
GOTC 2023
全球开源技术峰会
THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE
OPEN SOURCE, INTO THE FUTURE
「eBPF」专场
本期议题：使用 eBPF 代替 iptables 实现服务网格加速
刘齐均 2023年05月28日

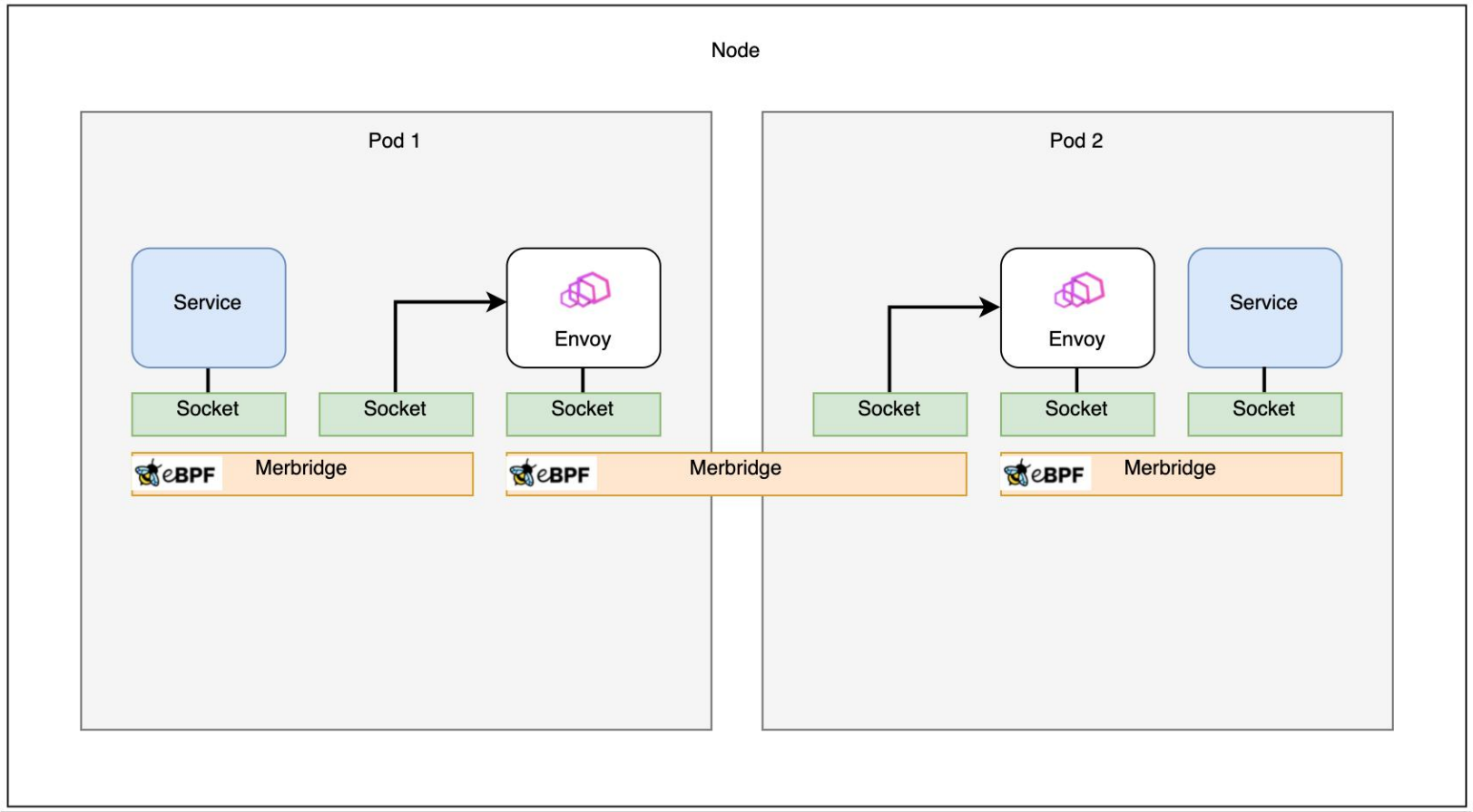
目录

Merbridge 项目介绍

Merbridge 是什么？

Merbridge 在服务网格中使用 eBPF 技术代替 iptables，实现流量拦截。

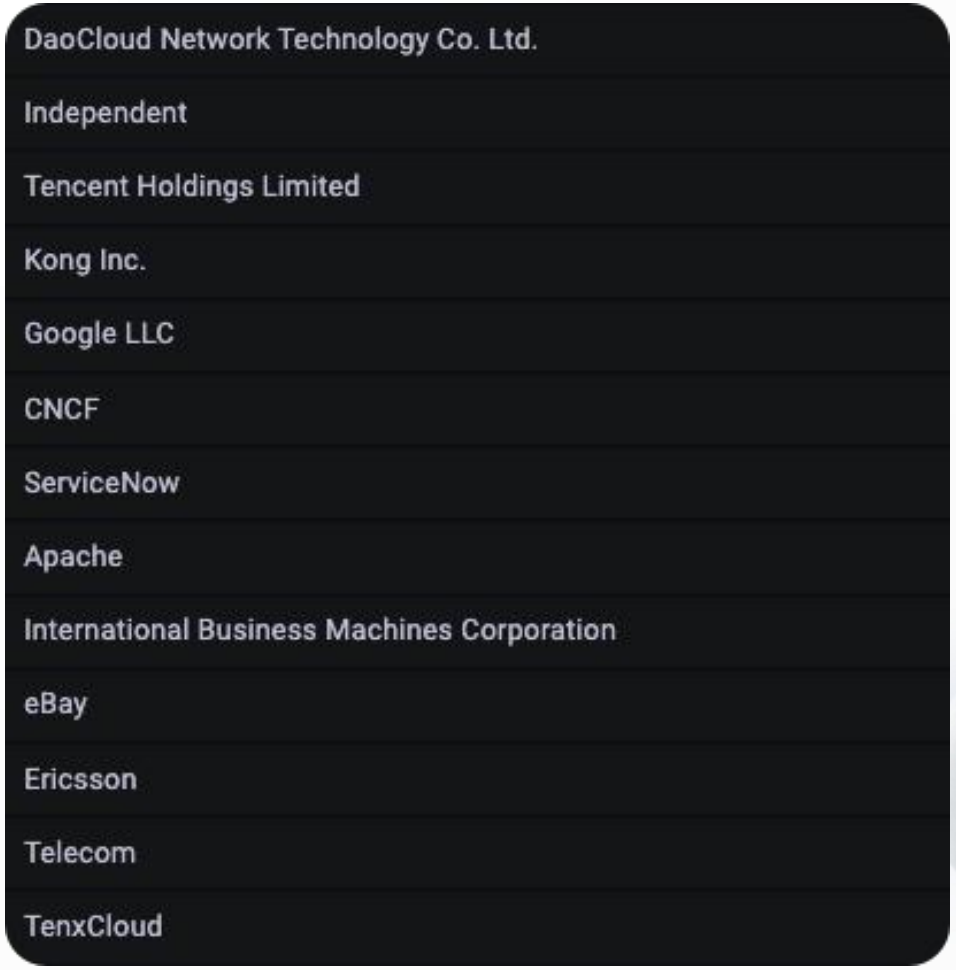
借助 eBPF 和 msg_redirect 技术，Merbridge 可以提高 Sidecar 和应用之间的传输速度，降低延迟。



Merbridge 项目介绍

目前生态

- 支持 Istio & Ambient 模式
- 支持 Kuma mesh
- 支持 Open Service Mesh
- 支持 Linkerd2

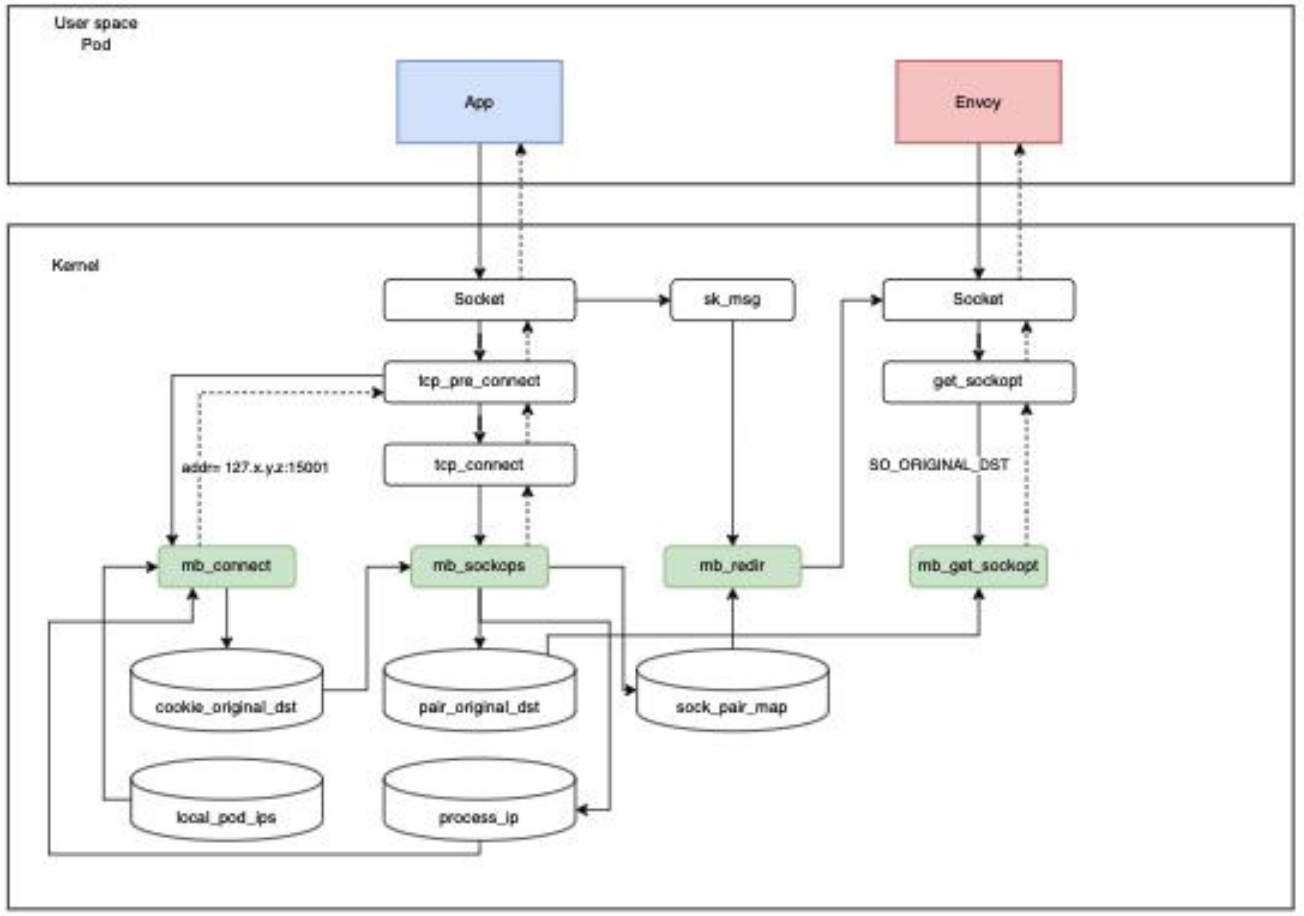


来自：devstats

实现原理

如何实现流量拦截

- 什么流量需要拦截？
  - App 发出的流量（通过 UID 判断）
  - 拦截到哪里？
- 怎么拦截？
  - 通过 connect 程序修改发起连接的目的地址为 127.0.0.1:15001
- 如何获取原地址？
  - 通过 get_sockopt 程序返回原始目的地址。



实现原理

如何加速？

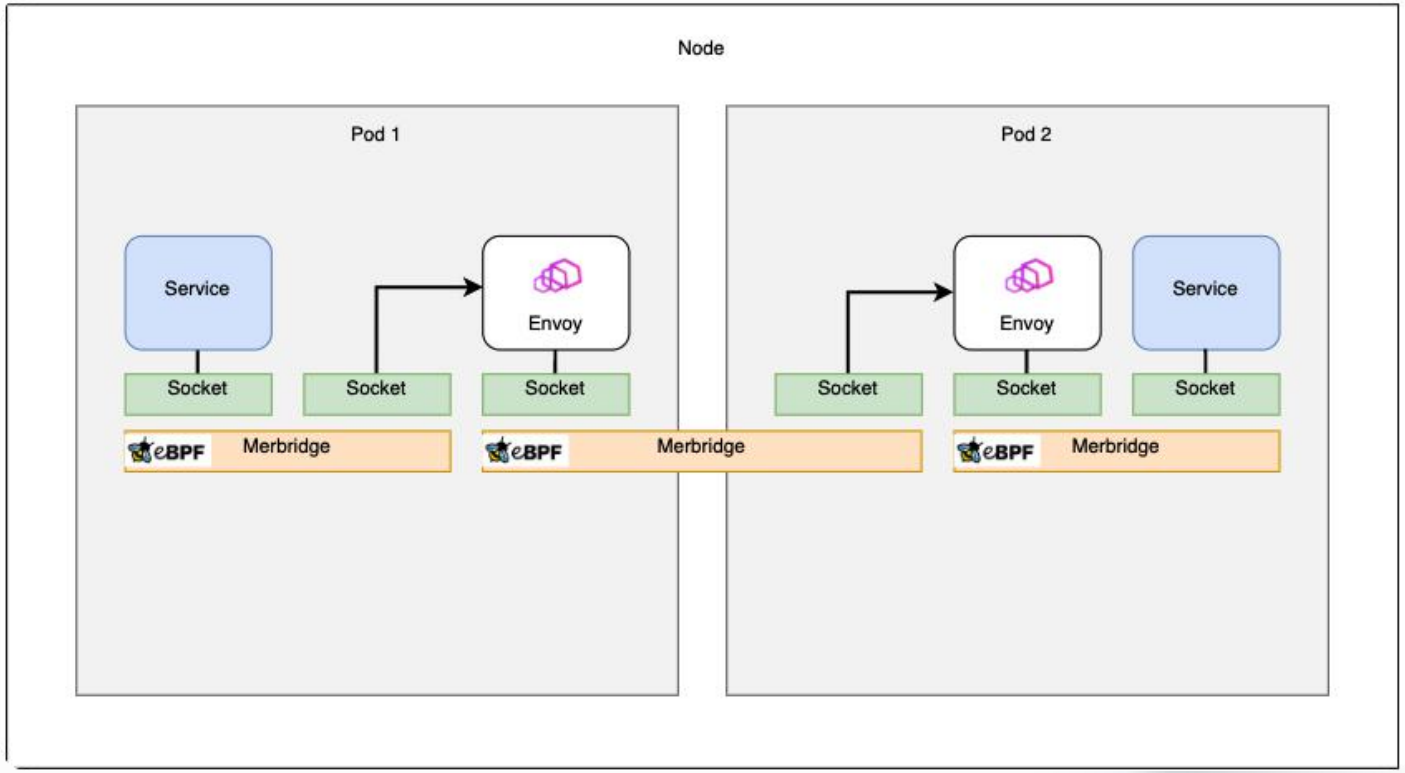
eBPF 提供了 bpf_msg_redirect_hash 相关的 helper 函数，可以将主机上的两个 socket 传输路径进行短接，绕过内核态协议栈，从而加速进程间的访问。

Helper 函数依赖一个内核级别的 map 保存连接信息，需要确保每个连接在 map 中的 key 互不冲突。

在容器这种场景下，如果四元组冲突怎么办？

P1：127.0.0.1:56789 => 127.0.0.1:15001

P2：127.0.0.1:56789 => 127.0.0.1:15001



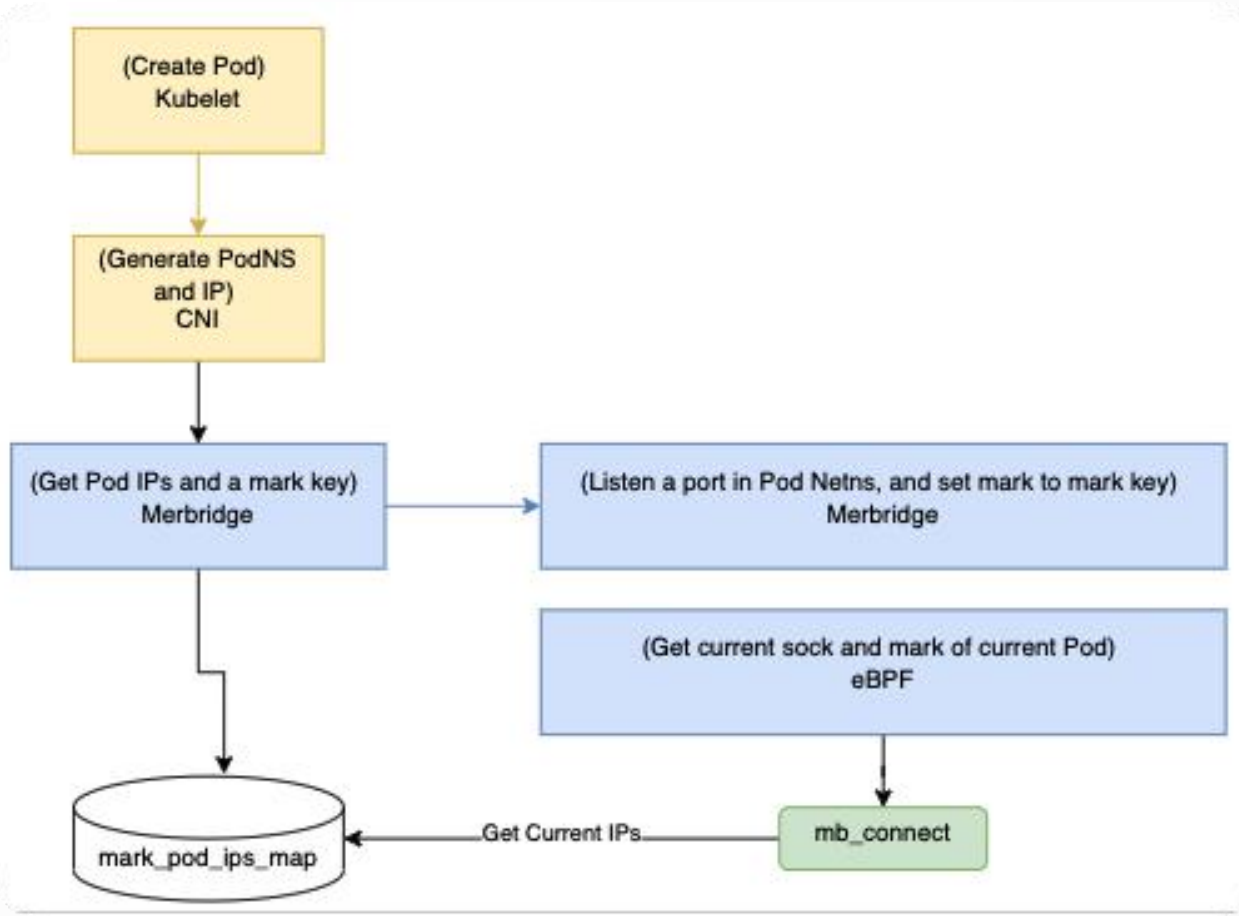
实现原理

通过 CNI 获取 Pod IP 解决四元组冲突问题

- Merbridge 会通过 CNI 插件的能力，在 Pod 创建的时候，将 Pod 的 IP 写入一个 Map，同时，在 Pod 的 Netns 监听一个特定的端口，并设置这个 socket 的 mark 为刚才 IP 的 Key。

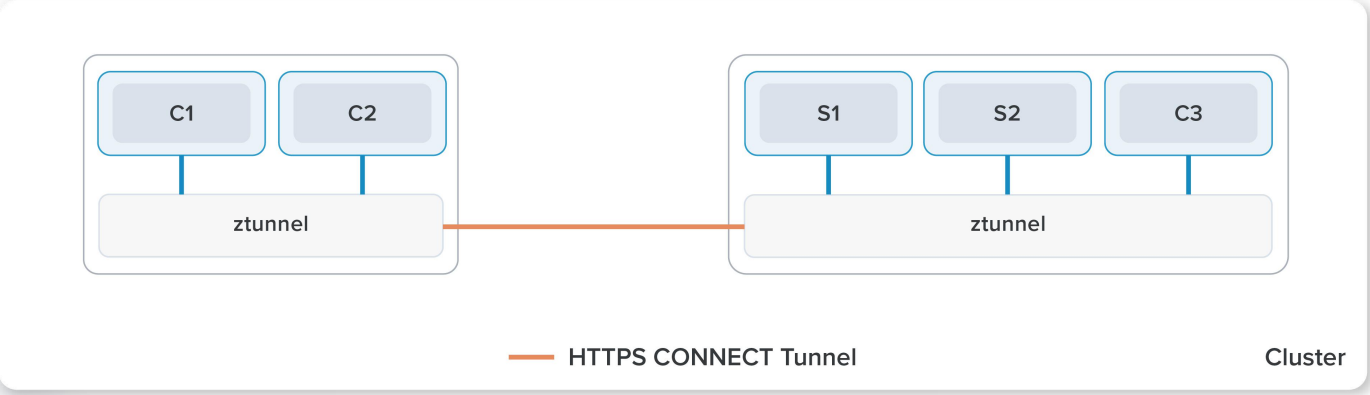
- 当 Merbridge 的 mb_connect 程序在处理请求时，通过获取当前 Netns 的指定端口 socket 的 mark，再根据 mark 从 map 中读取 IP，即可获得当前 IP。

- 通过将 source ip 替换成 PodIP 即可解决四元组冲突问题。



实现原理

Ambient 模式支持难题



- Ambient 模式下，加入/移除网格将更加灵活，CNI 模式将不再适用（CNI 只会在 Pod 创建的时候生效，Ambient 模式允许通过修改 annotation 的方式将 Pod 纳入网格）。
- Container 出口流量拦截的地址将不再是 127.0.0.1:15001，而需要转发到当前节点 ztunnel 实例。
- 由于 Pod 中不存在 Sidecar，之前通过判断是否监听 15001 端口等方案来确定是否需要拦截的方案将不再生效。

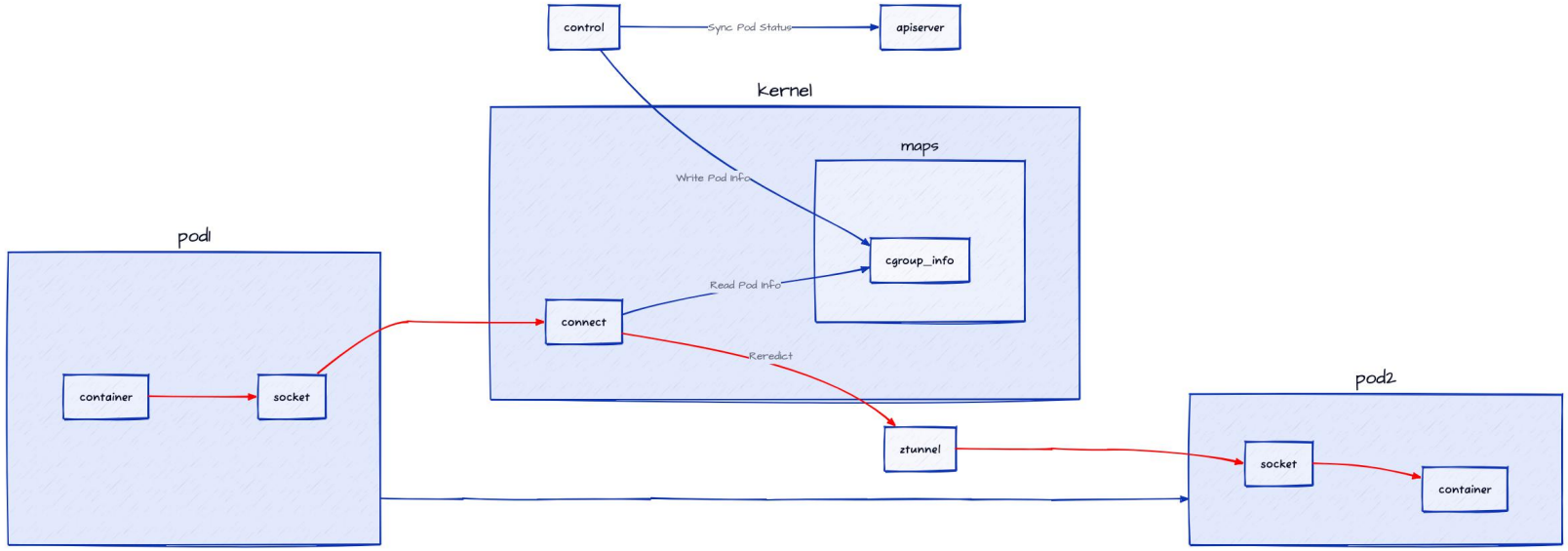
实现原理

Ambient 模式支持方案

1. 通过 eBPF 观察进程创建事件，在用户态关联 cgroup id 和 Pod IP。同时，将 Pod 信息等进行储存。

2. 在 eBPF 程序中，通过当前进程的 cgroup id 查找当前进程的 Pod IP。（一个容器中的进程应该具有相同的 cgroup id，且不会变更）

3. 如果是 Ambient 模式的 Pod 发出的流量，将流量转发到 ztunnel。

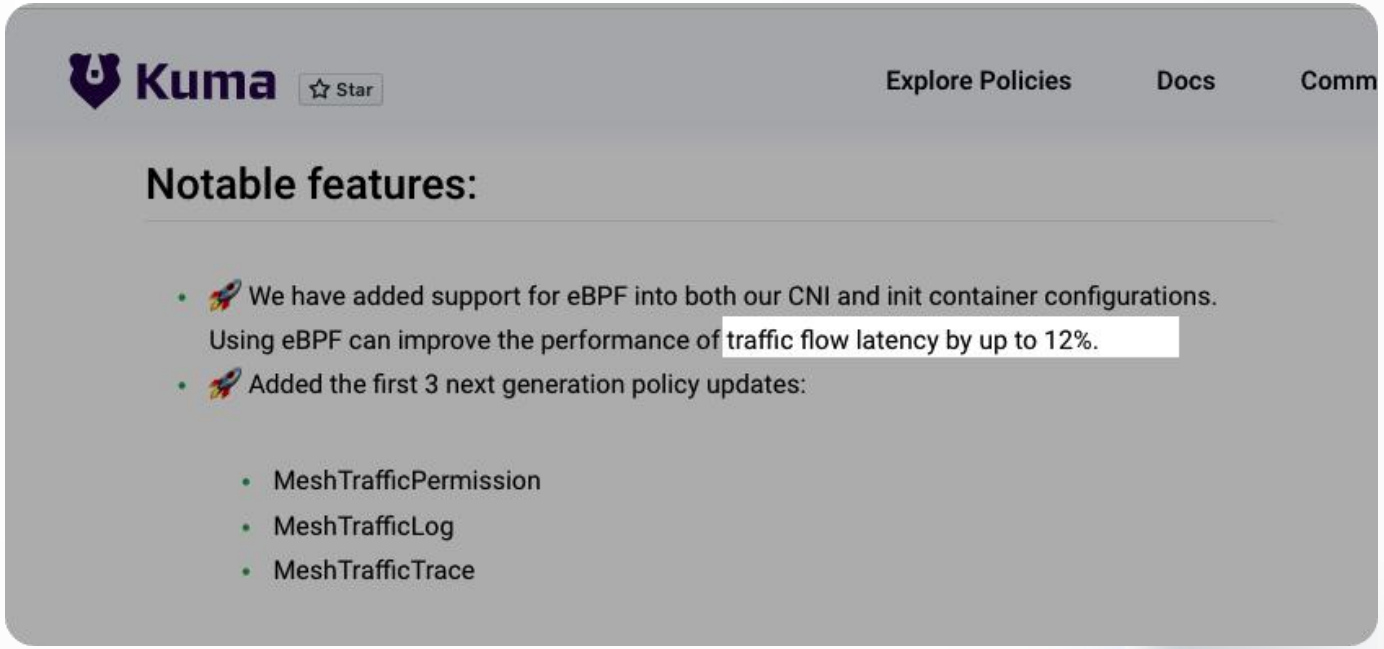


实现效果

一般情况下，可以实现 5-12% 的延迟提升。

可应对高并发场景。

无其它性能损耗或要求。



未来展望

持续性能测试；

更低的内核版本要求；

跨节点加速；

双栈支持；

Ambient 生产就绪；

……

社区

- 网址：https://merbridge.io/zh/
- 项目地址：
  https://github.com/merbridge/merbridge
- Slack 交流：
  https://join.slack.com/t/merbridge/shared_invite/zt-11uc3z0w7-DMyv42eQ6s5YUxO5mZ5hwQ

GOTC

THANKS

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE