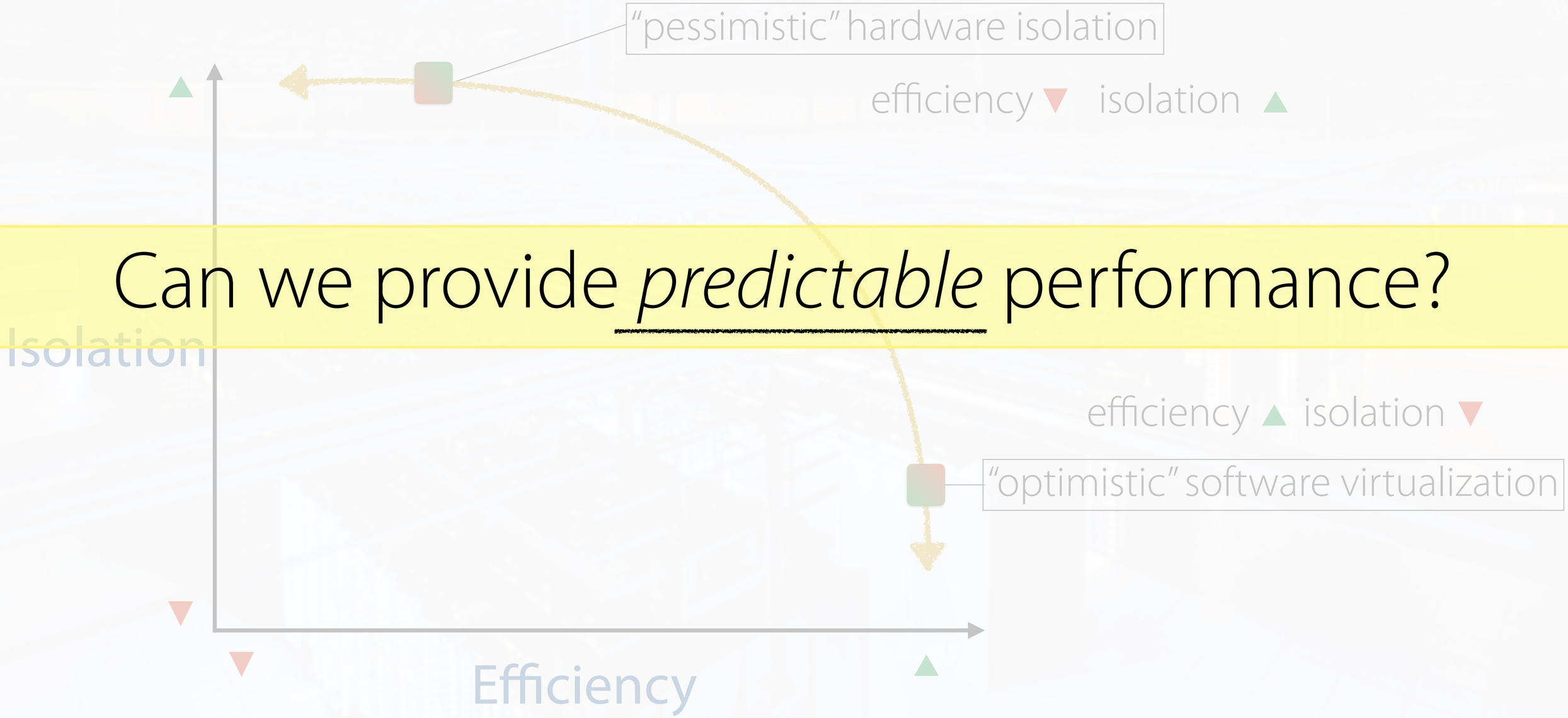




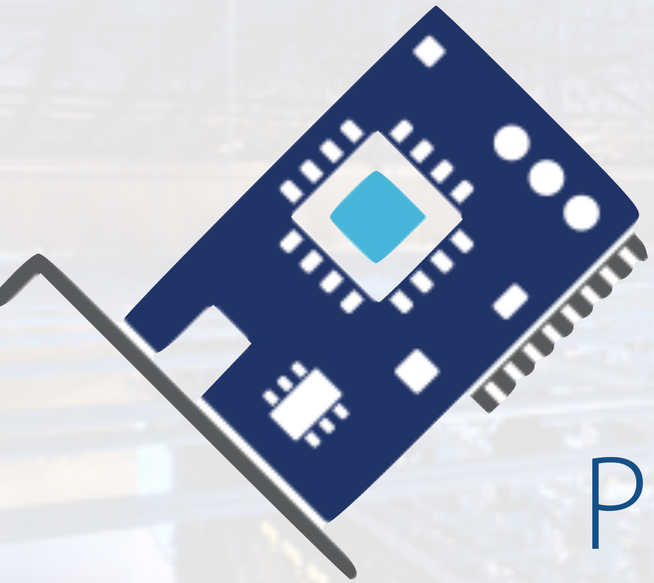
Example: Unpredictable Performance



Navigating the tradeoff

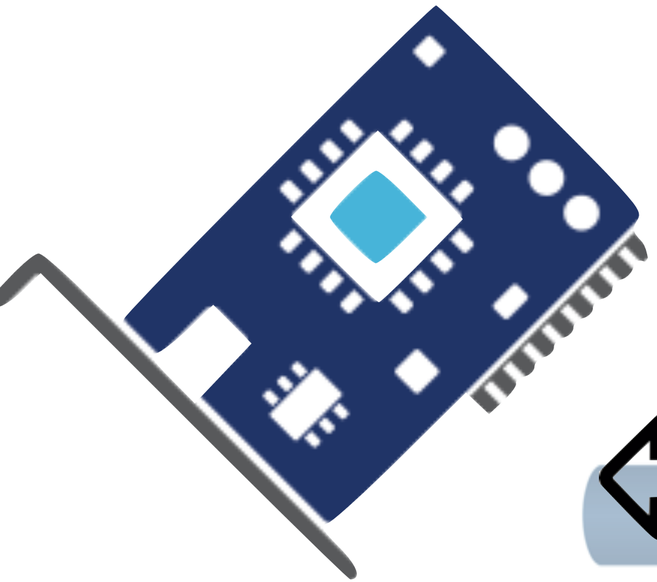


Can we provide *predictable* performance?



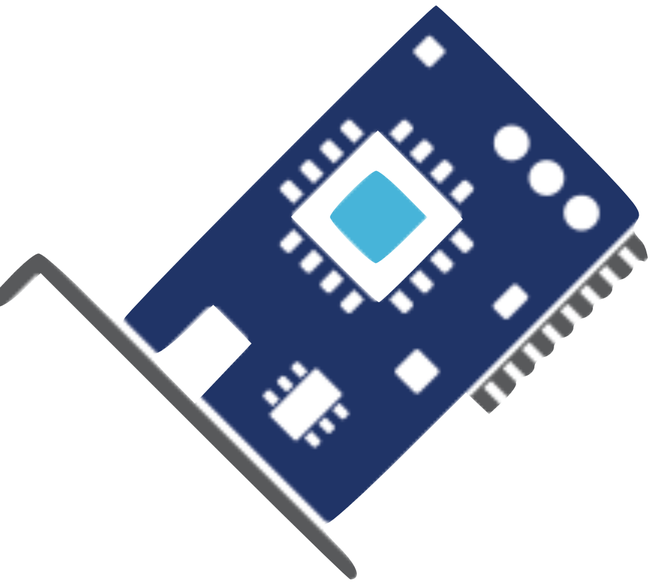
Predictable Virtualized NIC

Predictable Virtualized NIC



Metric	Predictable vNIC SLO
Bandwidth	*min*. and *max*. envelope (hose model)
Delay	Low, predictable*
Loss	No drops for cooperating traffic*

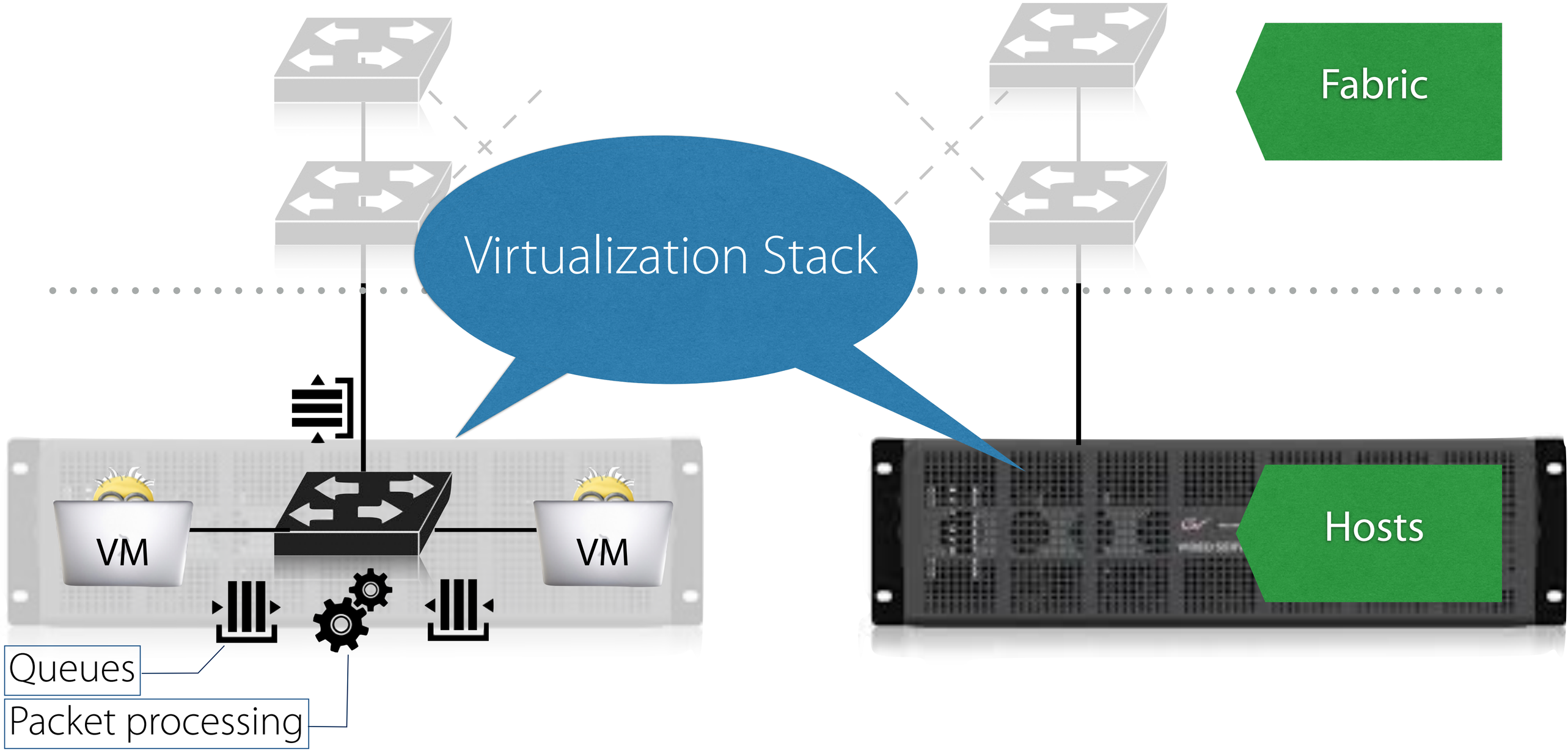
* for traffic within its bandwidth envelope



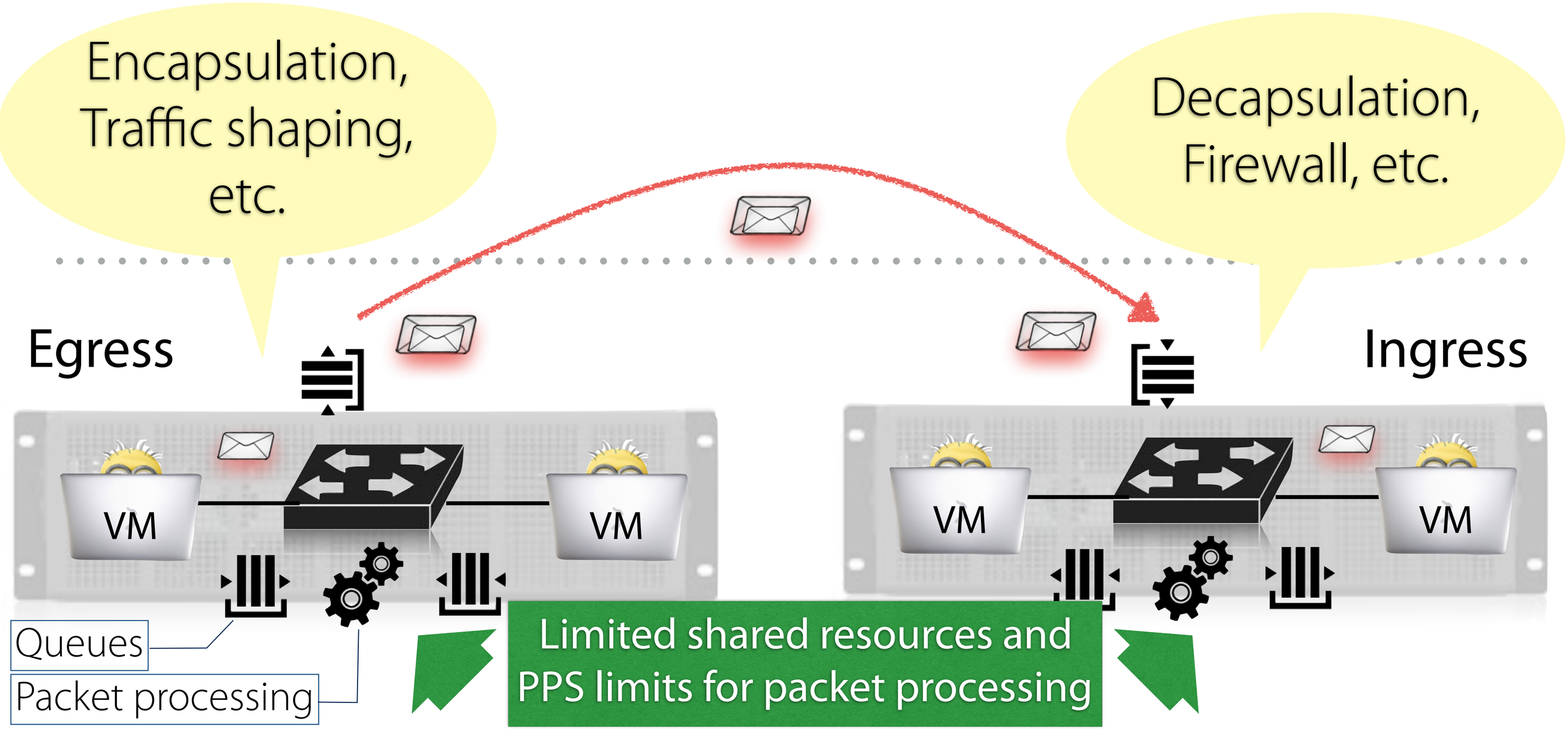
Key Challenge

Isolation breakdown in virtualization stacks at the end-hosts

Network Virtualization

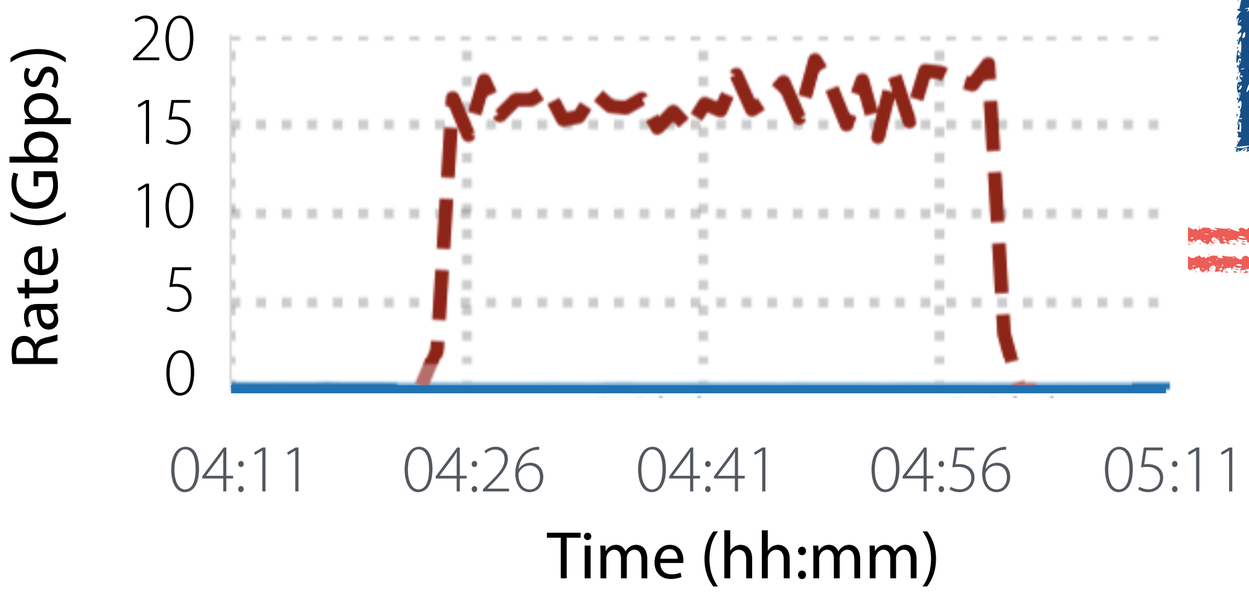


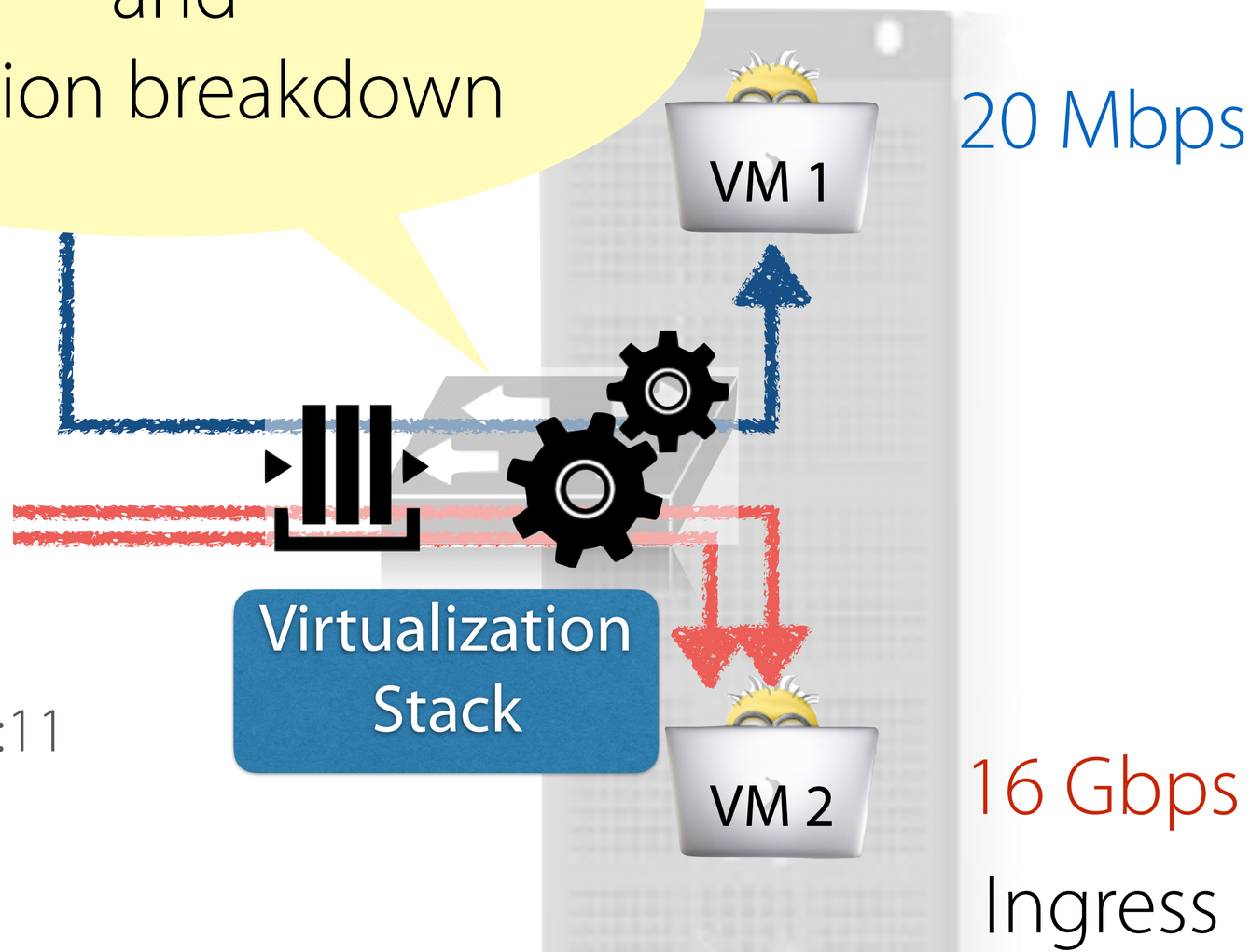
Virtualization Stack



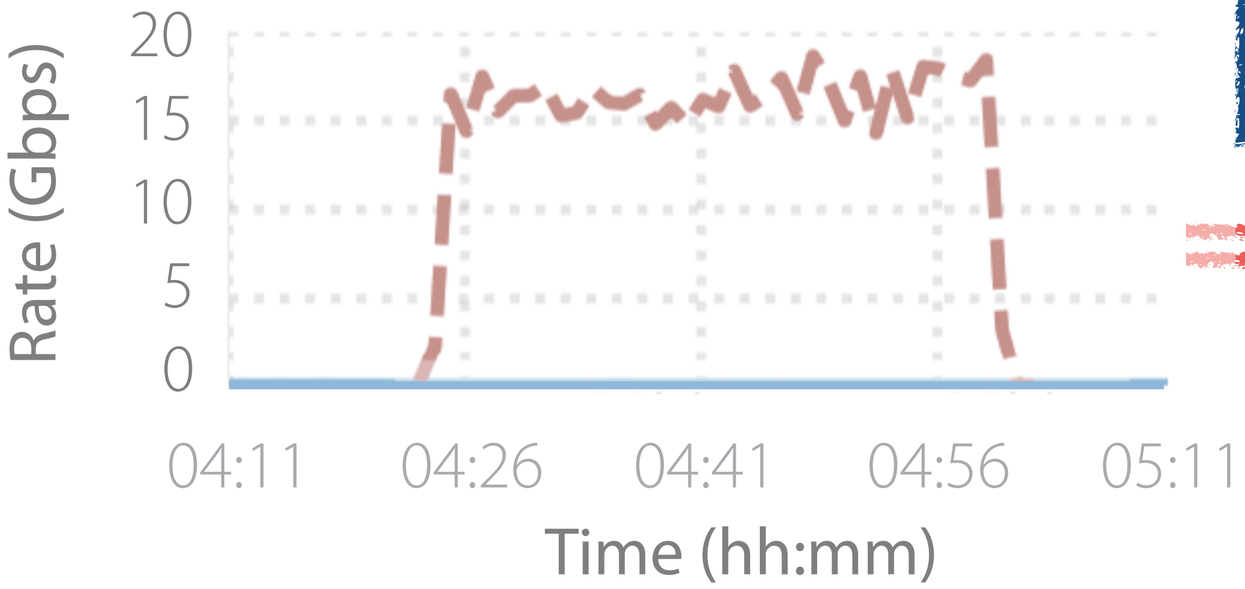
Example: Isolation Breakdown

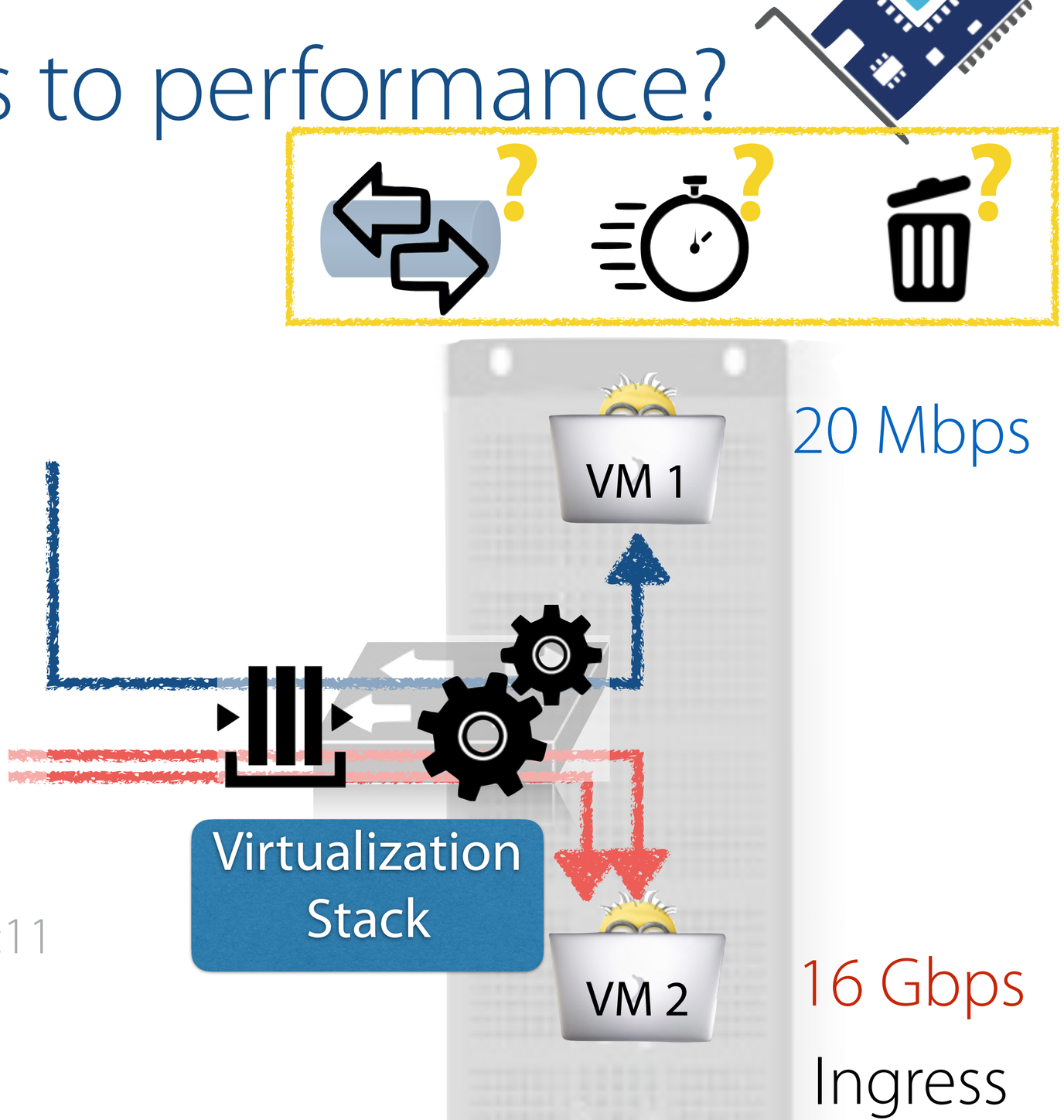
Resource contention
and
isolation breakdown



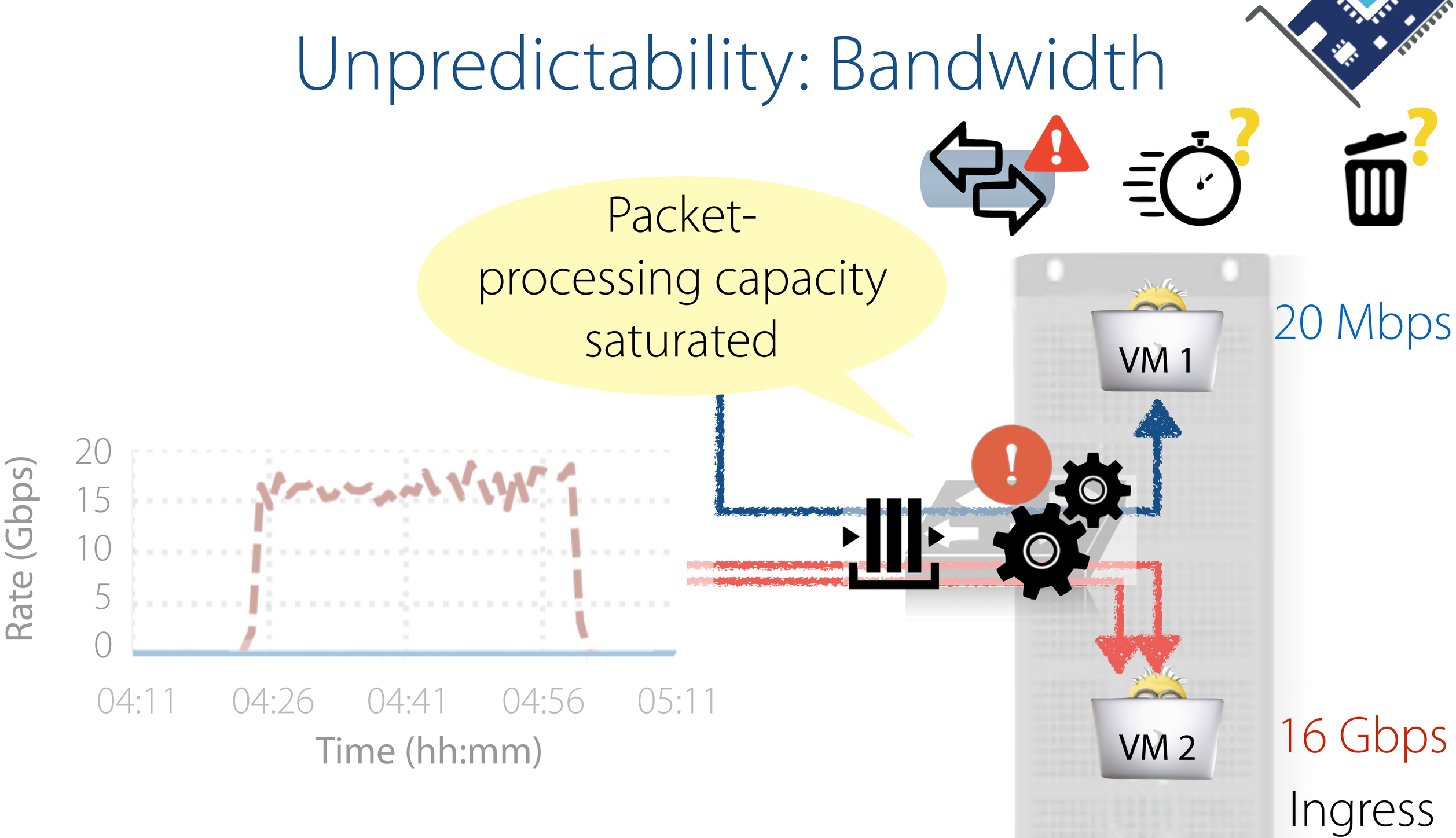


What happens to performance?

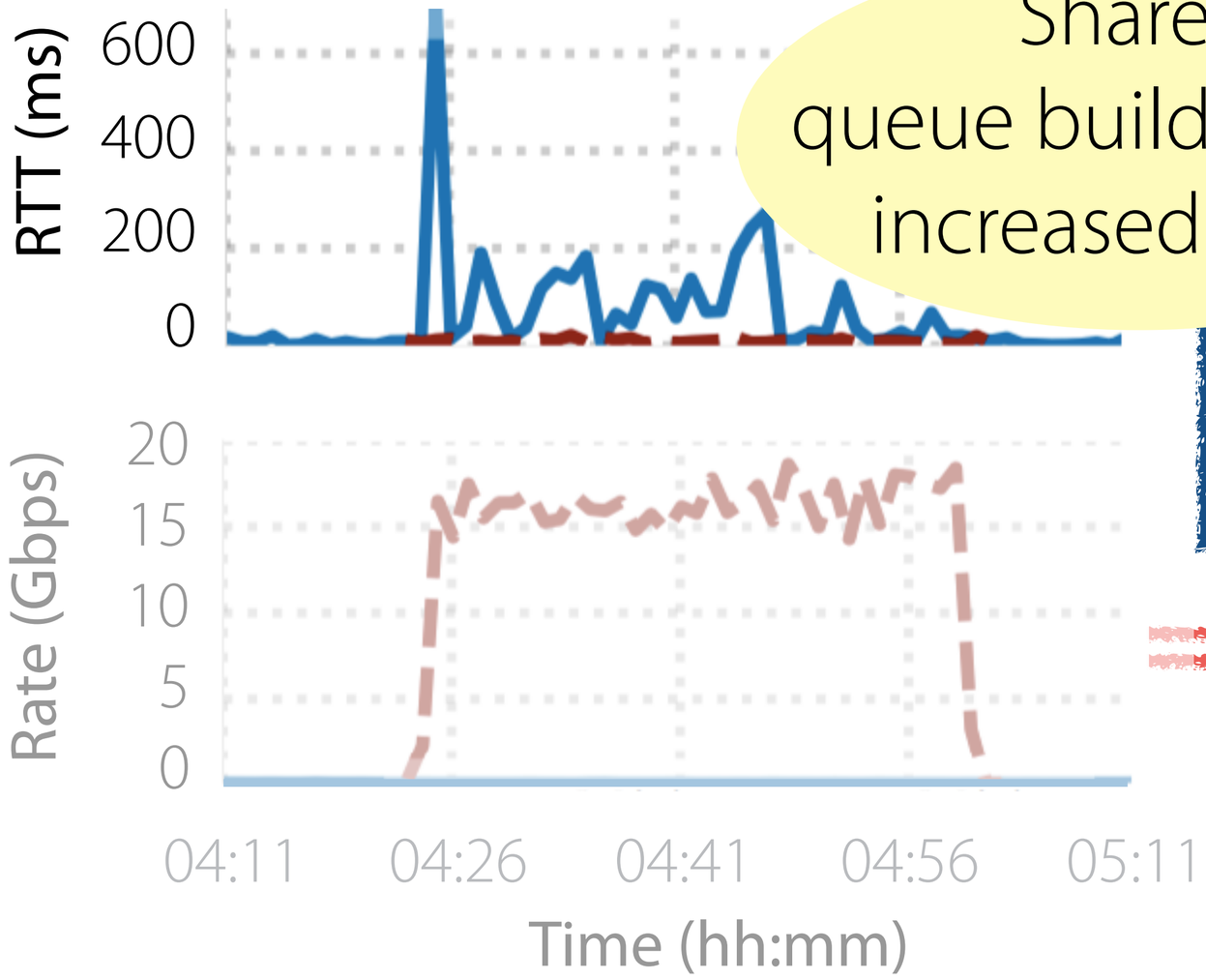




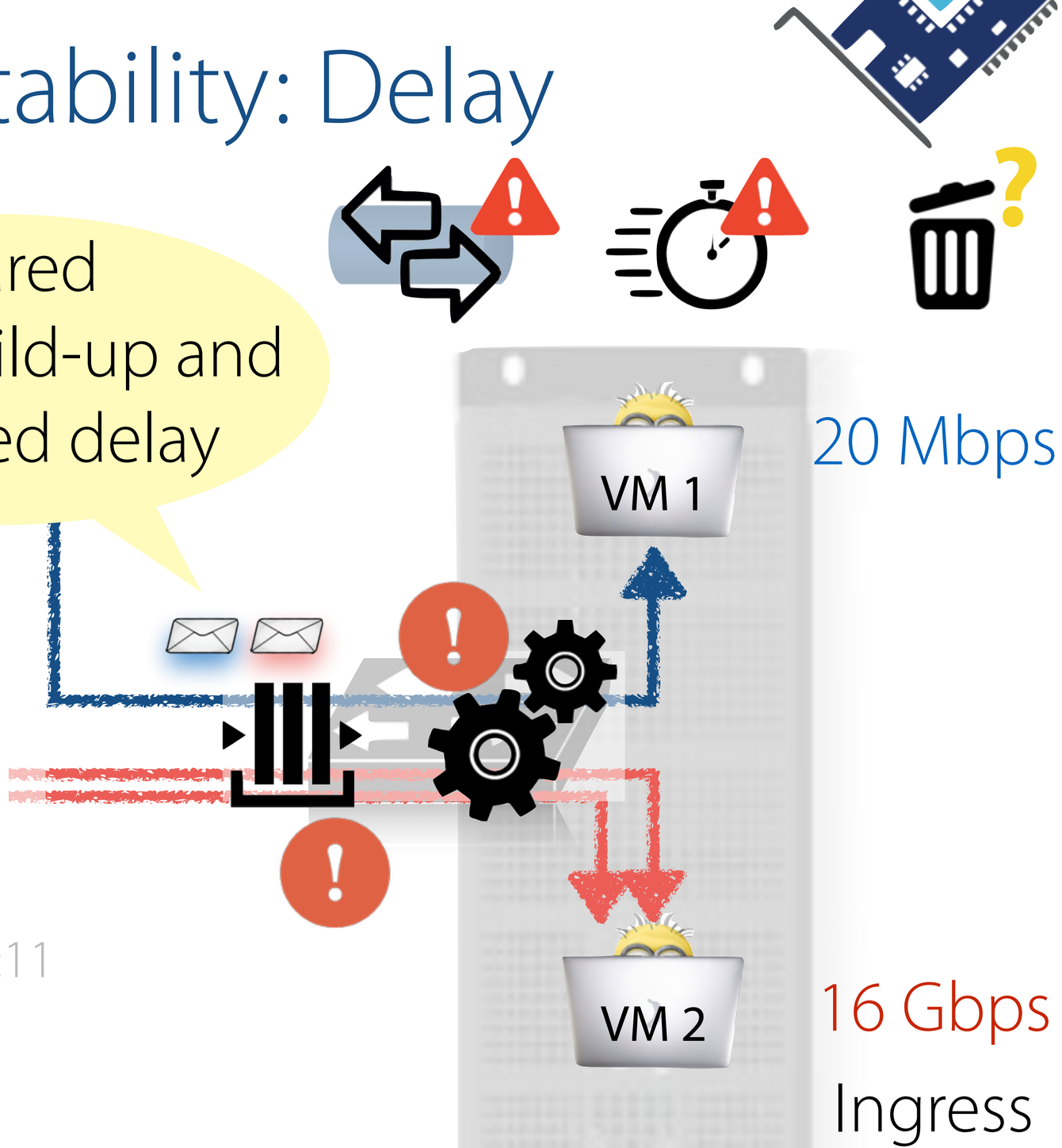
Unpredictability: Bandwidth



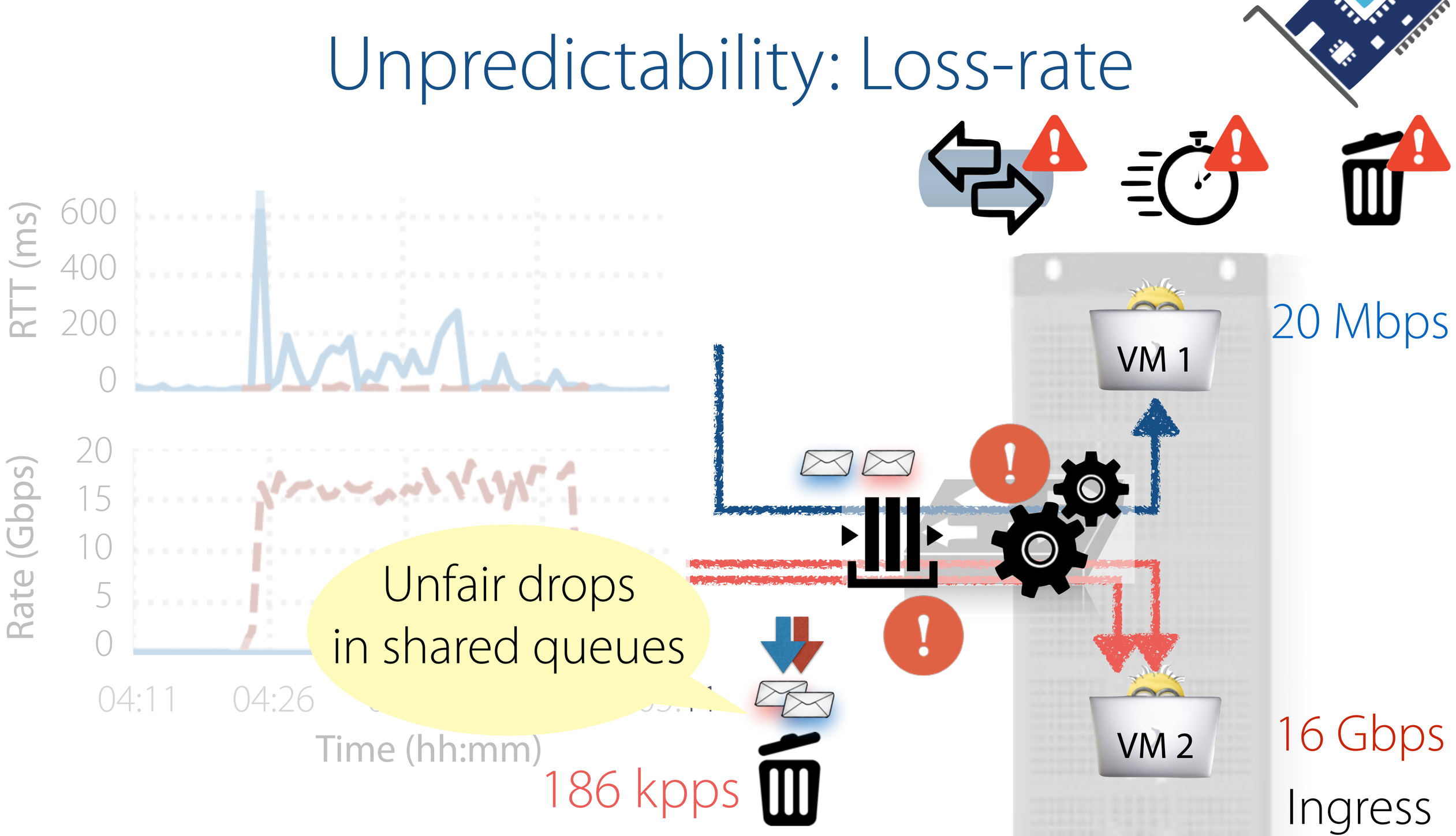
Unpredictability: Delay



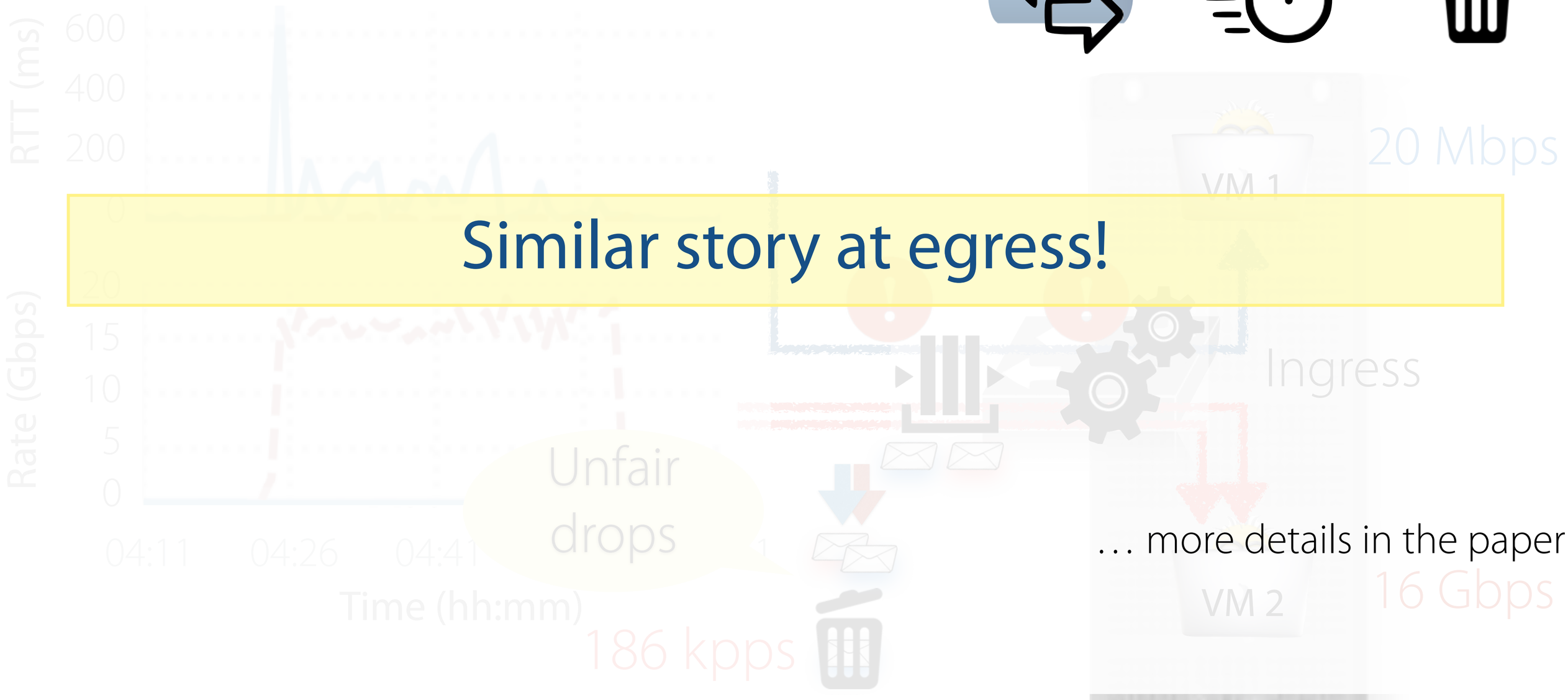
Shared queue build-up and increased delay



Unpredictability: Loss-rate



Example: Isolation Breakdown

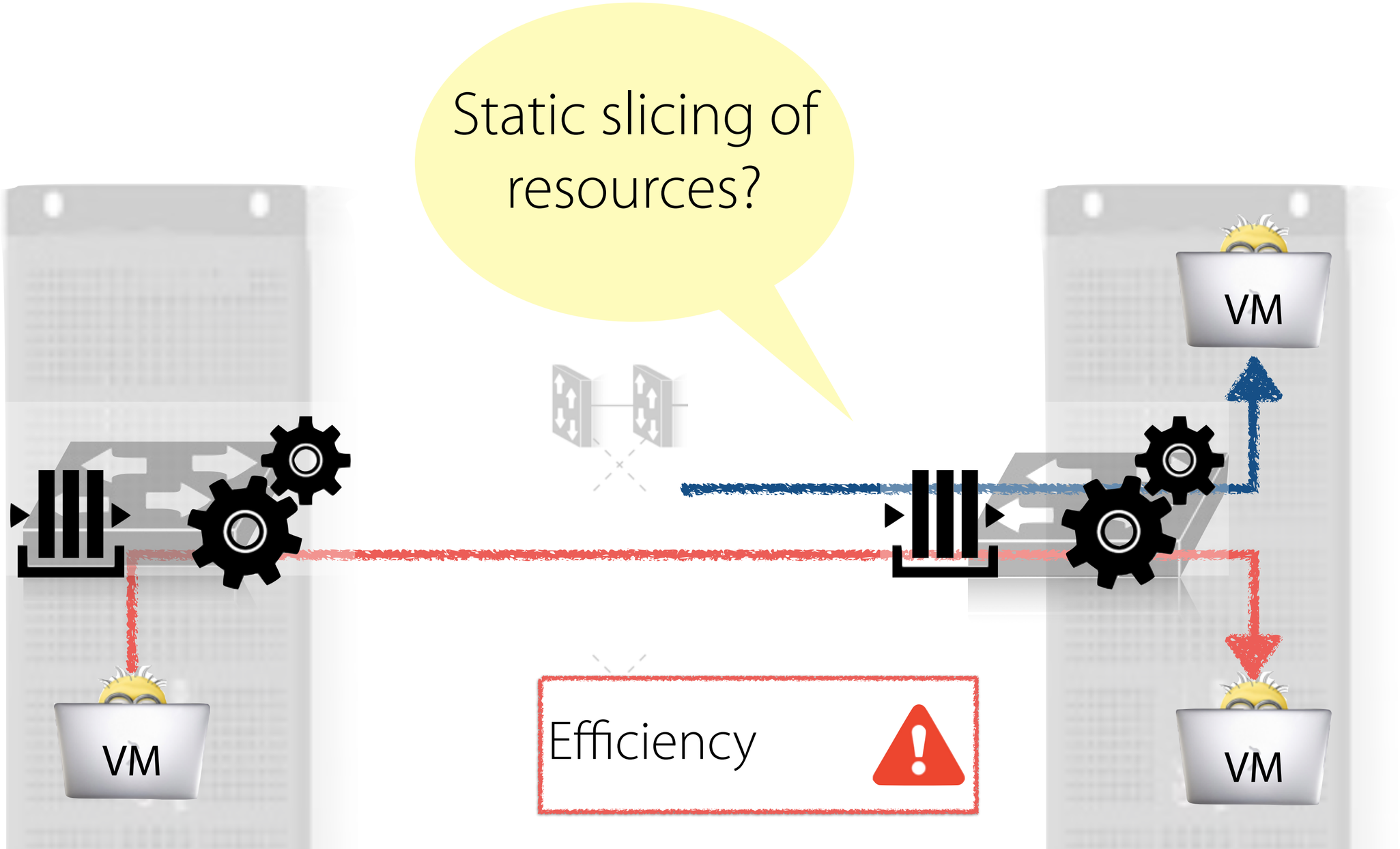


Similar story at egress!

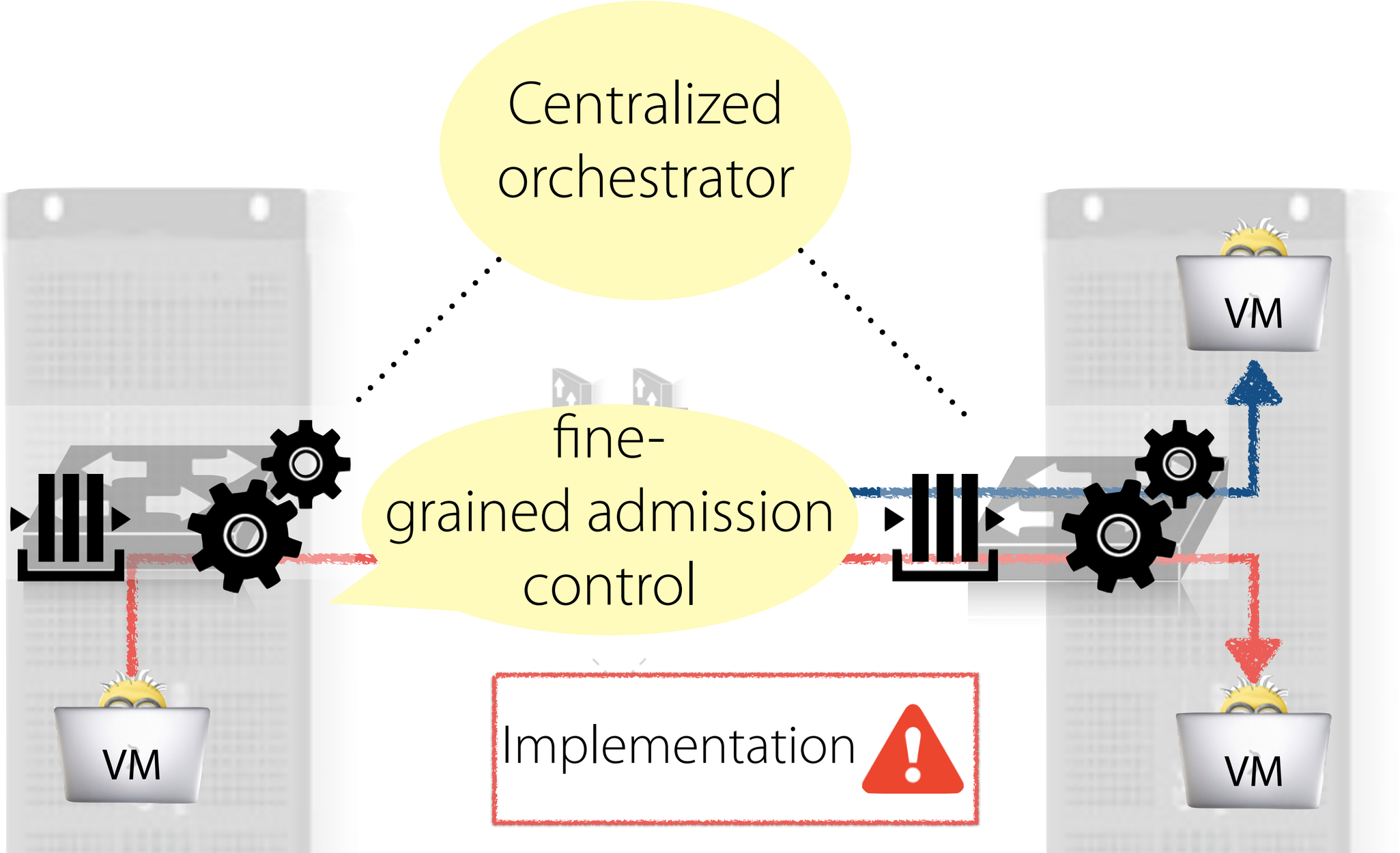
… more details in the paper

Solution Space

Solution Space



Solution Space



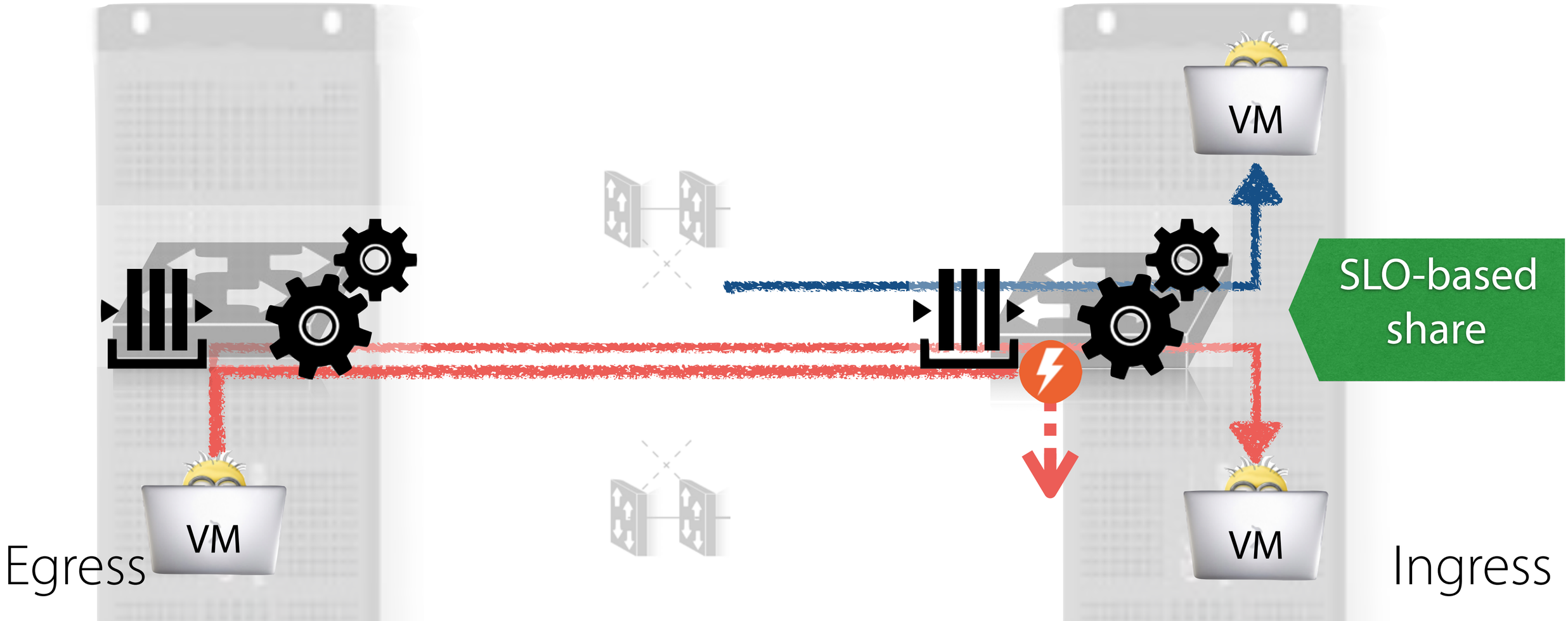
Approach

Insight

Share resources based on SLOs

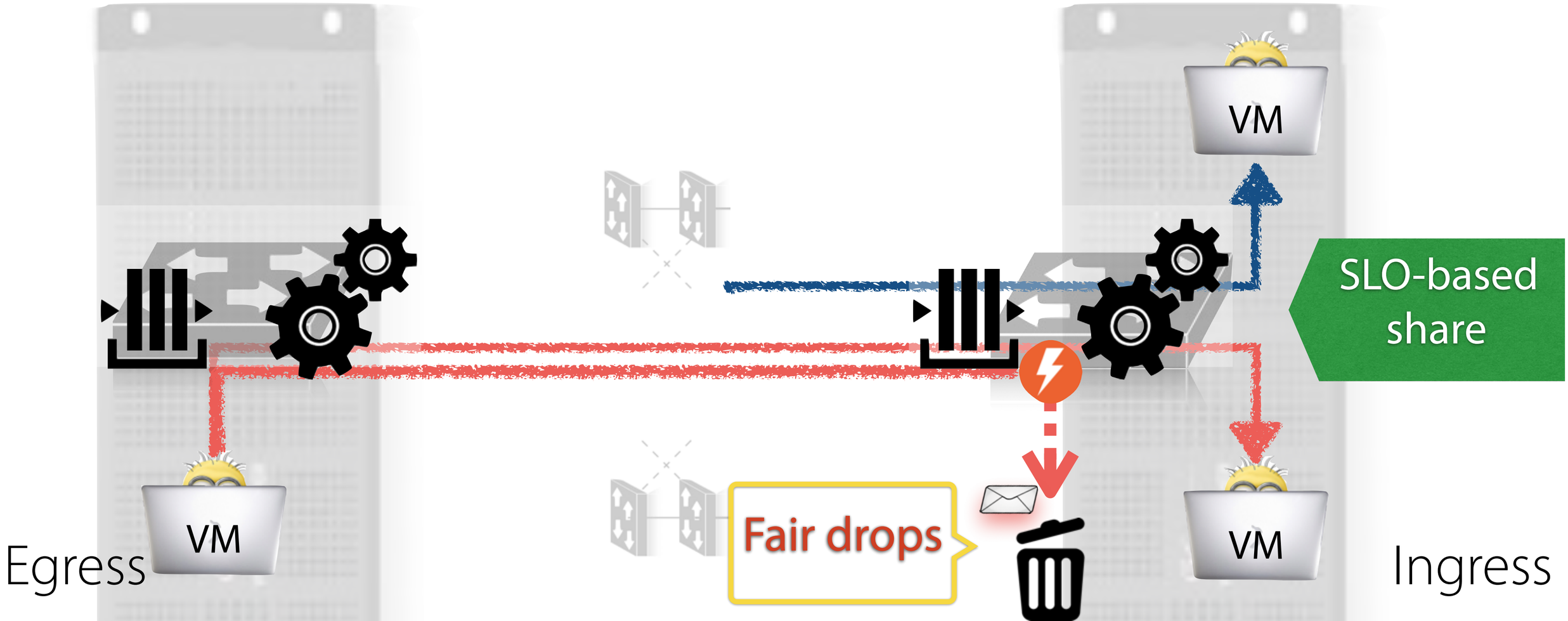
Approach

Share resources based on SLOs



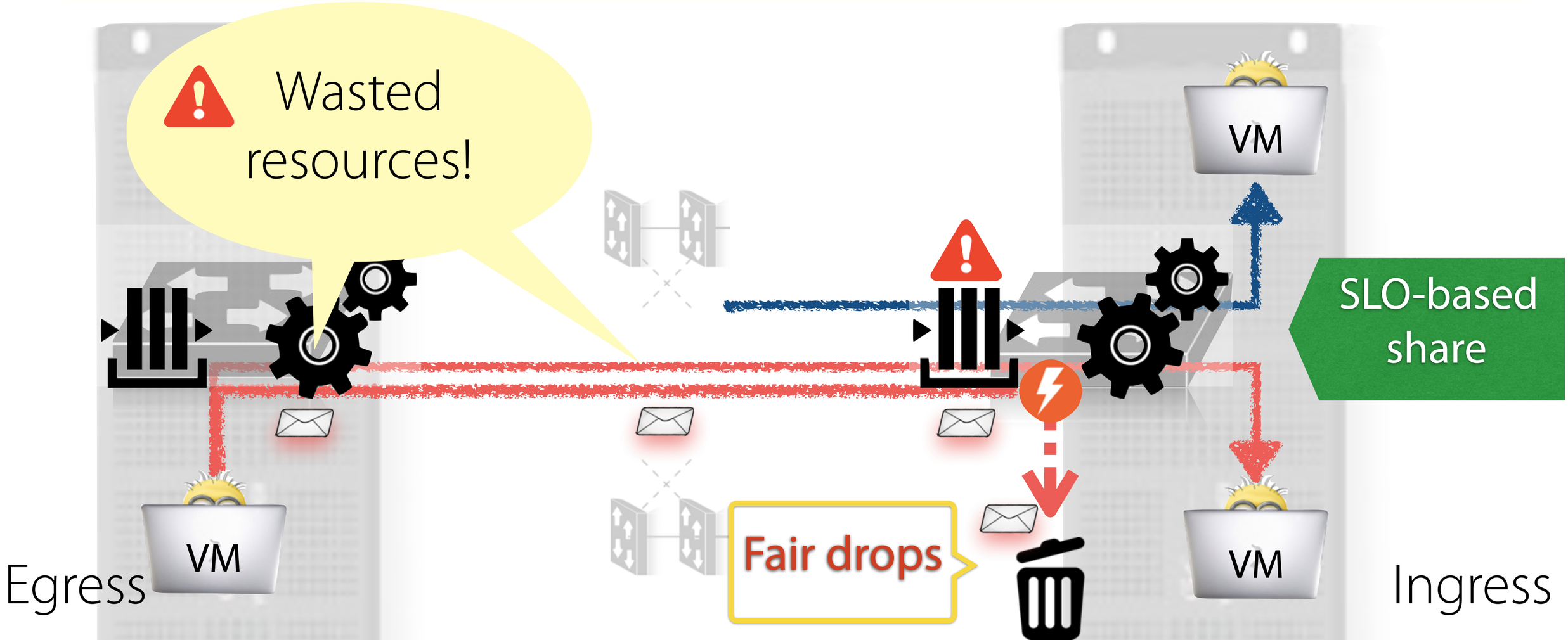
Approach

Share resources based on SLOs



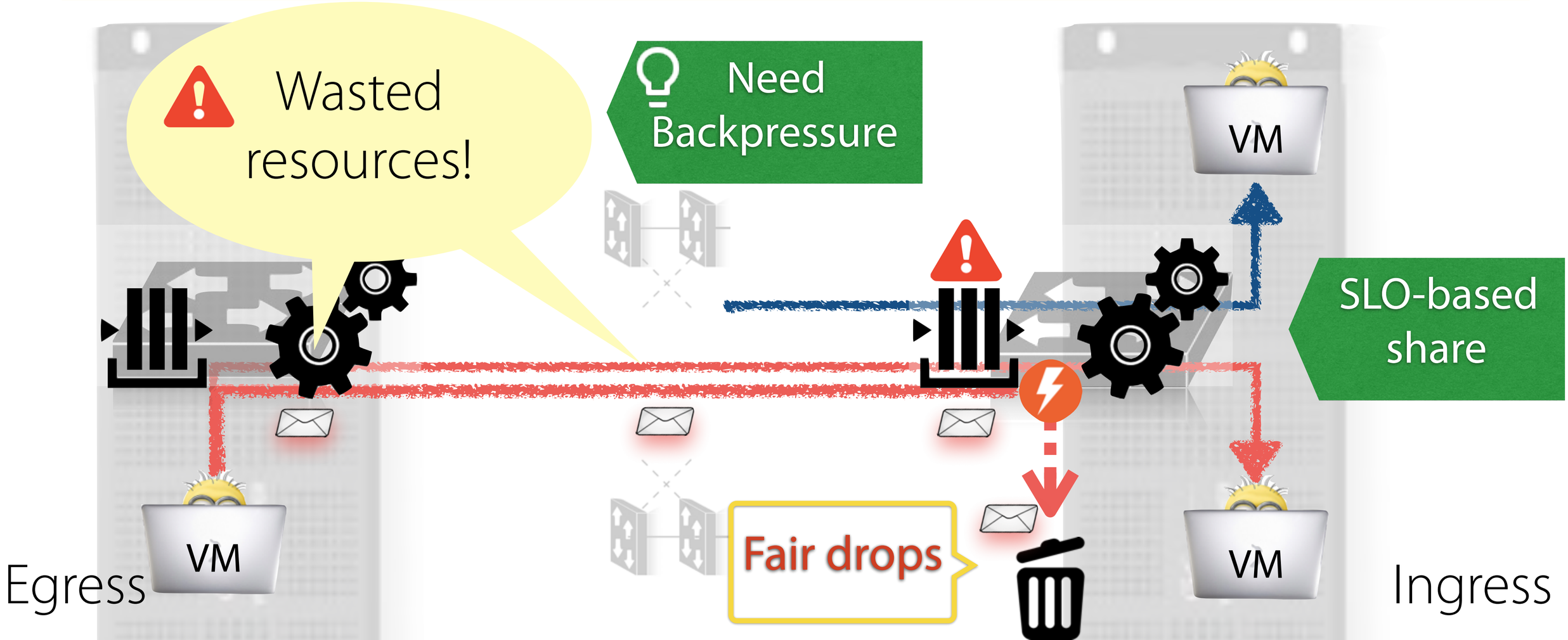
Approach

Share resources based on SLOs



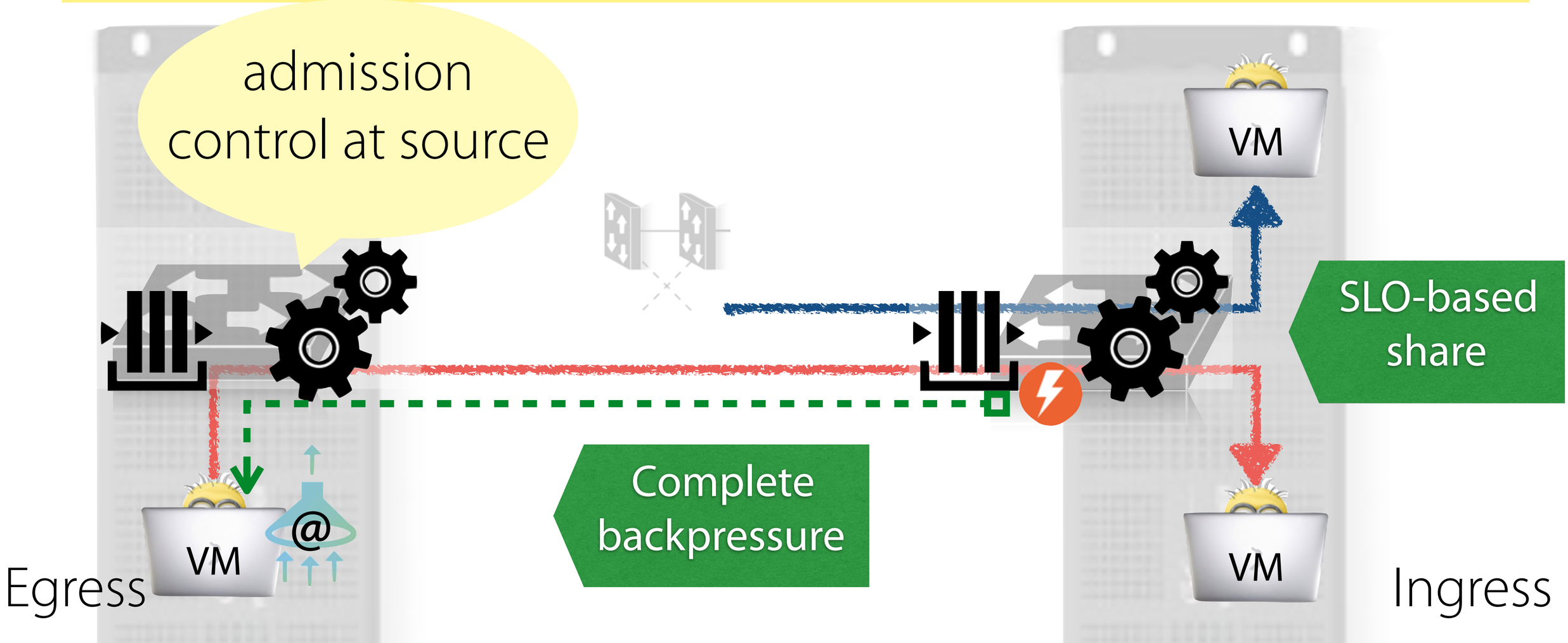
Approach

Share resources based on SLOs



Approach

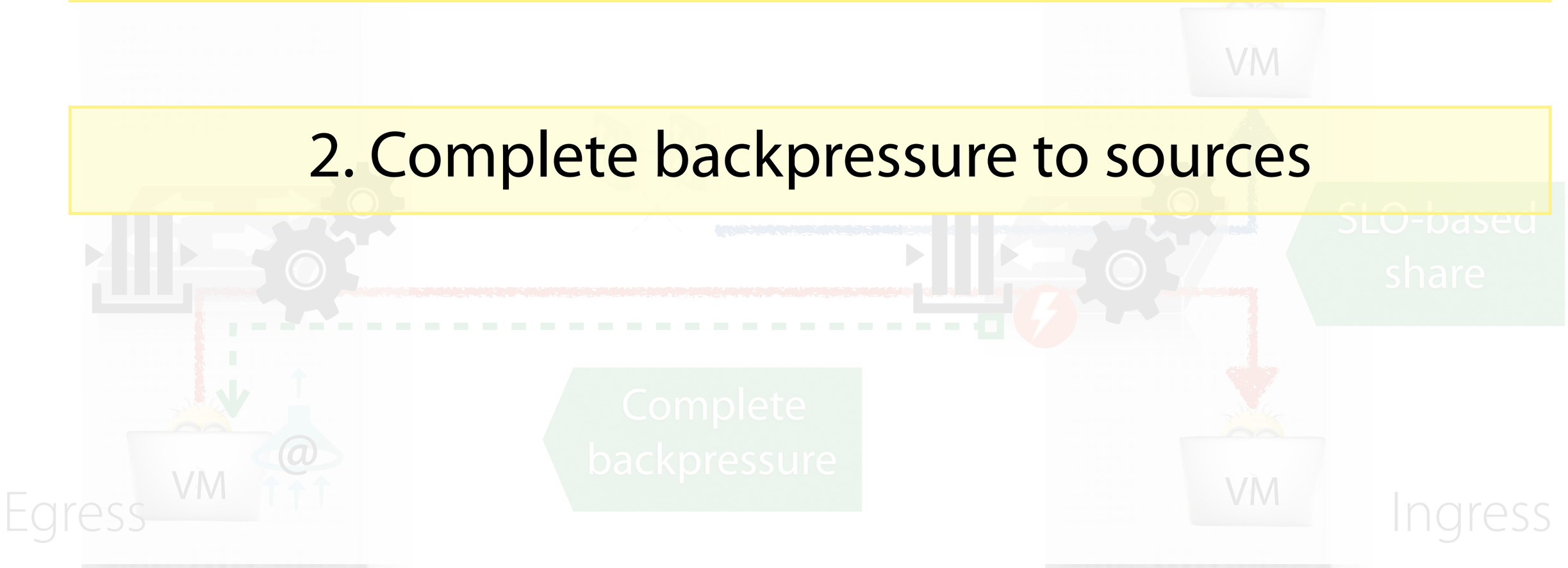
Share resources based on SLOs



Design Principles

1. Share resources based on SLOs

2. Complete backpressure to sources



PicNIC

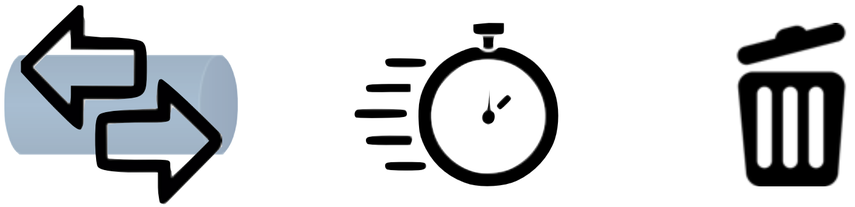
PicNIC Design

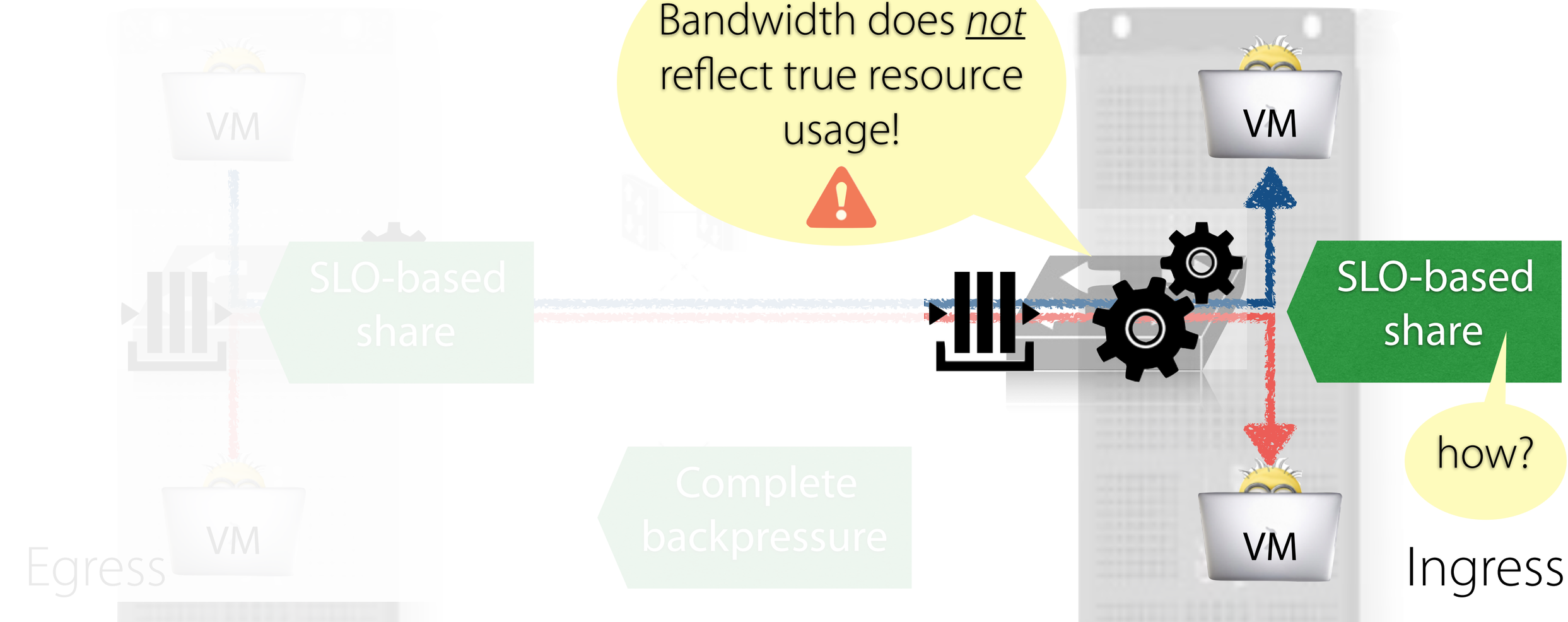
Ingress CPU-fair Queues

Congestion Control

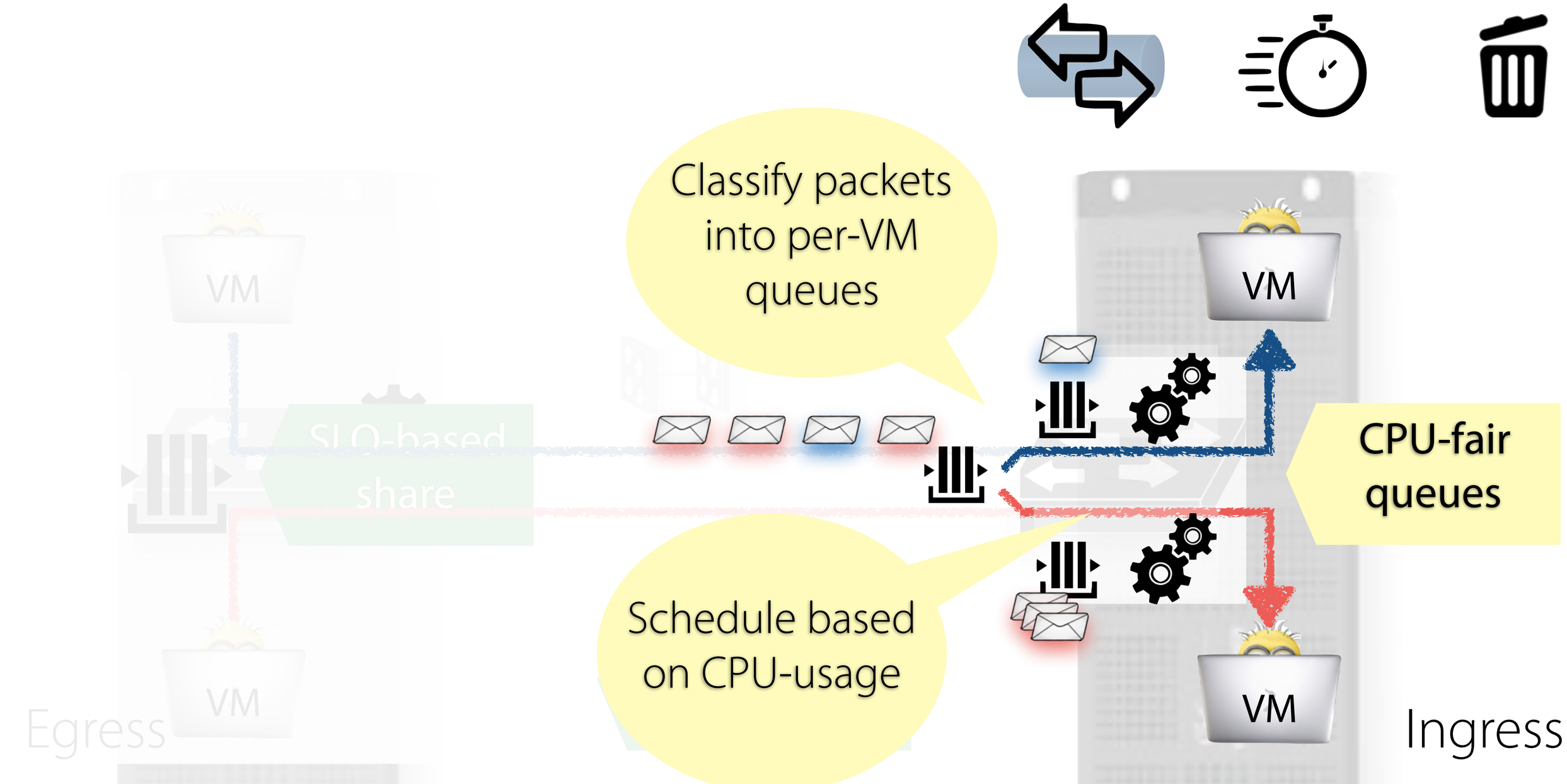
Sender-side Admission Control

Ingress CPU-fair Queues

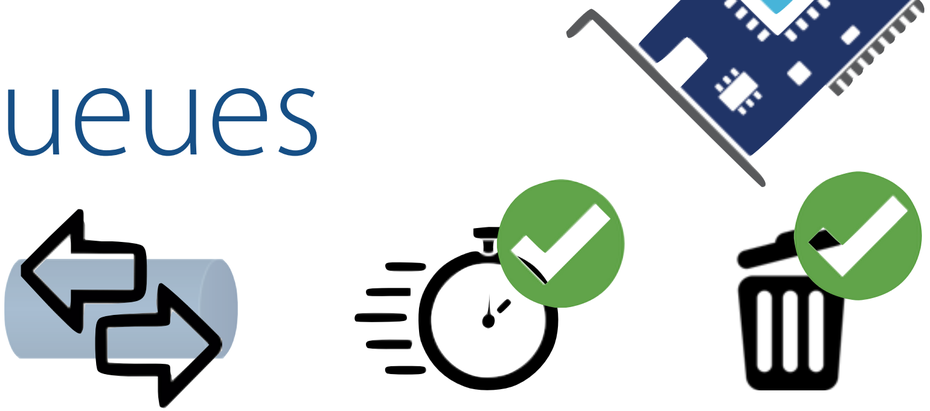


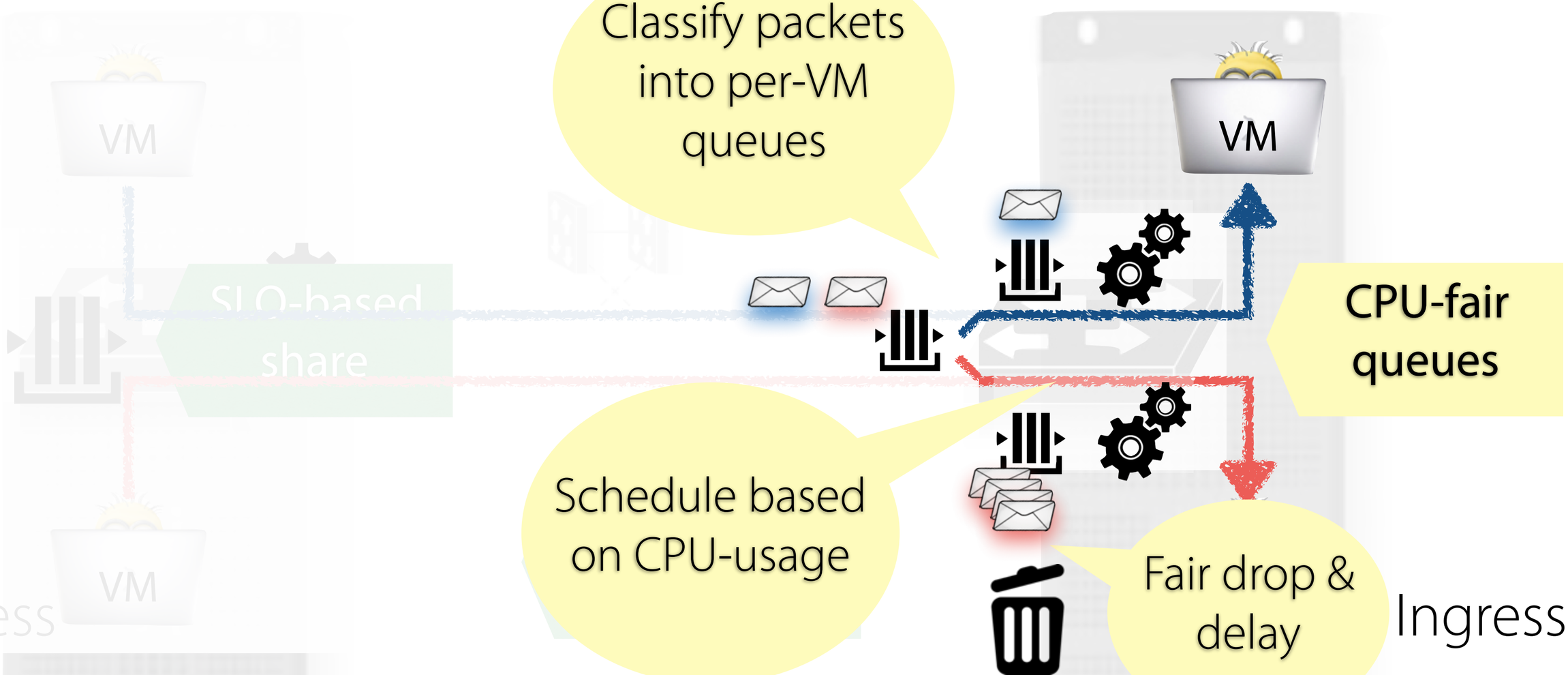


Ingress CPU-fair Queues

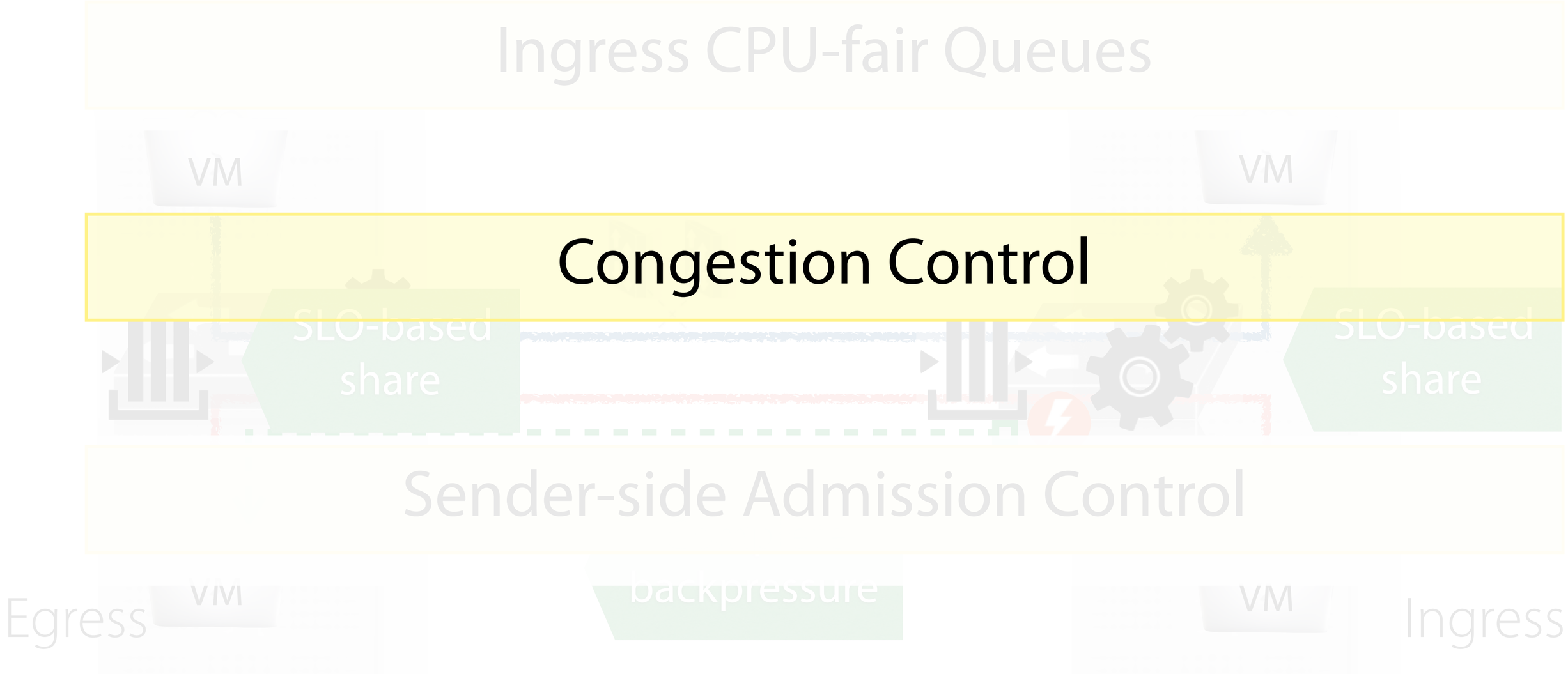


Ingress CPU-fair Queues

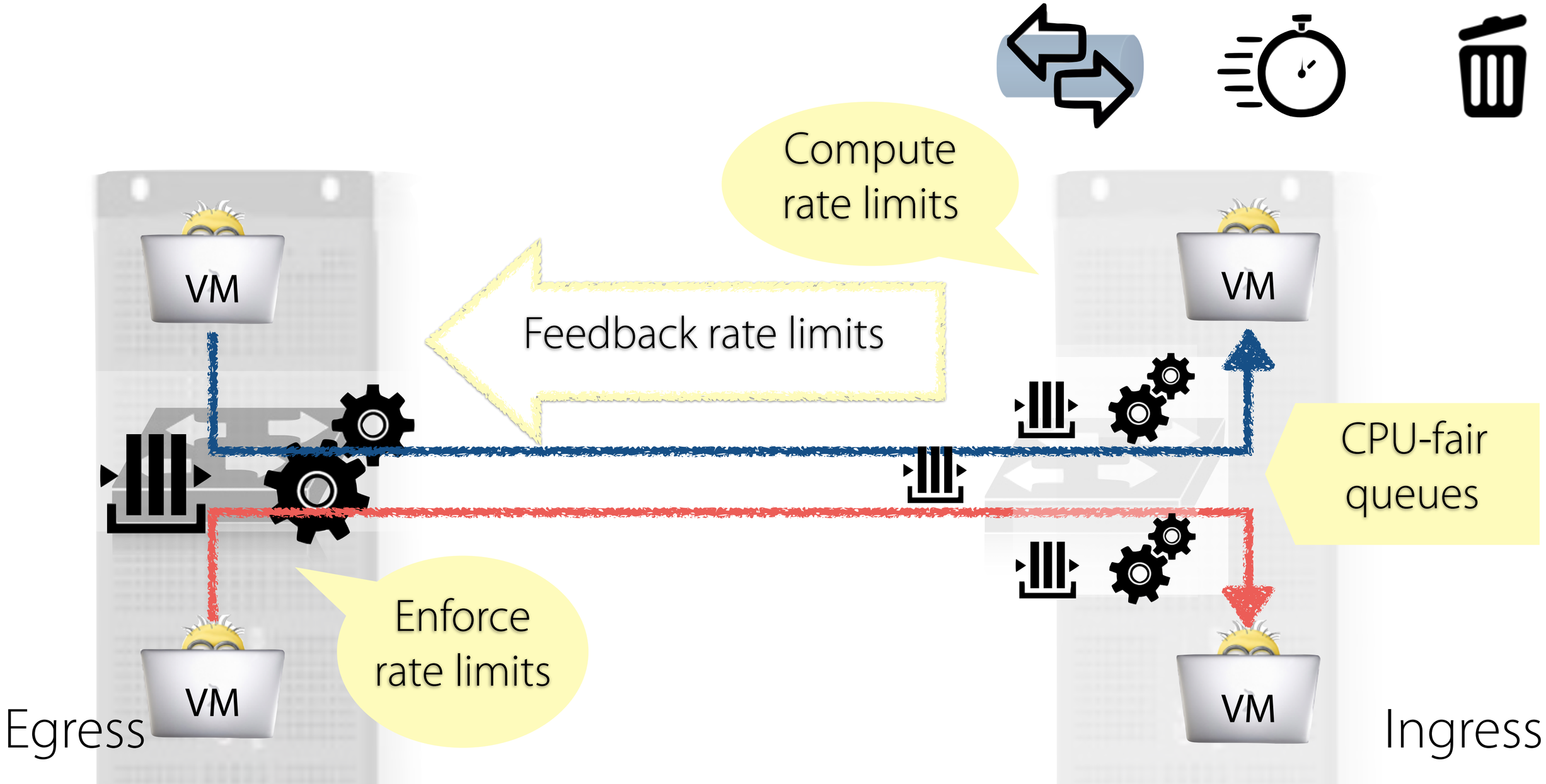




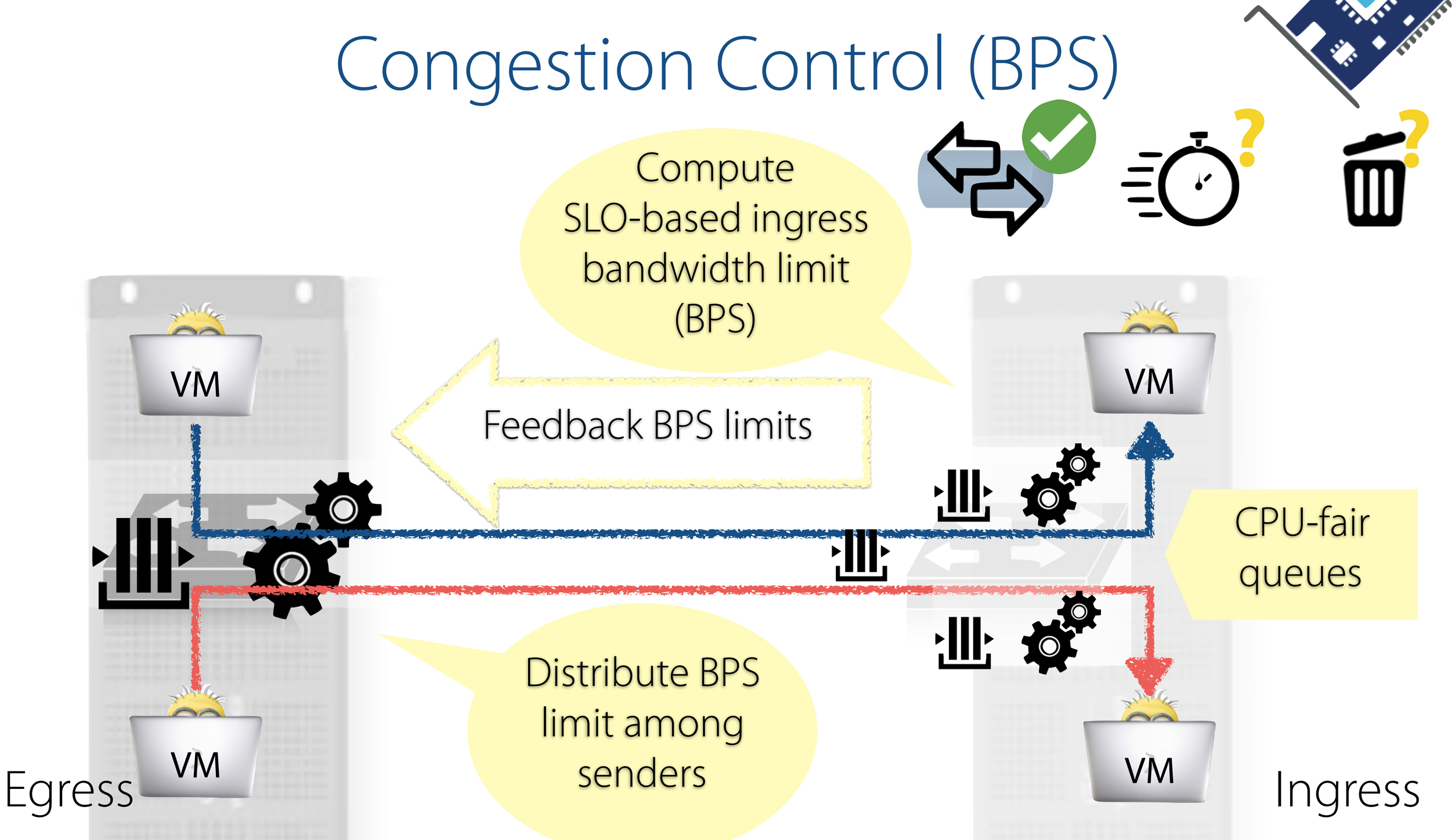
PicNIC Design



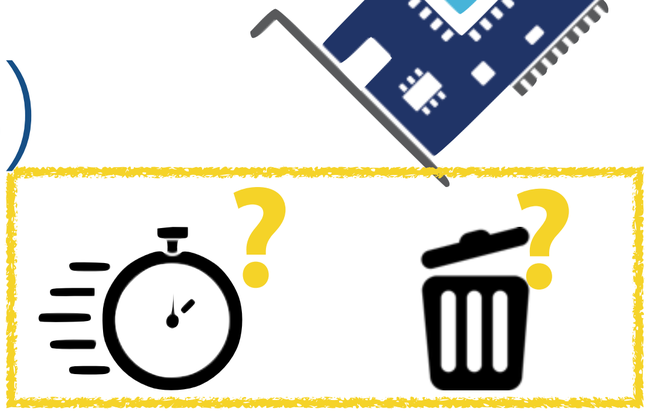
Receiver-driven Congestion Control



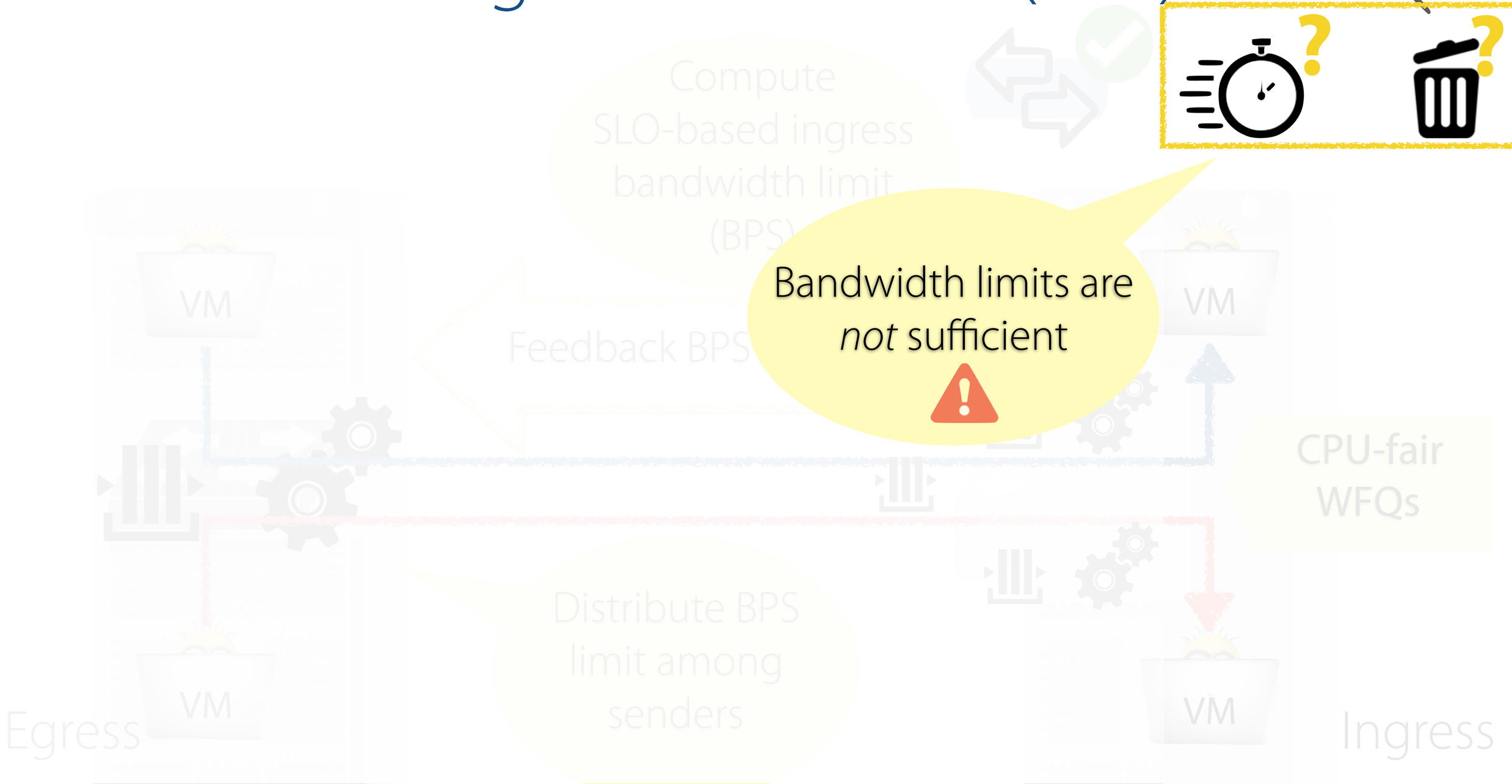
Congestion Control (BPS)



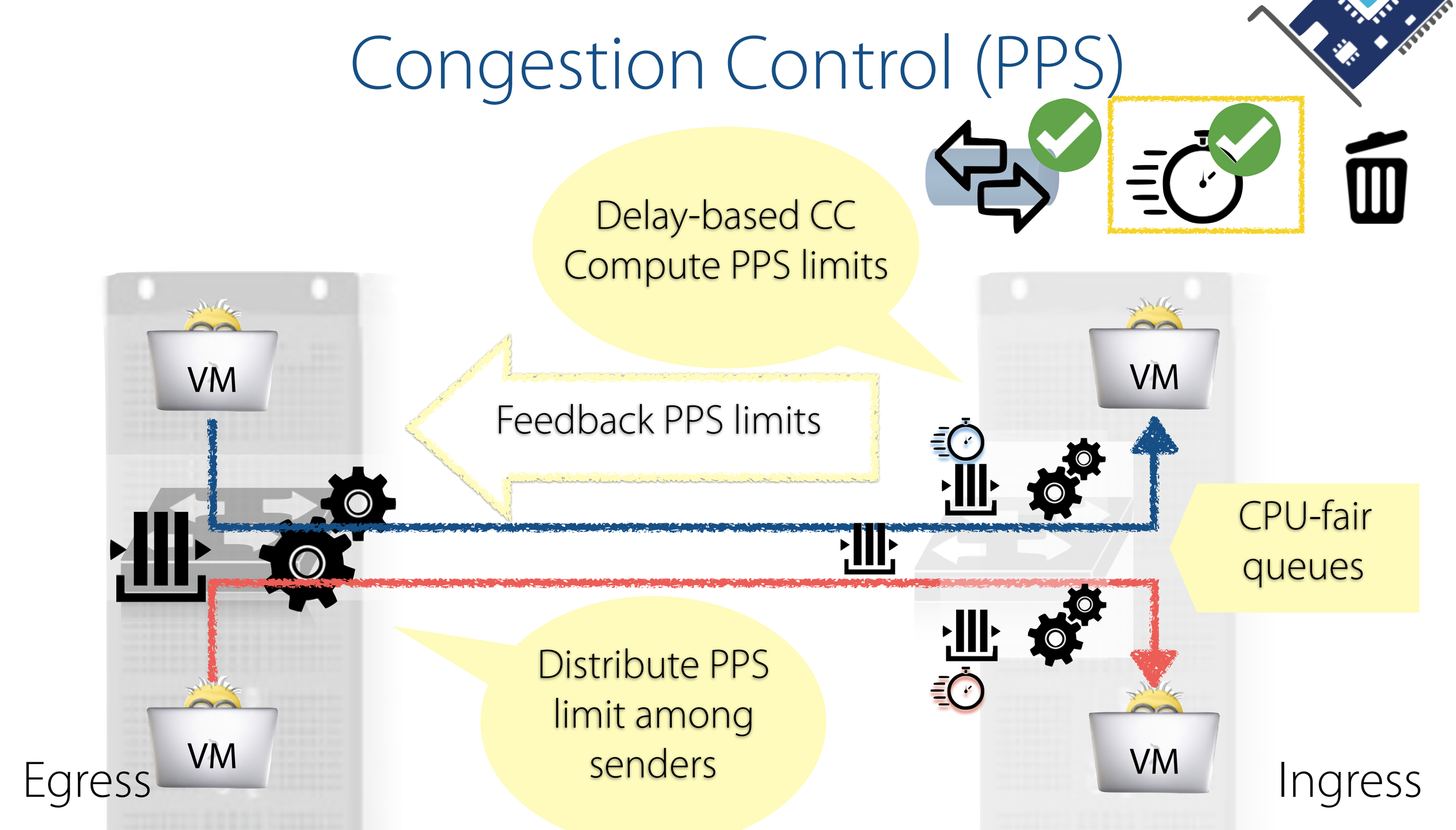
Congestion Control (BPS)



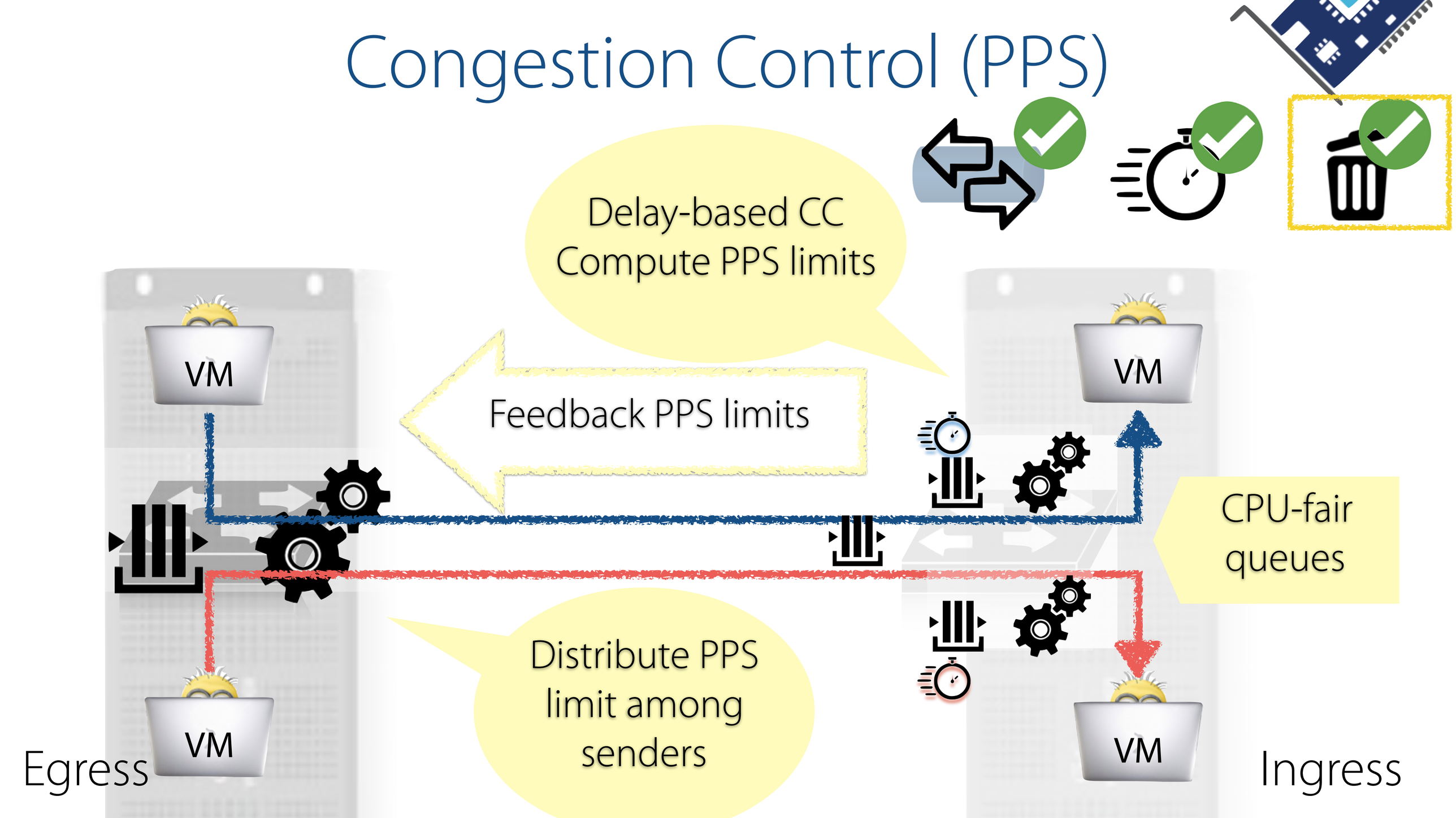
Bandwidth limits are *not* sufficient



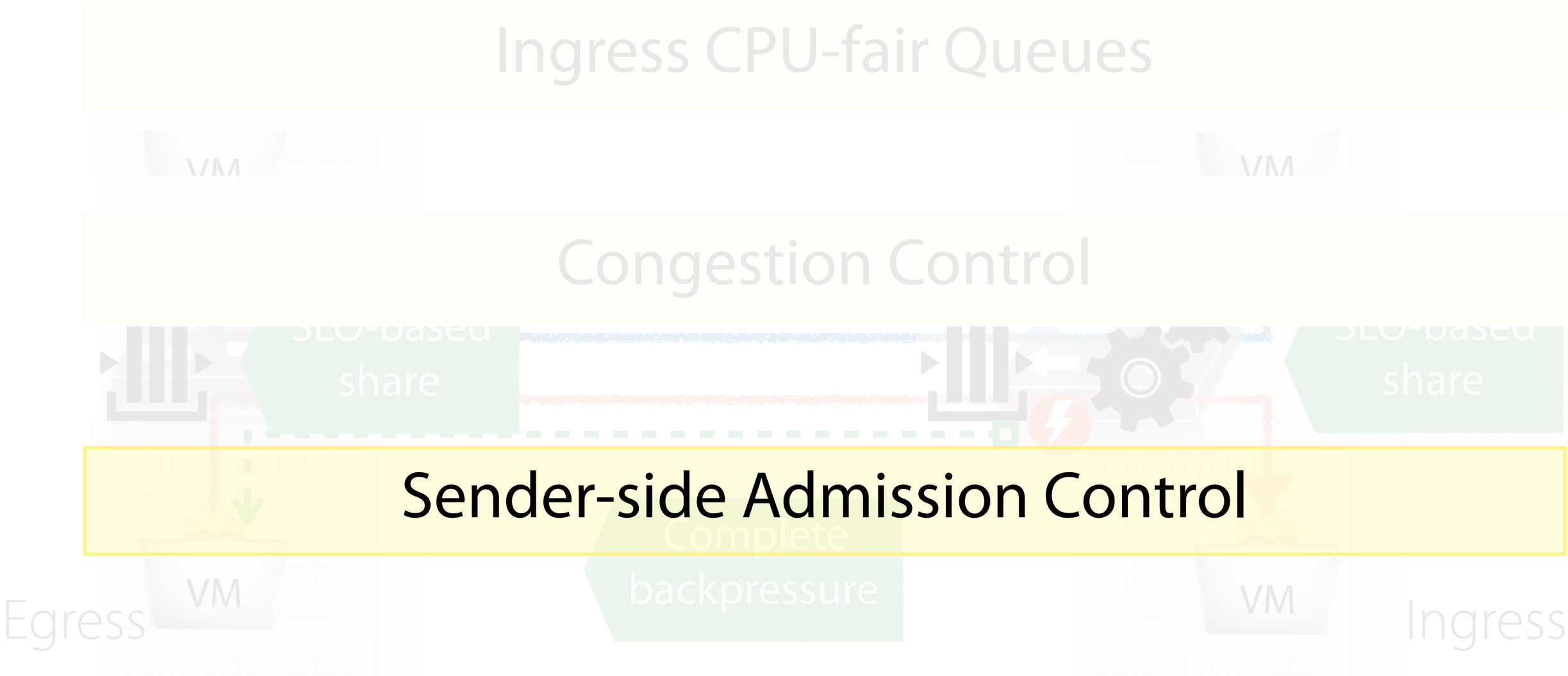
Congestion Control (PPS)



Congestion Control (PPS)

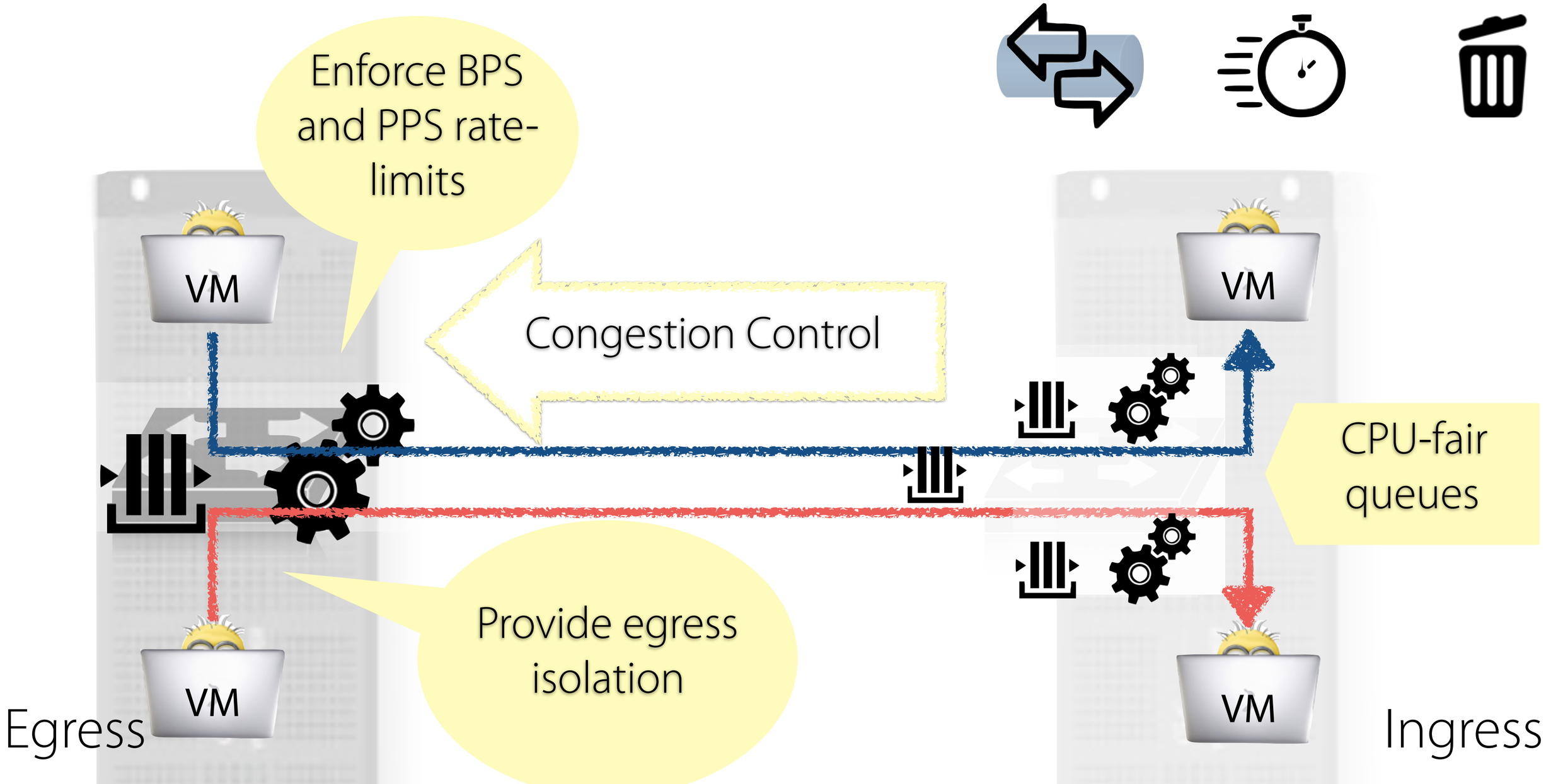


PicNIC Design

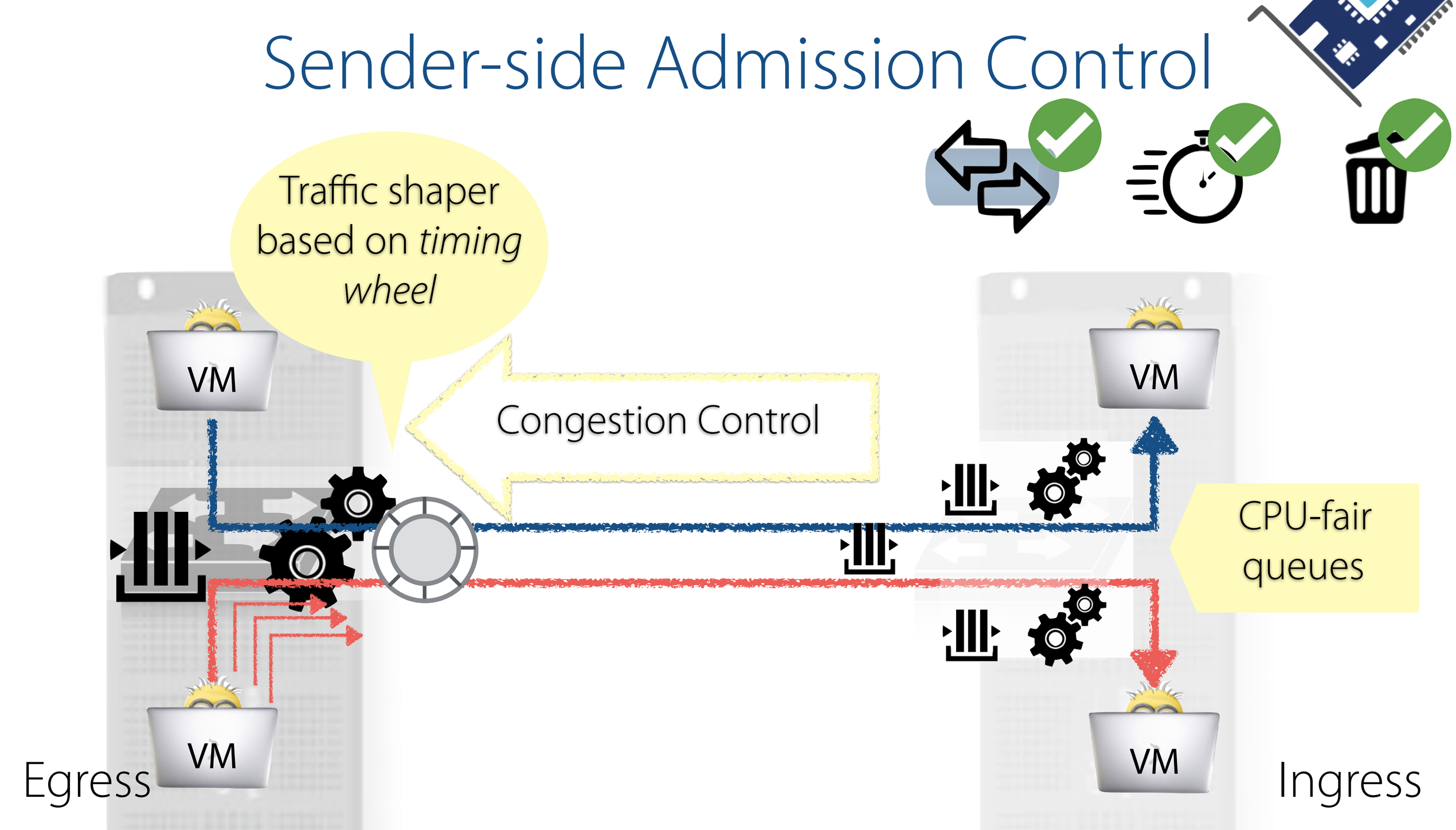


Sender-side Admission Control

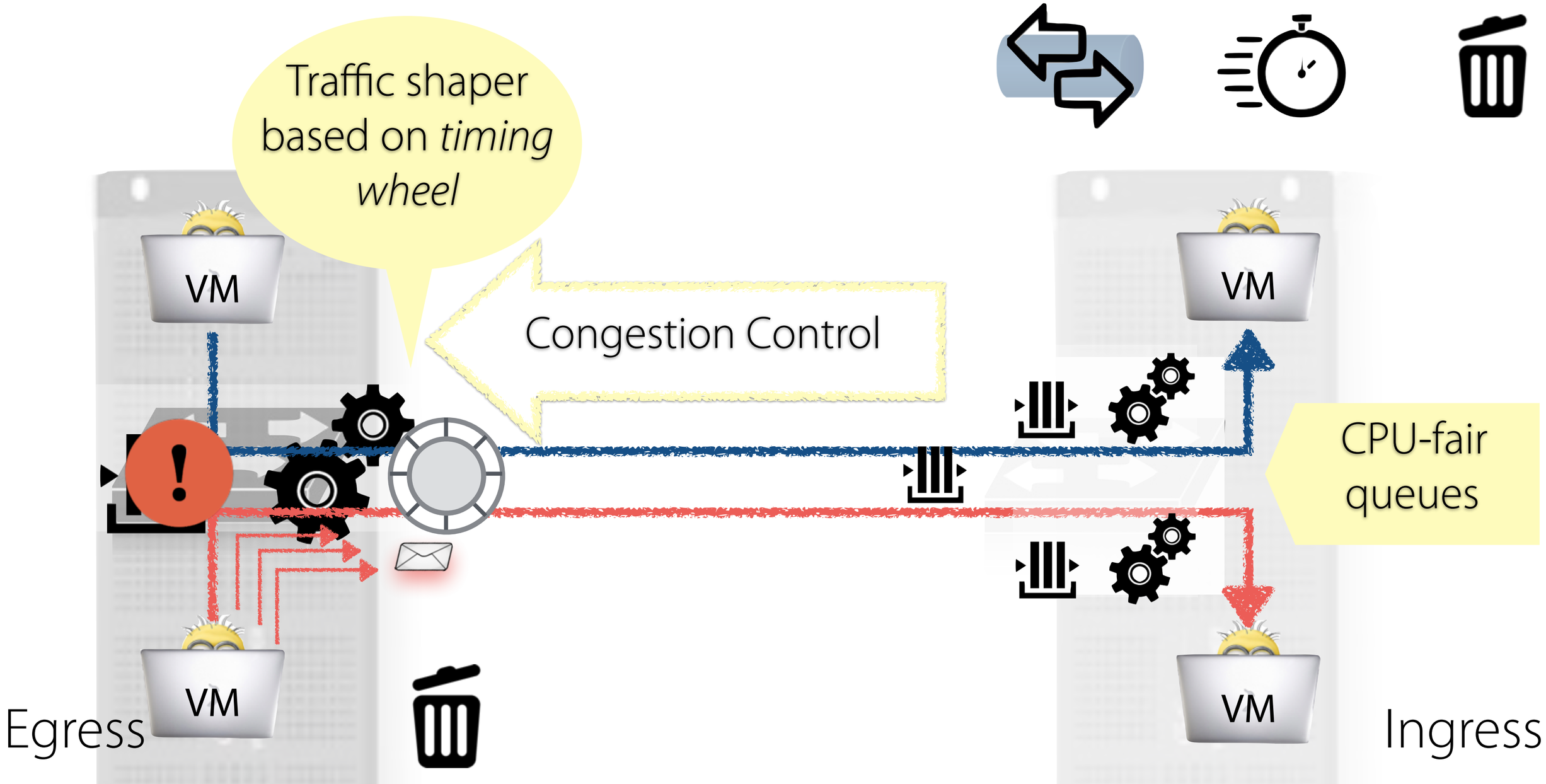
Sender-side Admission Control



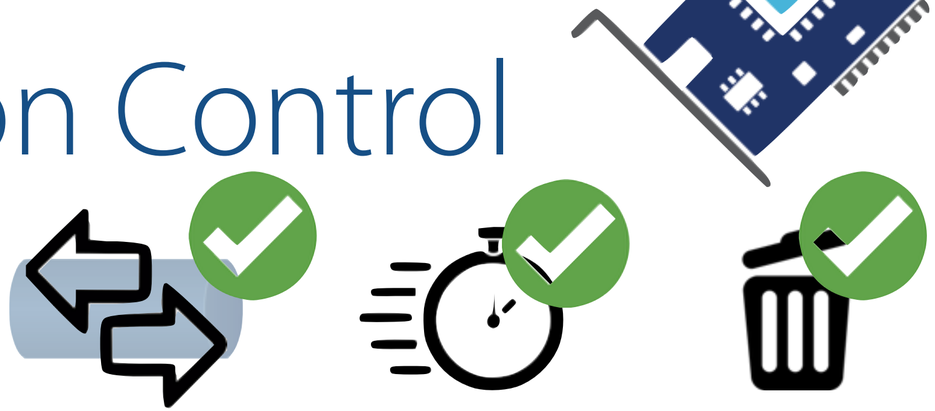
Sender-side Admission Control

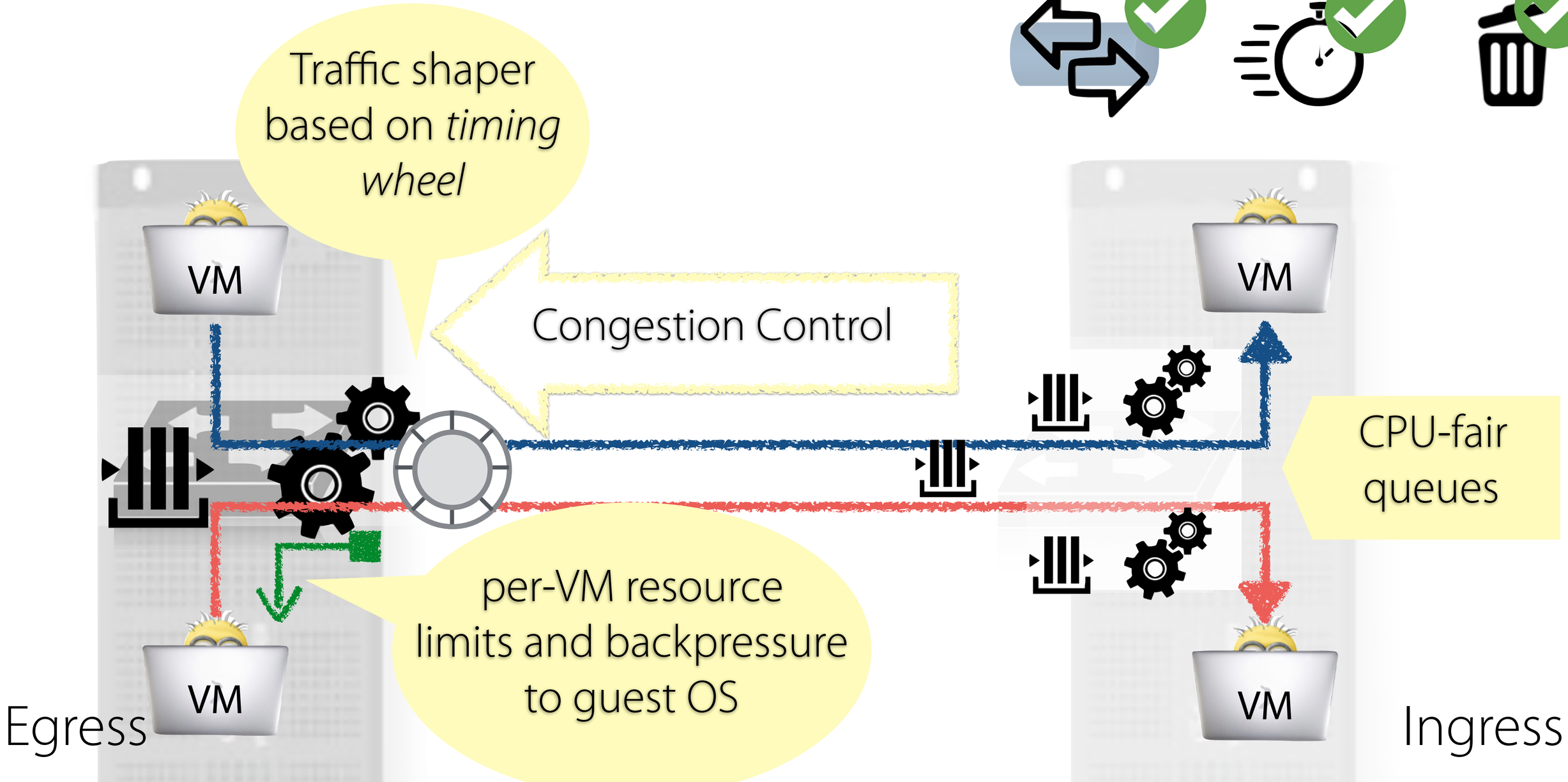


Sender-side Admission Control



Sender-side Admission Control





Summary

Principles

SLO-based sharing

Complete backpressure

Mechanisms

Local

Sender-side admission control

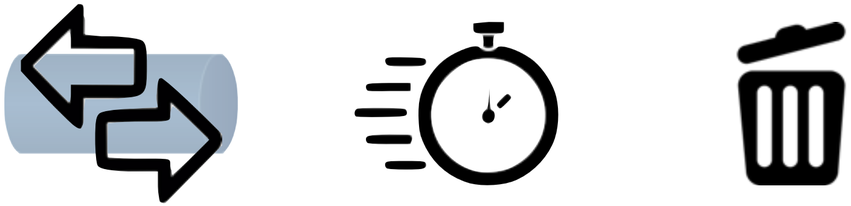
Ingress CPU-fair queues

BPS and PPS limits

Global

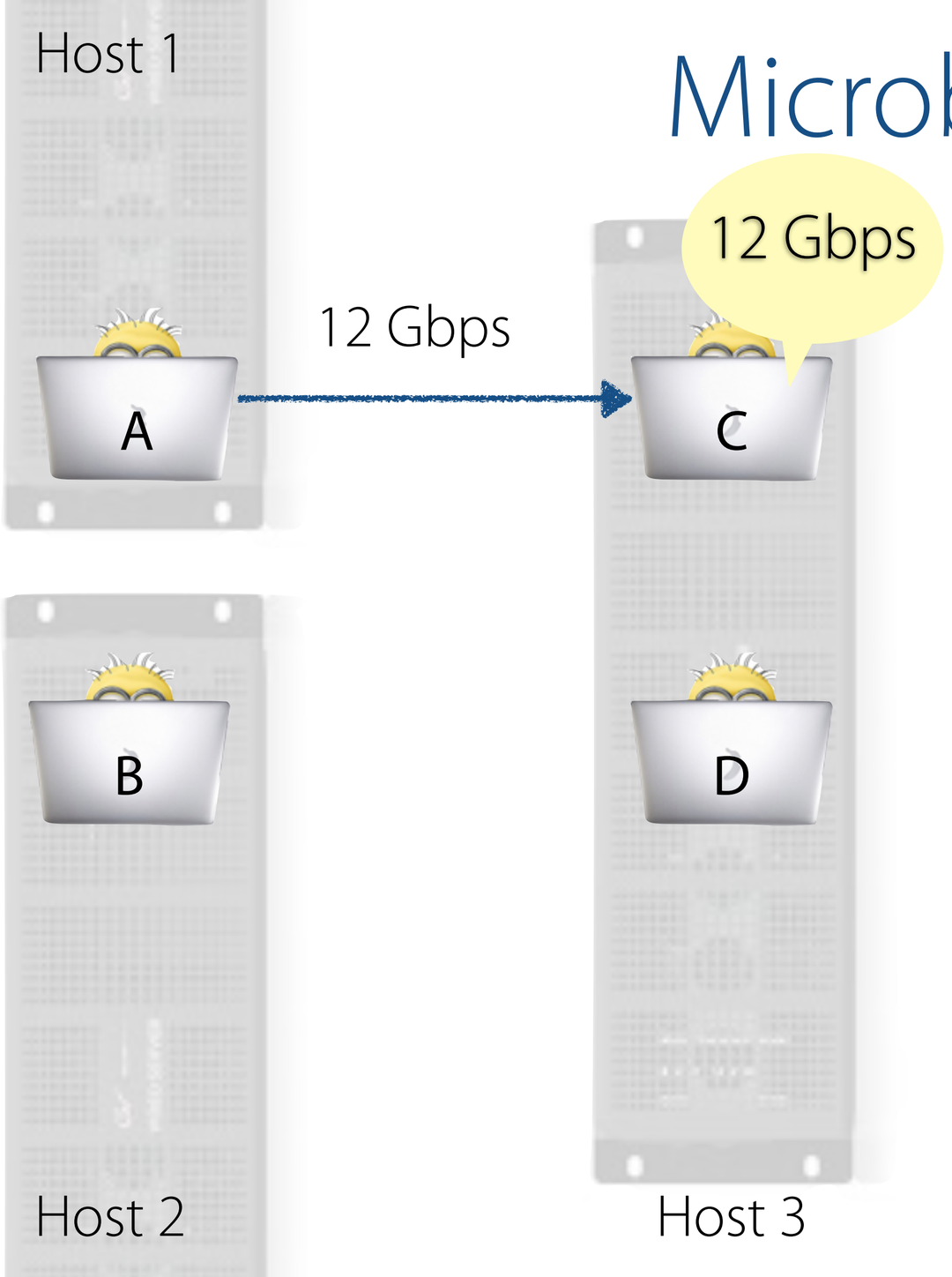
Congestion control

Evaluation



Does PicNIC provide predictable performance?

Microbenchmarks



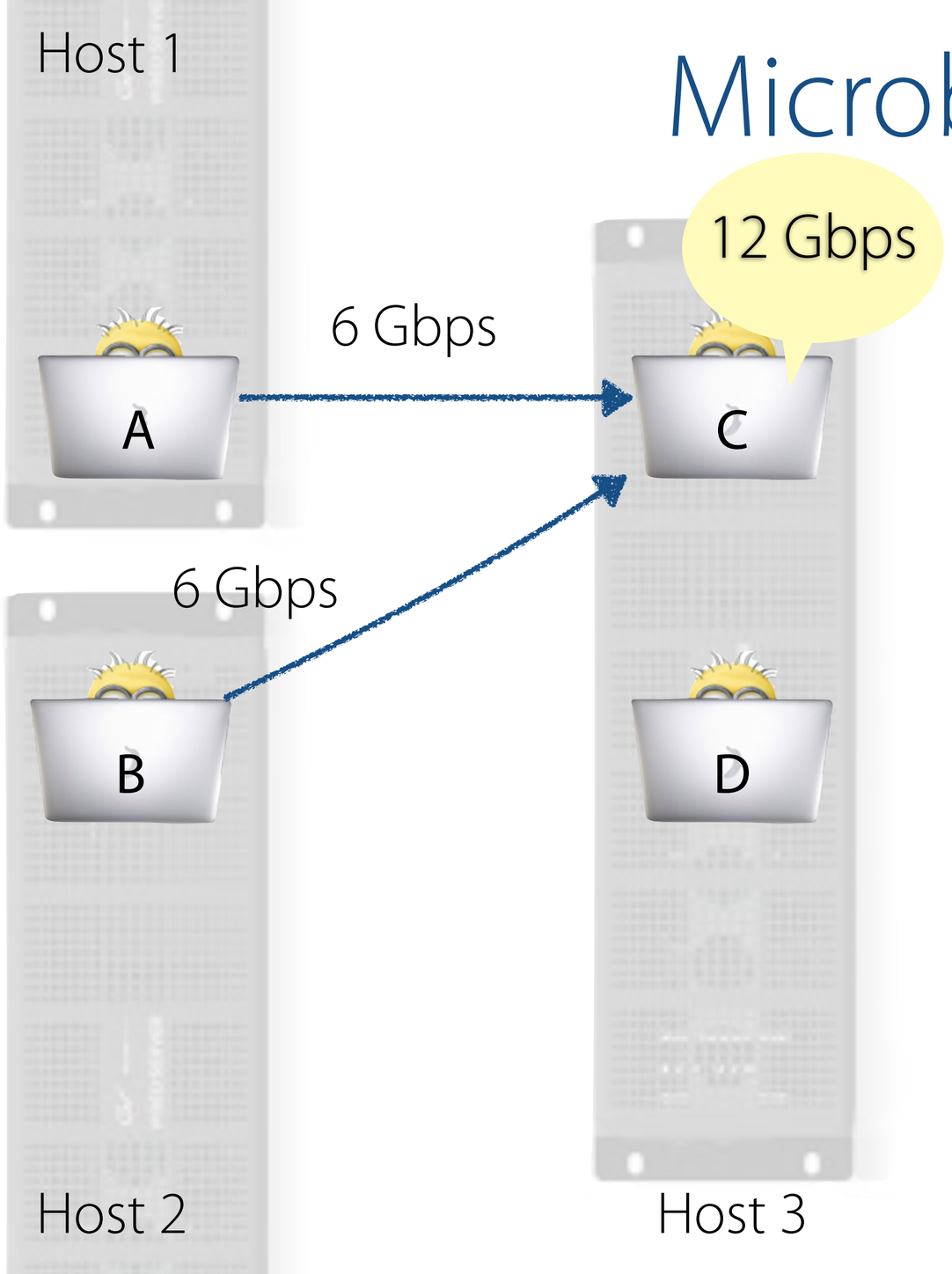
Bandwidth envelope per VM
min: 4 Gbps
max: 12 Gbps

Hose model

Host NIC: 20 Gbps

Enforce maximum bandwidth limit

Microbenchmarks



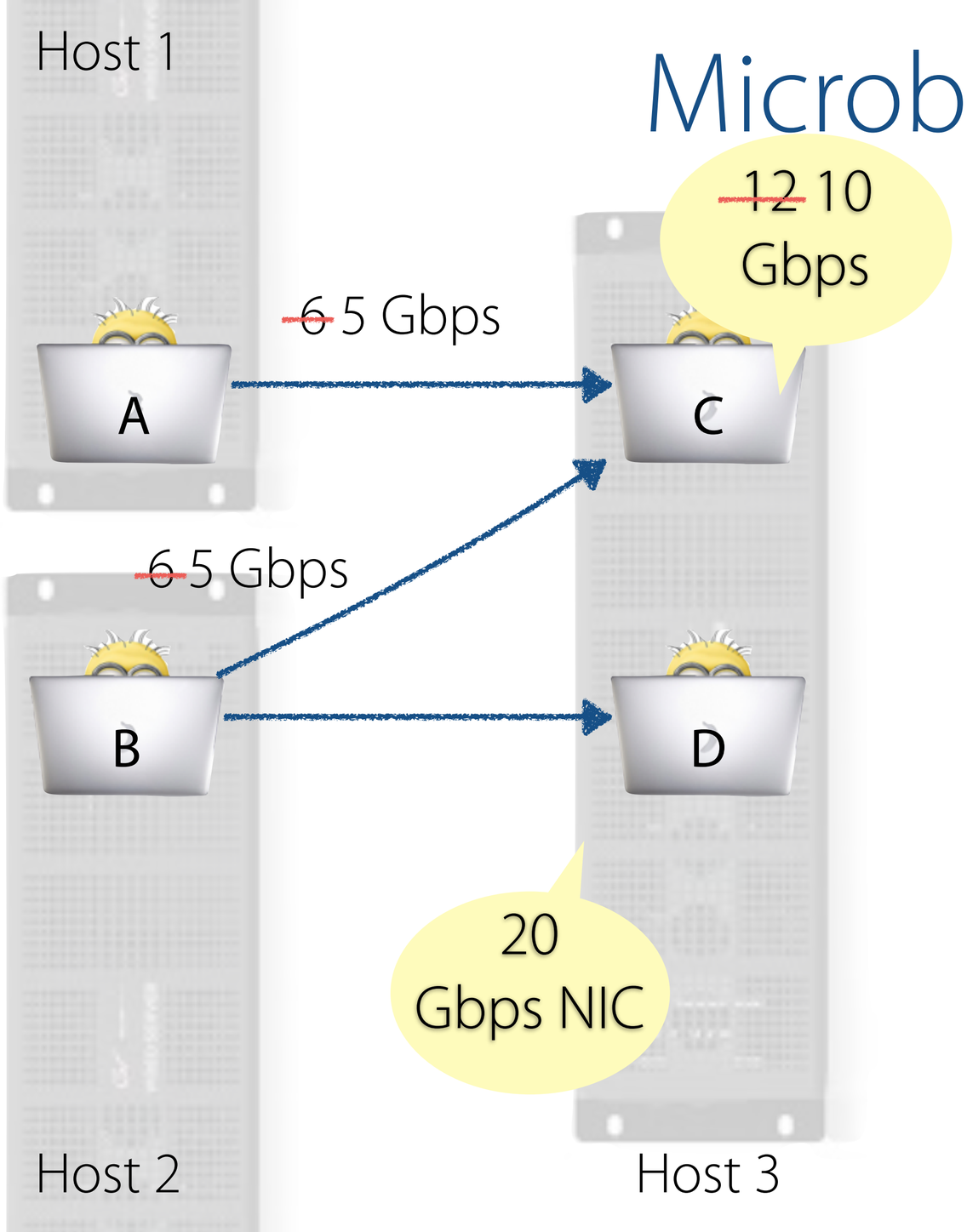
Bandwidth envelope per VM
min: 4 Gbps
max: 12 Gbps

Hose model

Host NIC: 20 Gbps

Contention at VM C (receiver)

Microbenchmarks



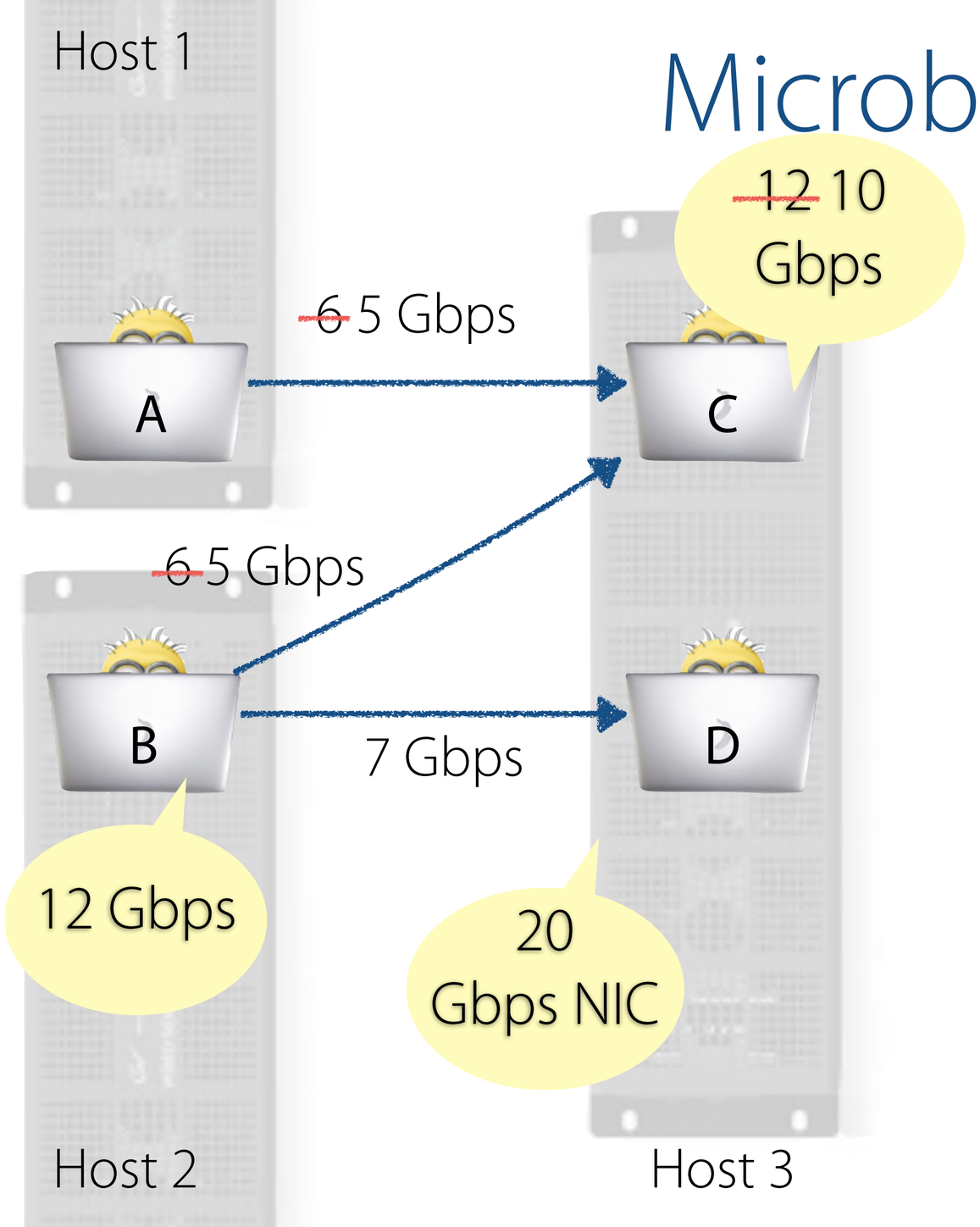
Bandwidth envelope per VM
min: 4 Gbps
max: 12 Gbps

Hose model

Host NIC: 20 Gbps

Contention at host 3 NIC (receiver)

Microbenchmarks



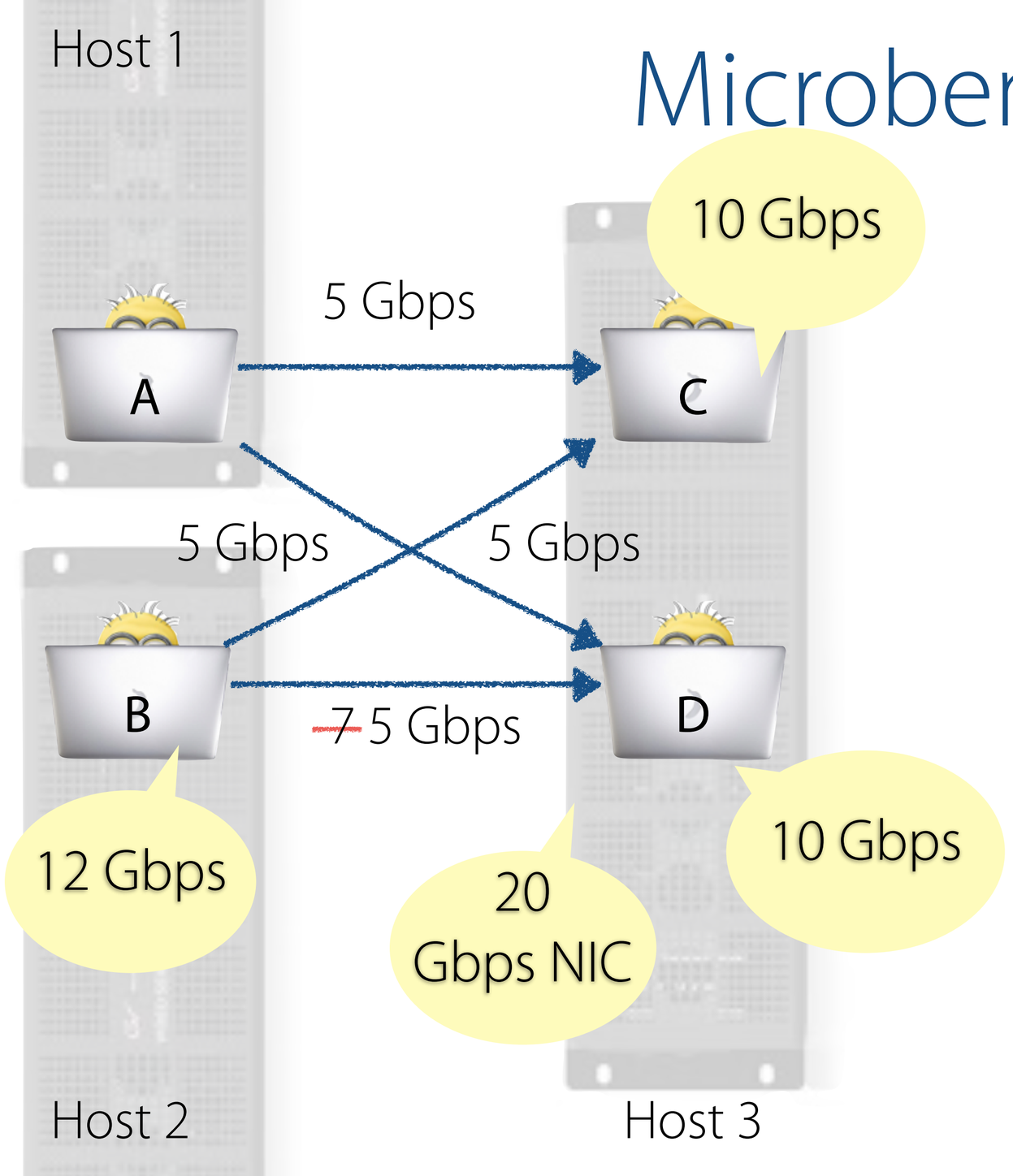
Bandwidth envelope per VM
min: 4 Gbps
max: 12 Gbps

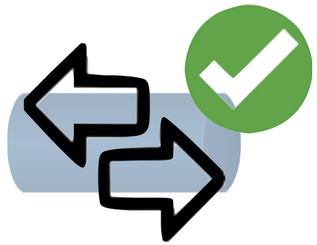
Hose model

Host NIC: 20 Gbps

Contention at VM B (sender)

Microbenchmarks





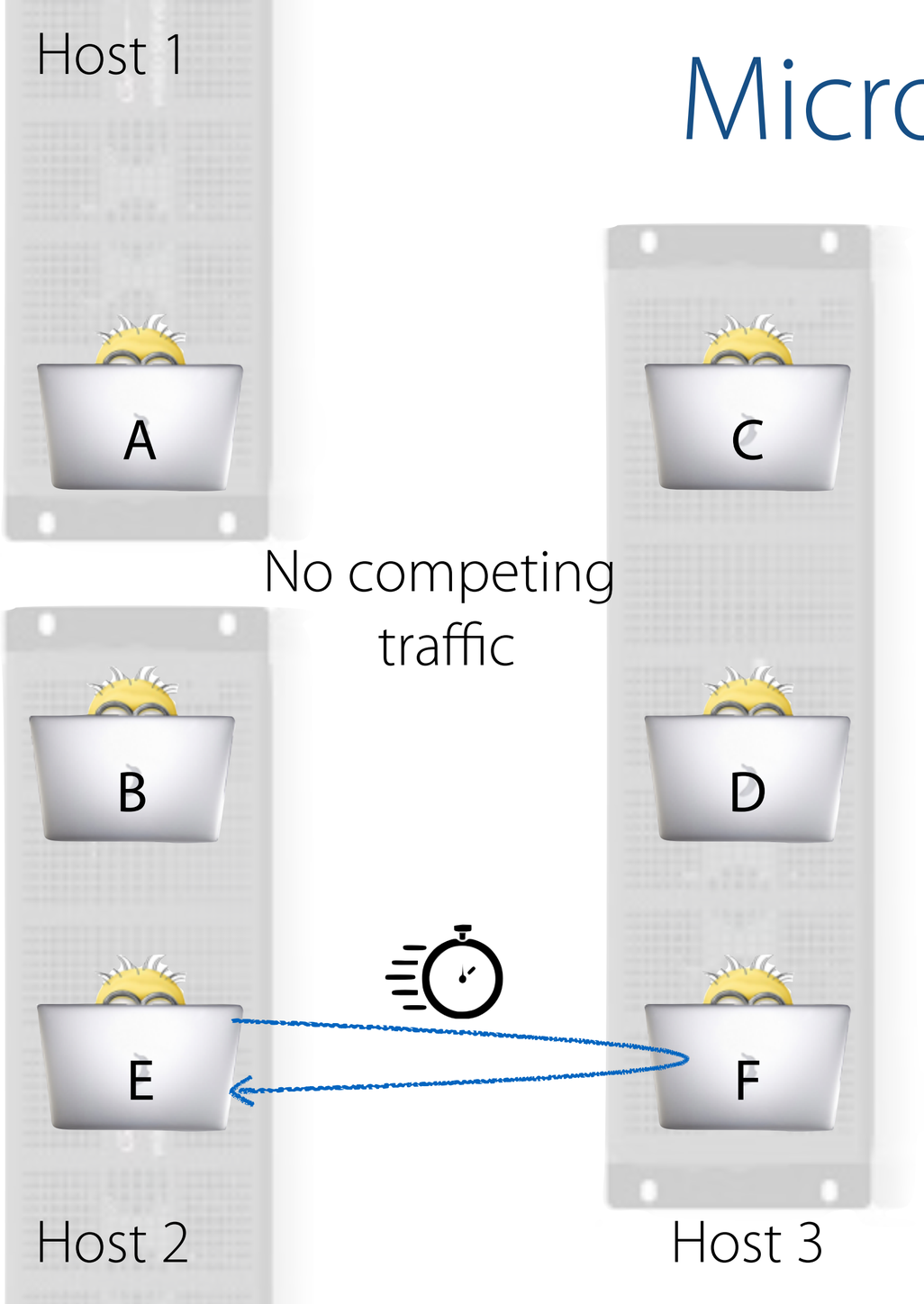
Bandwidth envelope per VM
min: 4 Gbps
max: 12 Gbps

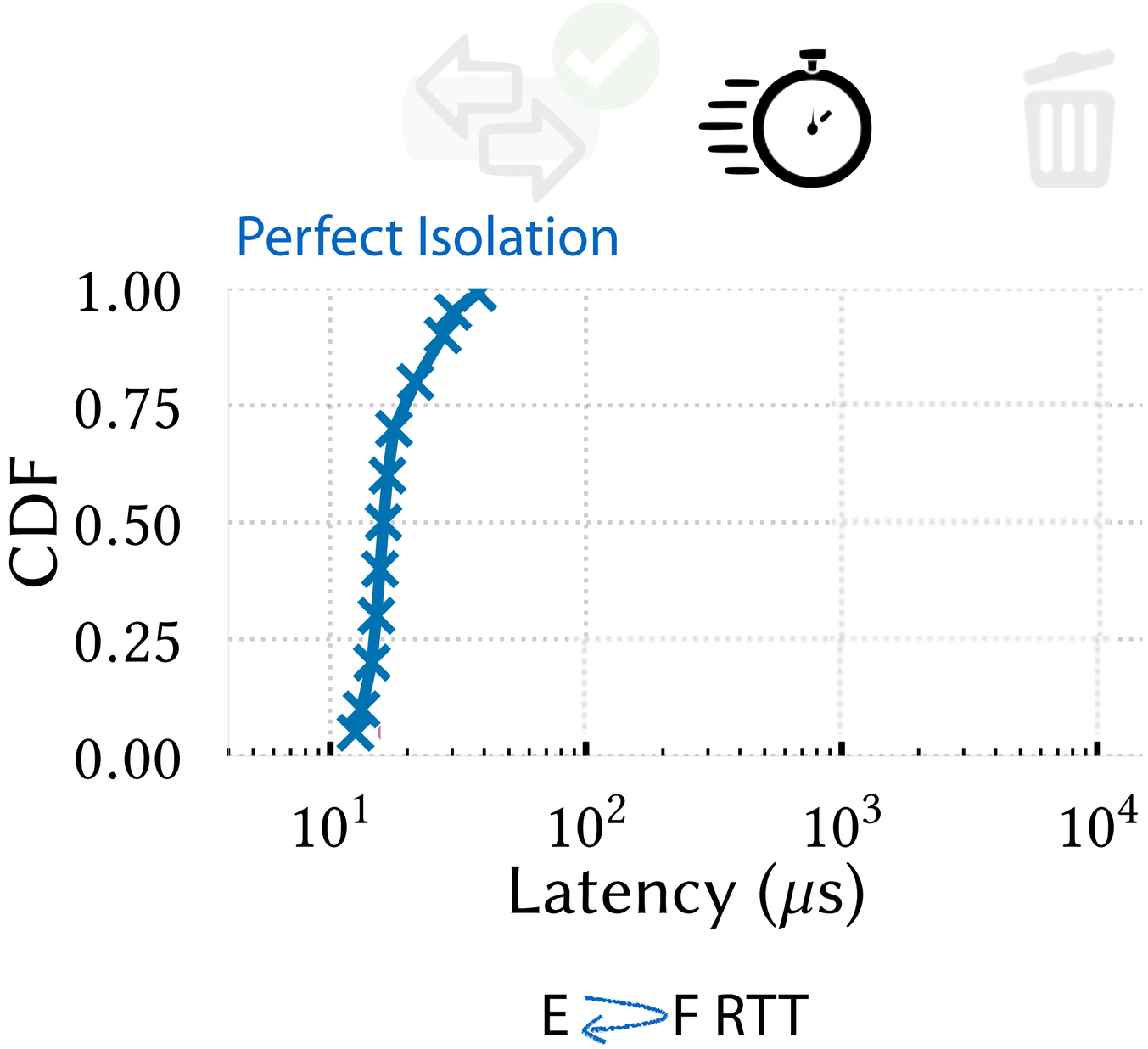
Hose model

Host NIC: 20 Gbps

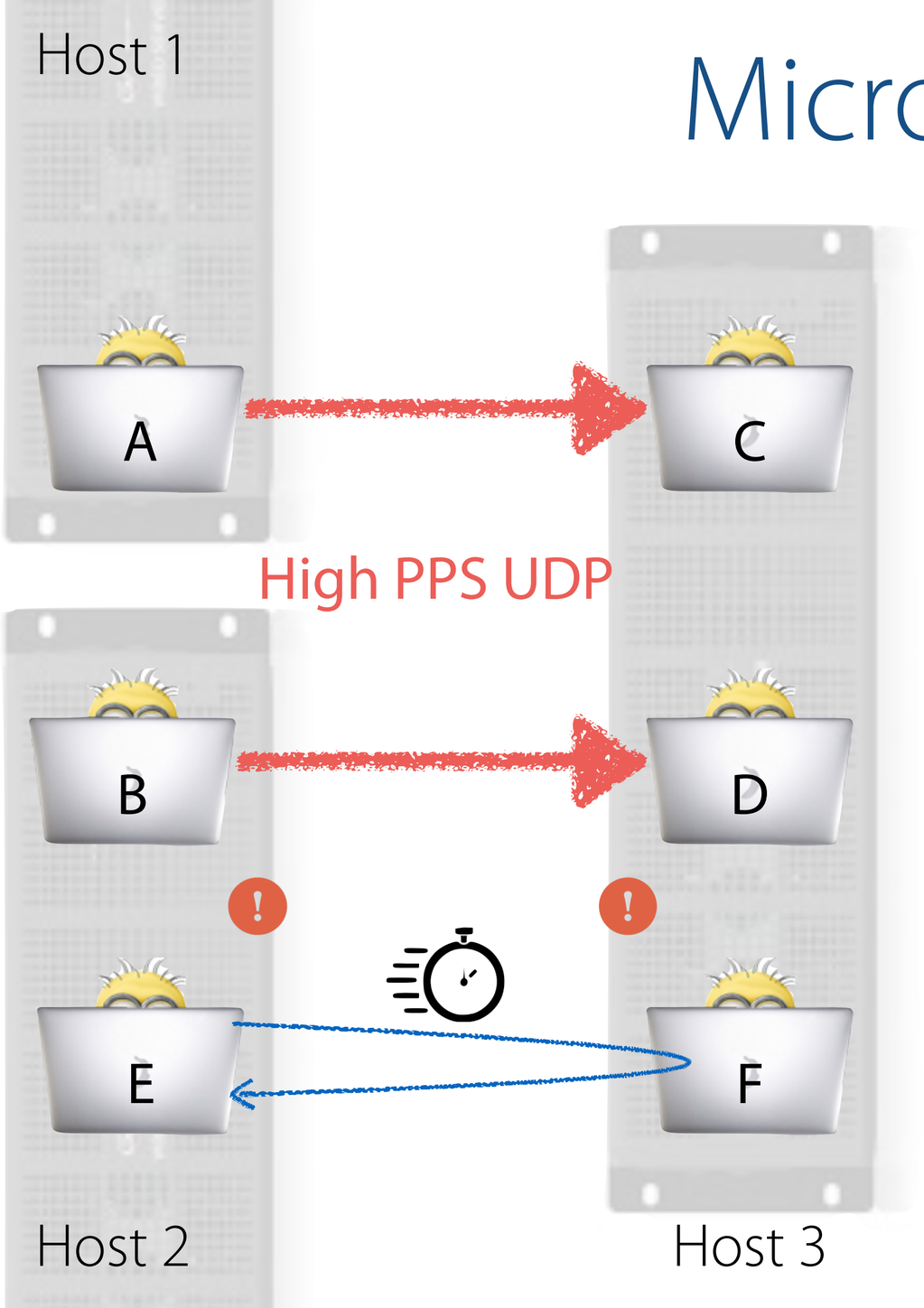
Contention at VM D (receiver)

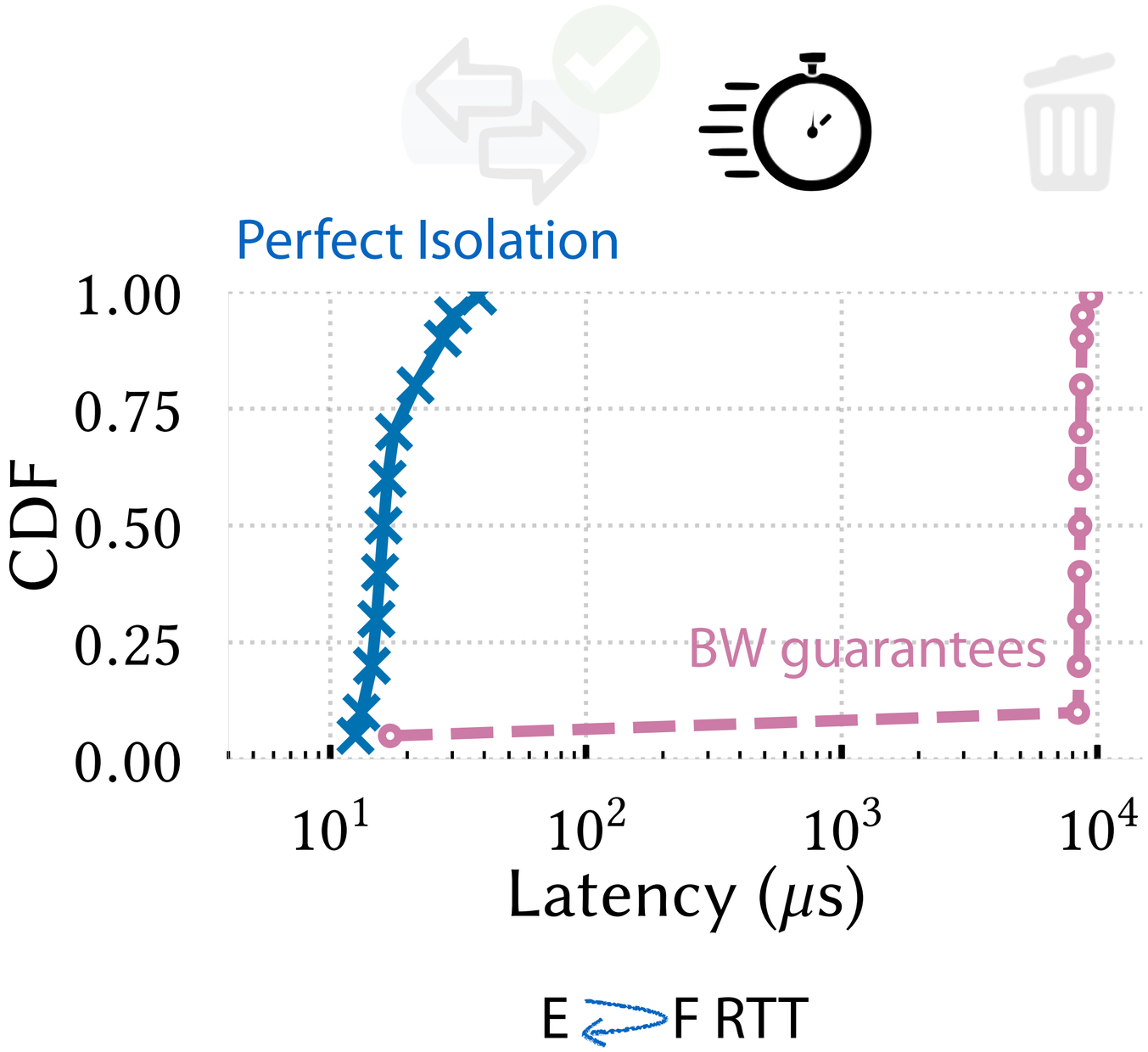
Microbenchmarks



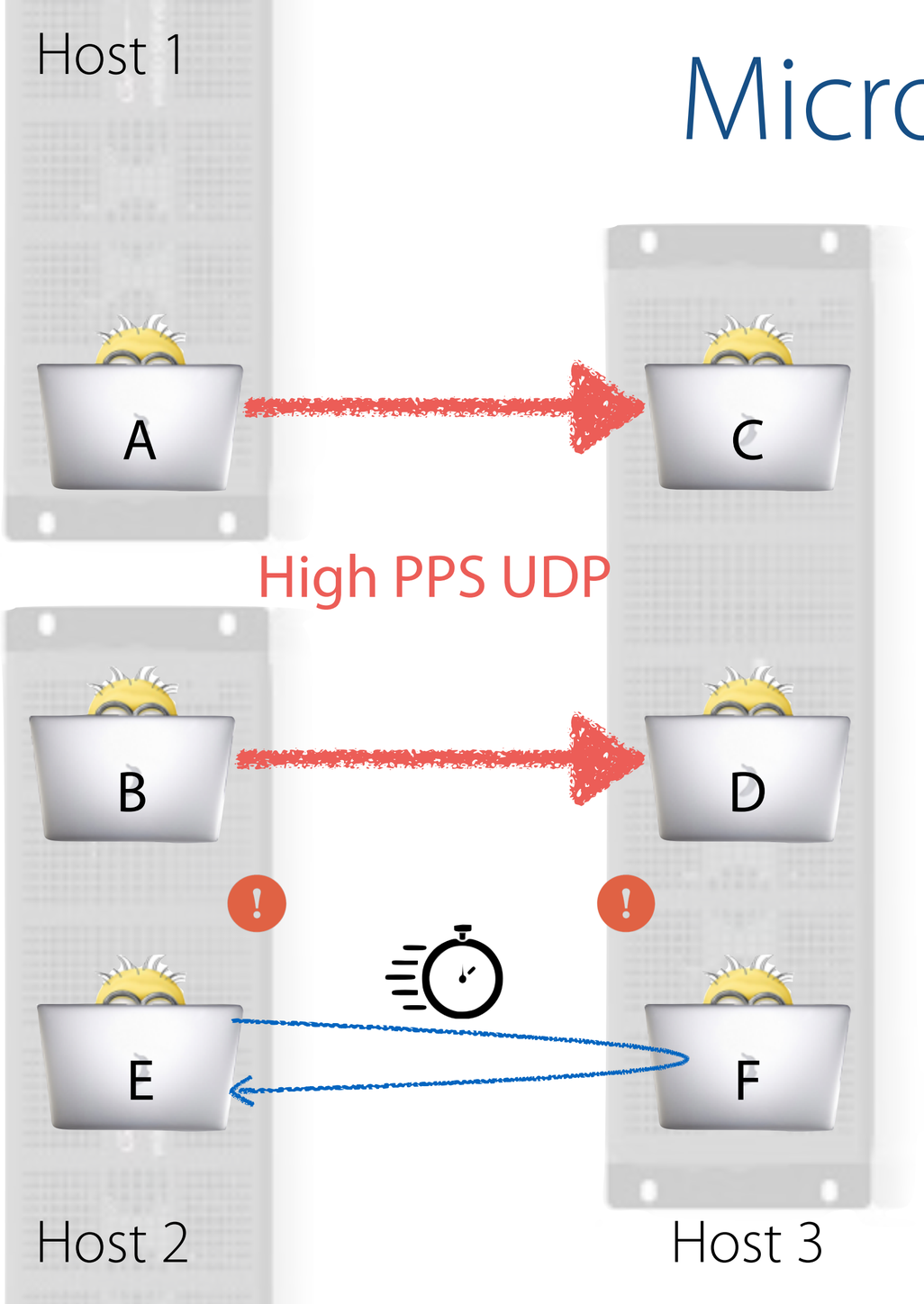


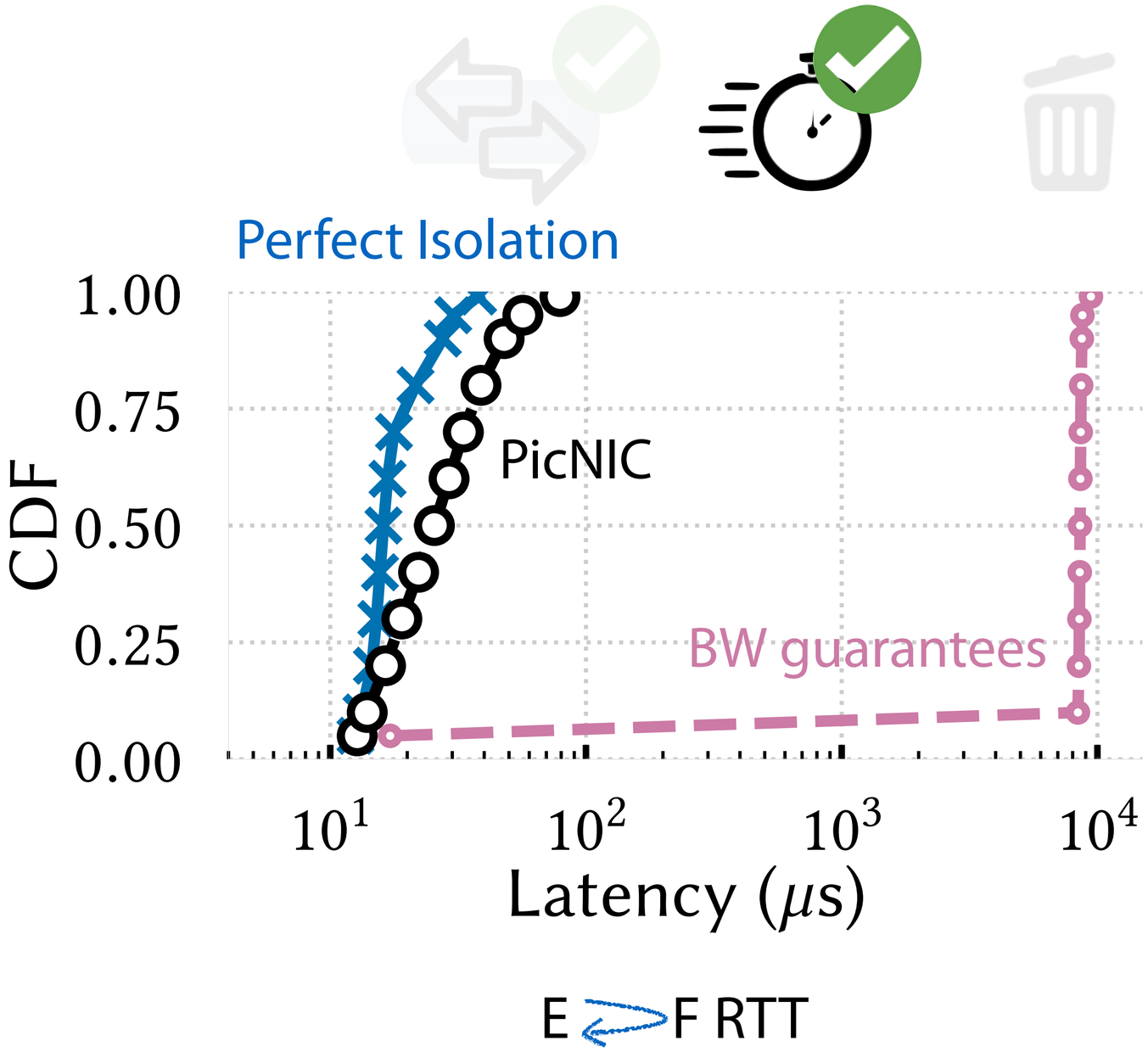
Microbenchmarks



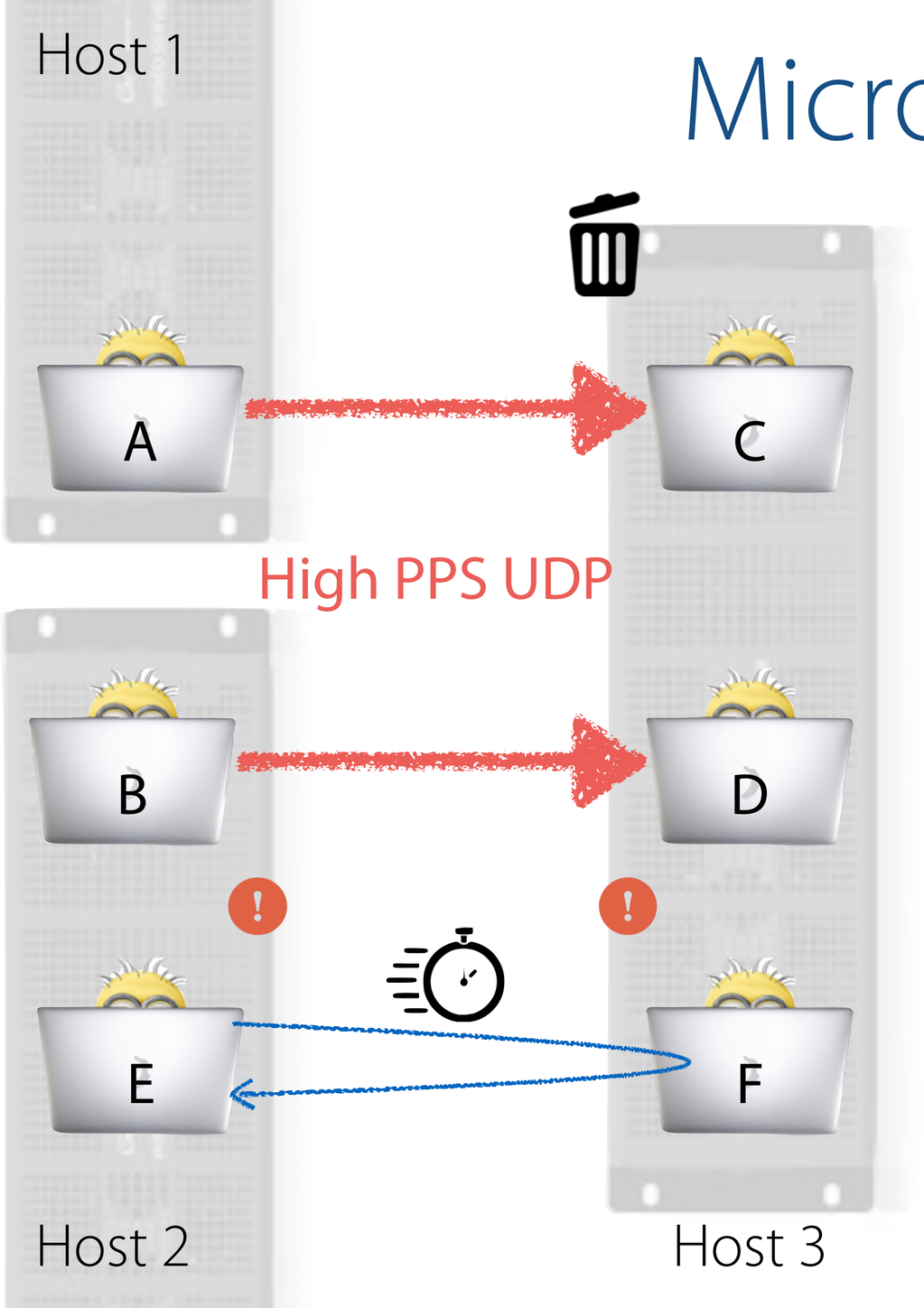


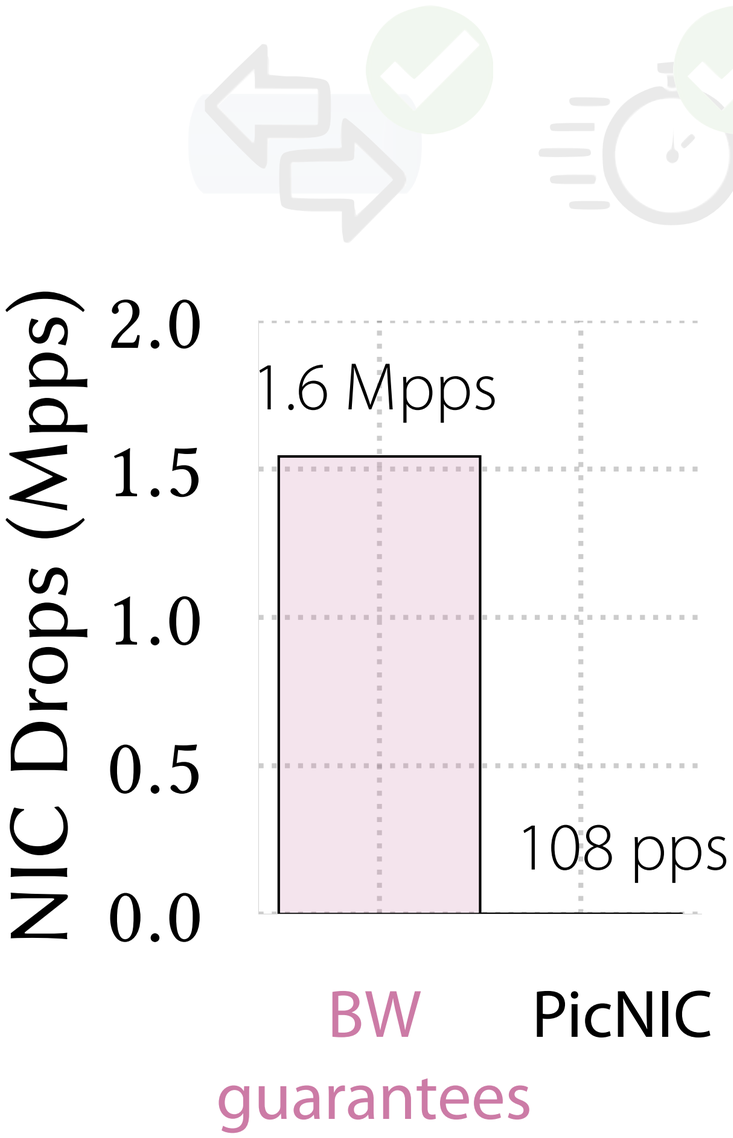
Microbenchmarks

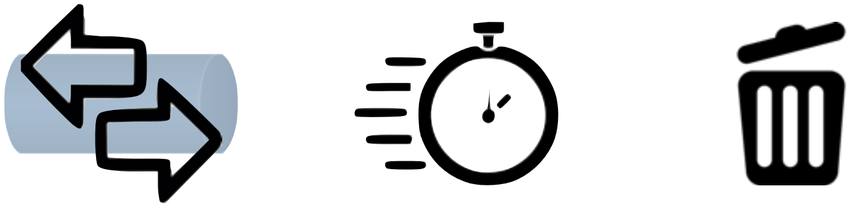




Microbenchmarks

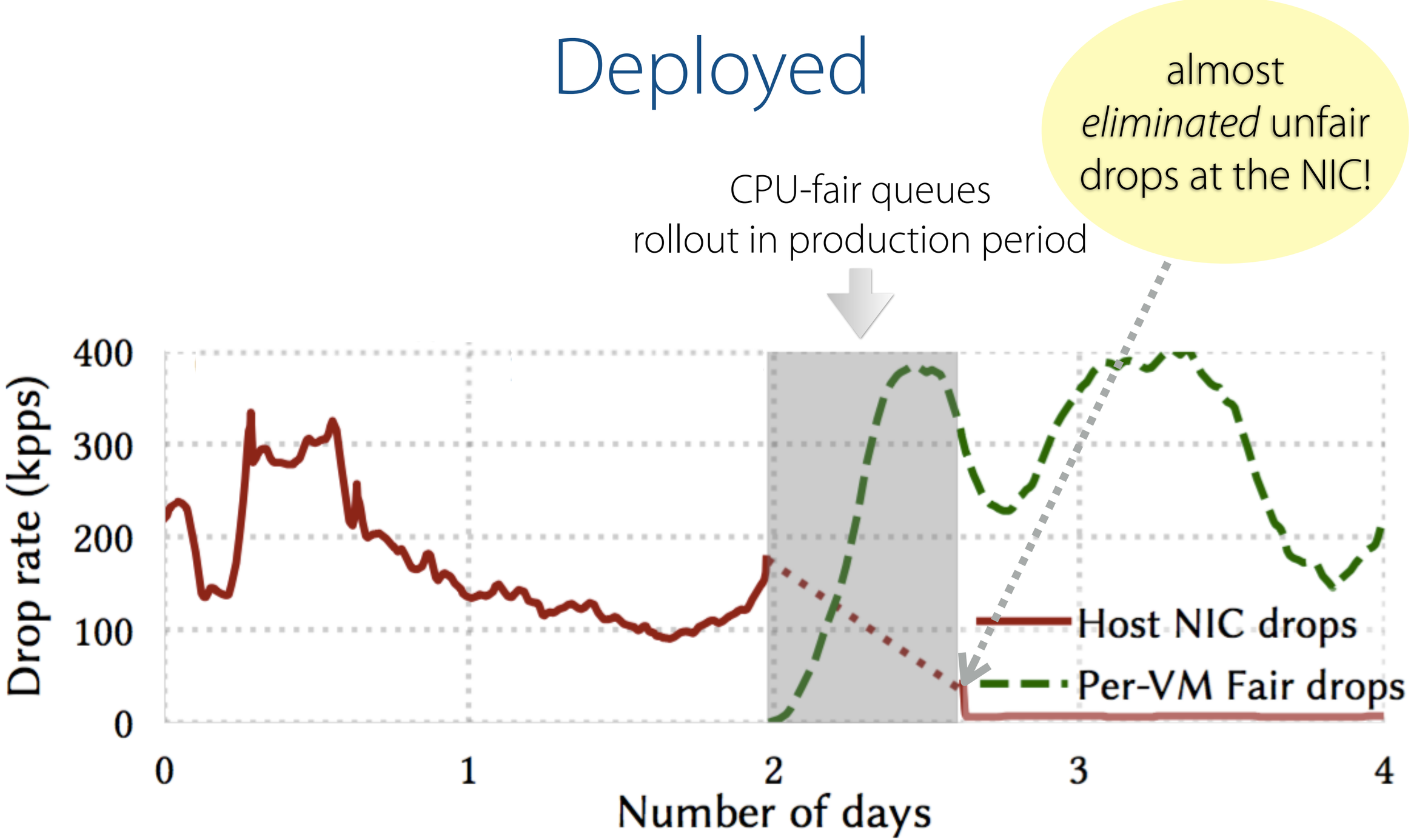


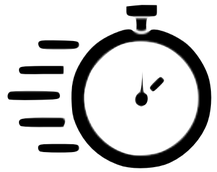




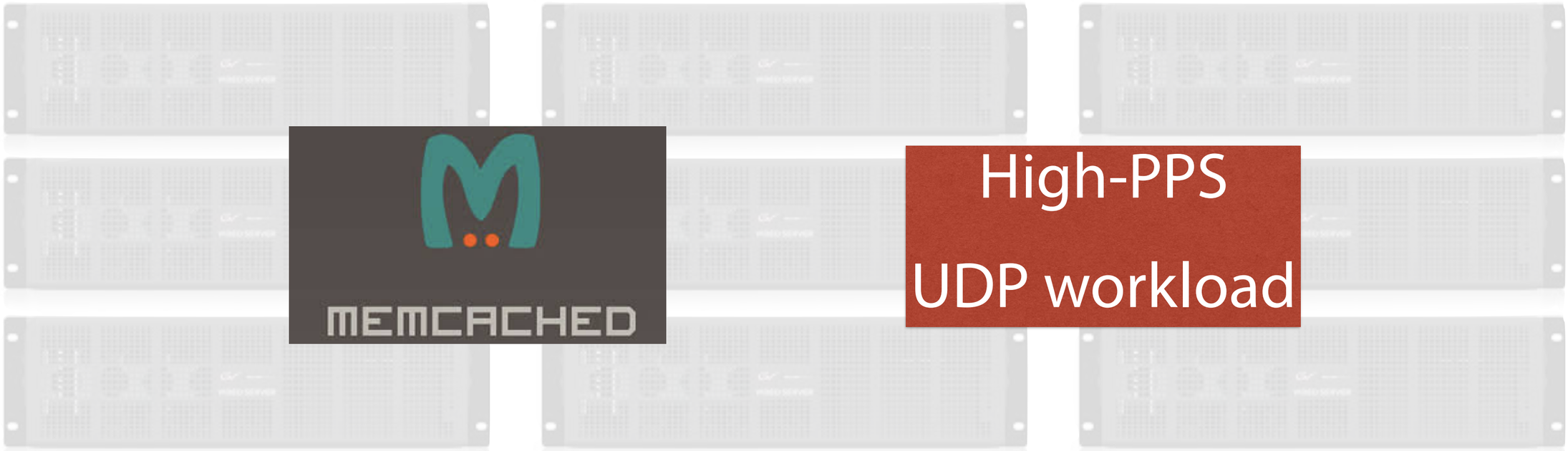
Is it deployable?

Deployed


almost
eliminated unfair
drops at the NIC!
CPU-fair queues
rollout in production period
Drop rate (kpps)
400
300
200
100
0
0
1
2
3
4
Number of days
Host NIC drops
Per-VM Fair drops



What about application performance?

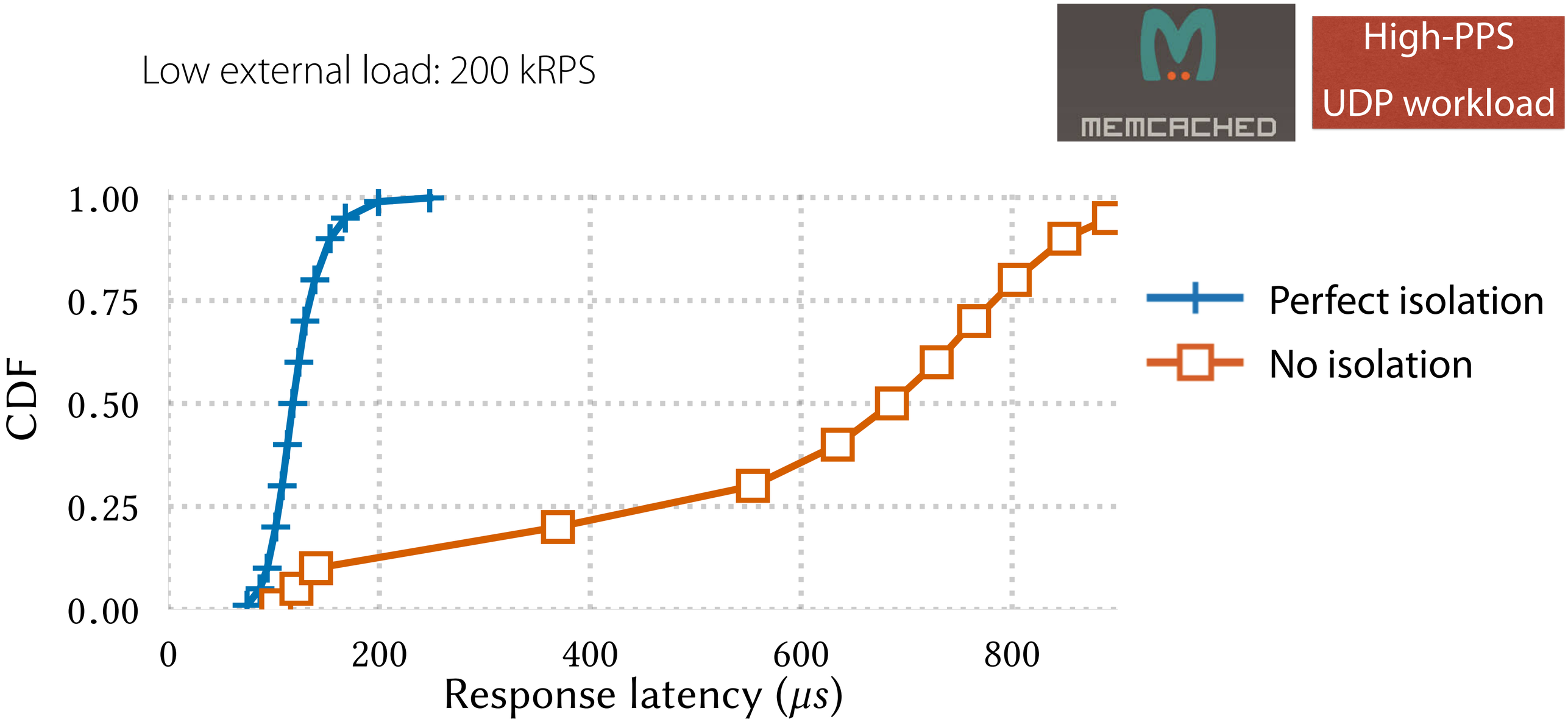


High-PPS

UDP workload

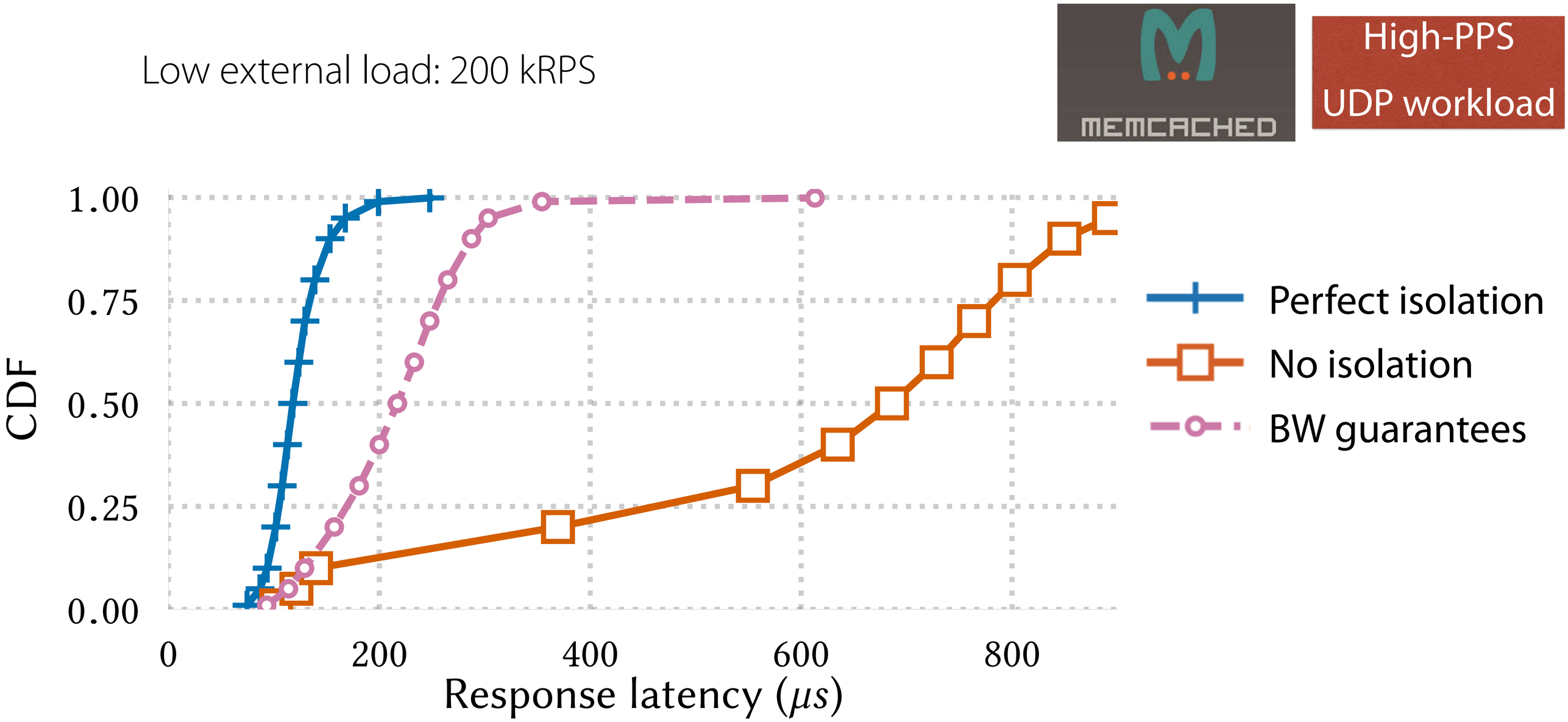
Memcached Latency

Low external load: 200 kRPS



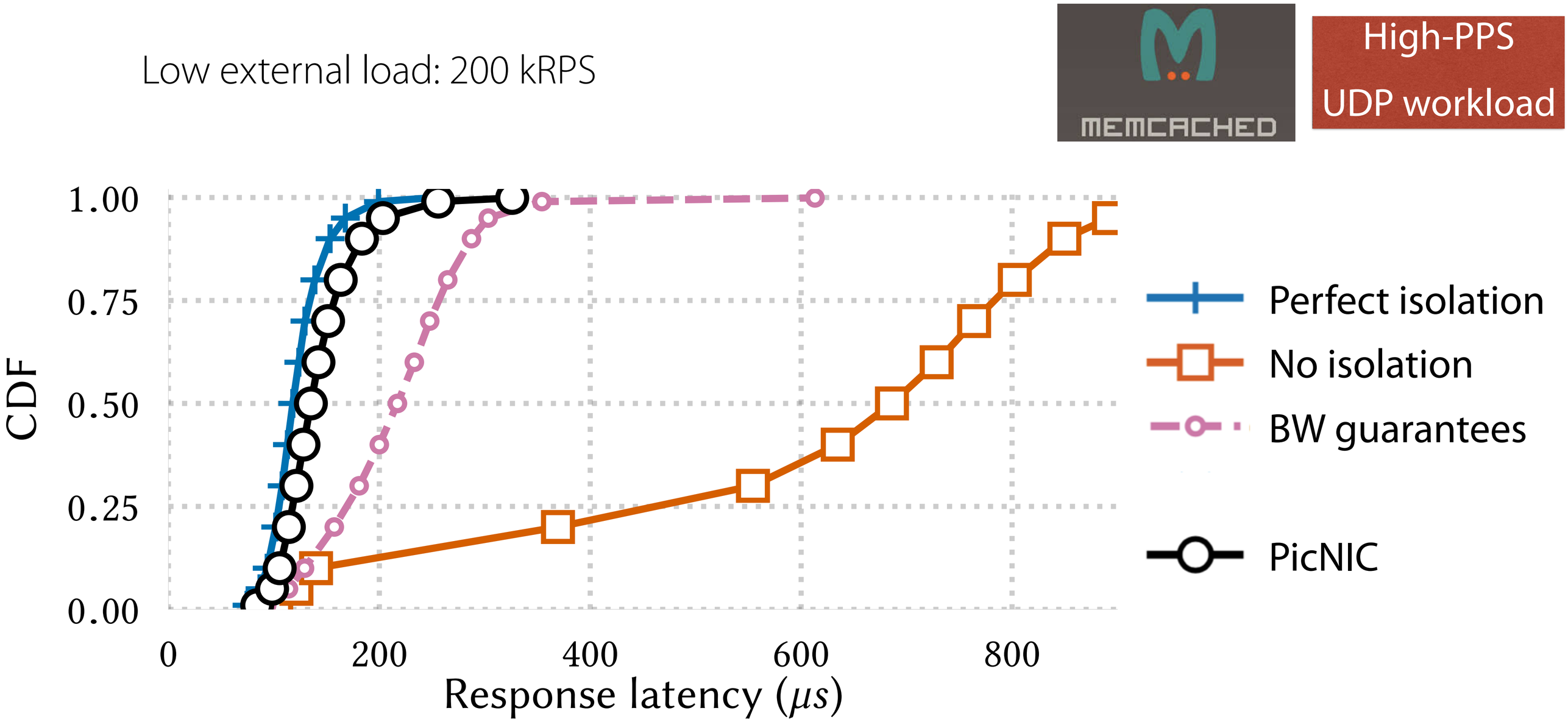
Memcached Latency

Low external load: 200 kRPS



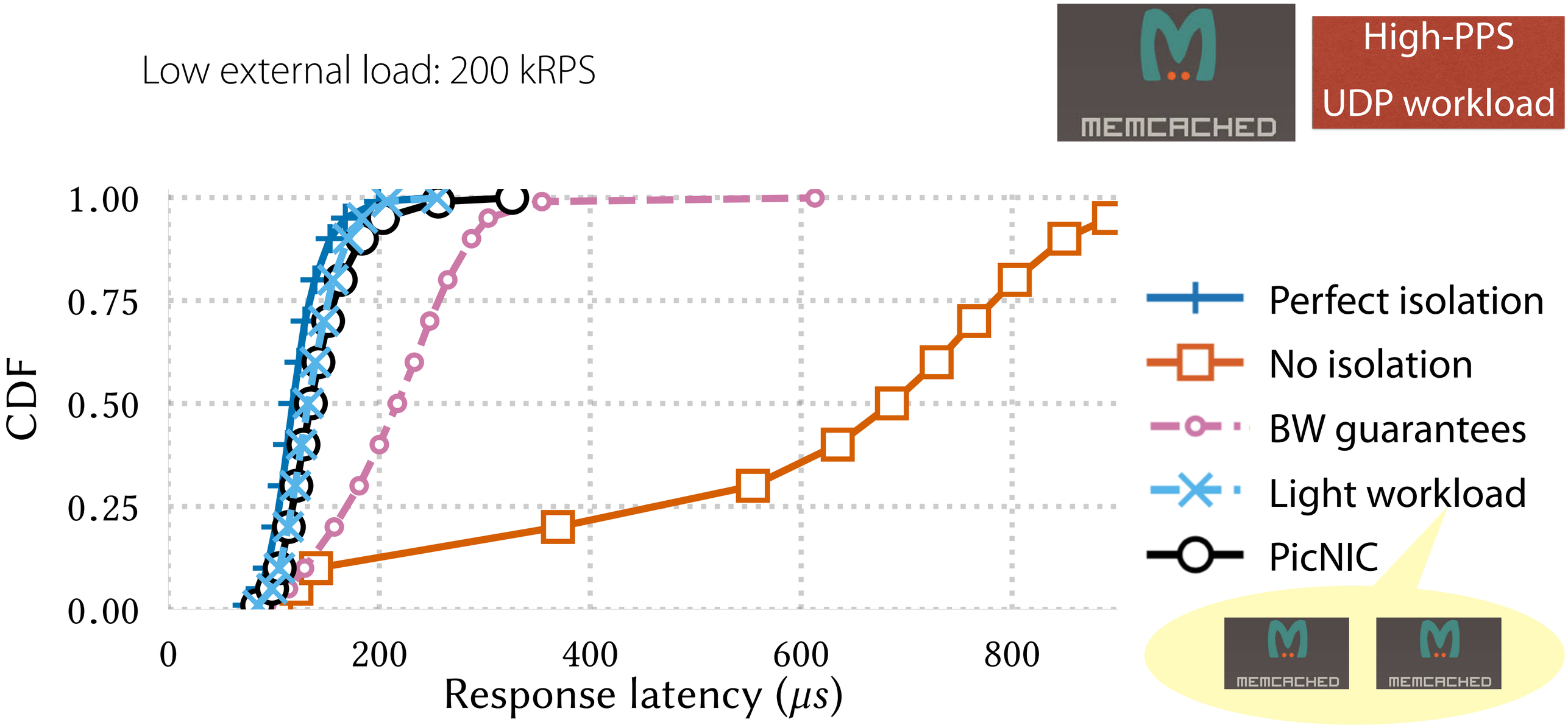
Memcached Latency

Low external load: 200 kRPS

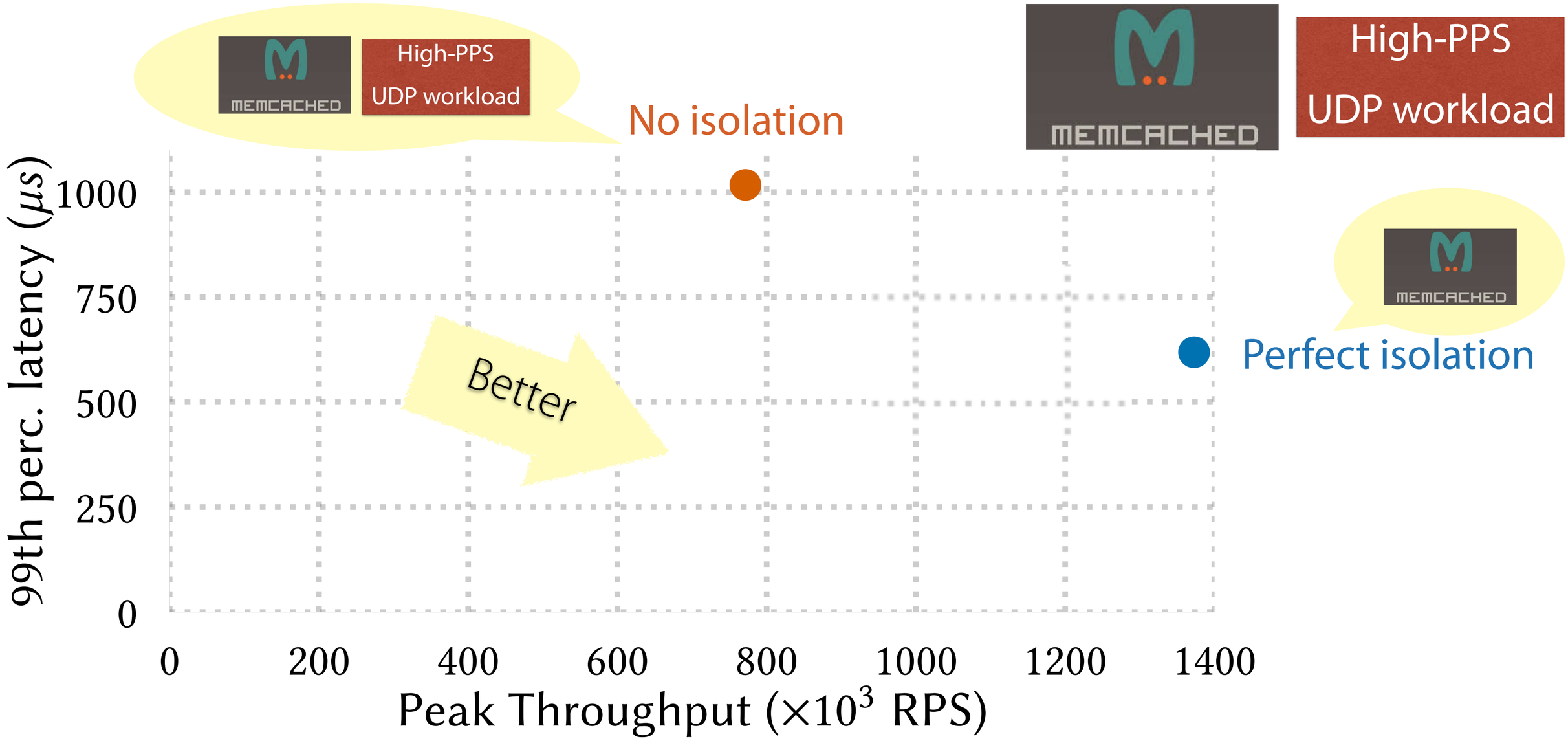


Memcached Latency

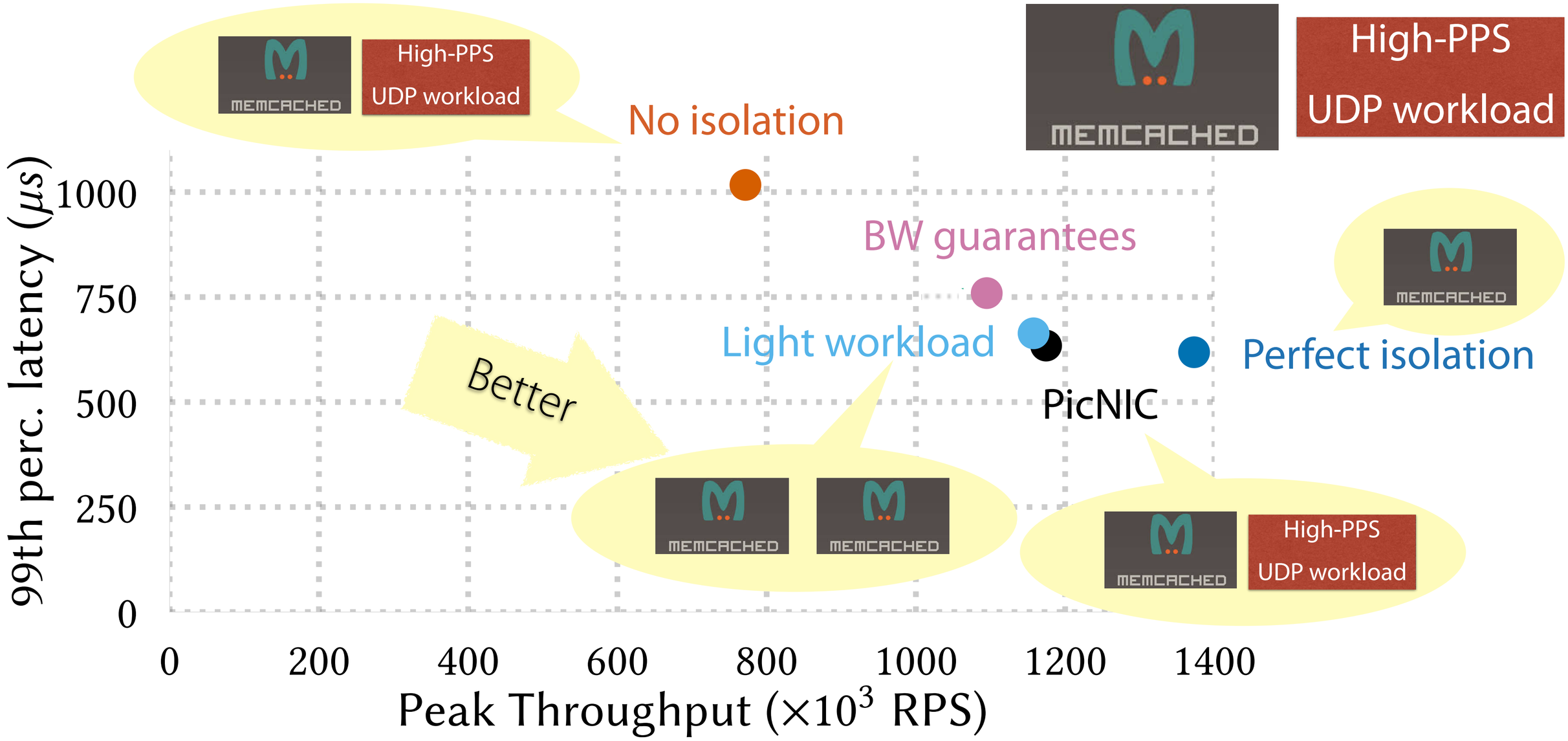
Low external load: 200 kRPS



Latency and Throughput



Latency and Throughput

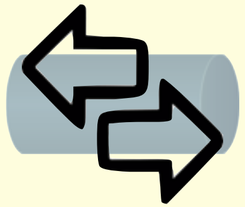
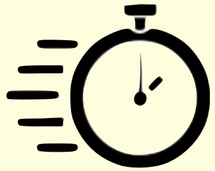


Takeaways

Isolation breakdown in end-host **virtualization stacks**: a key contributor to unpredictable performance

Predictable Virtualized NICs: enrich the behavior of virtual NICs with SLOs

- **SLO-based sharing**
- **complete back-pressure**

Future directions

- Co-operate with mechanisms for fabric isolation
- NIC design and HW offloads

Thank you!

PicNIC: Predictable Virtualized NIC

Praveen Kumar, Nandita Dukkipati, Nathan Lewis, Yi Cui, Yaogong Wang, Chonggang Li, Valas Valancius, Jake Adriaens, Steve Gribble, Nate Foster, Amin Vahdat

BEIJING
SIGCOMM
2019