《Istio大咖说》第2期

从微服务架构到 Istio——架构升级实践分享

主持人：宋净超（Tetrate）　　嘉宾：潘天颖（小电科技）

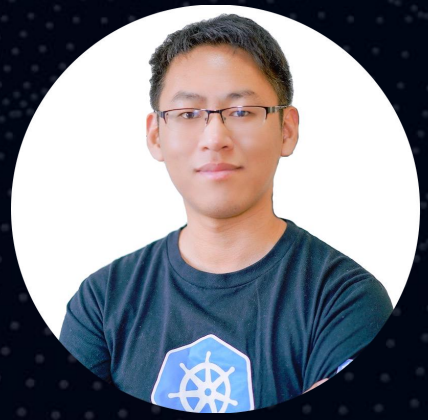
6月2日　晚8:00 – 9:00

联合主办方：tetrate　云原生社区 Cloud Native Community　ServiceMesher


tetrate
THE ENTERPRISE SERVICE MESH COMPANY

Istio 大咖说第 2 期

tetrate



宋净超（Jimmy Song）

Tetrate 布道师、云原生社区创始人
https://jimmysong.io

主持人

潘天颖

云原生爱好者、小电科技工程师兼
Istio布道者、Kubernetes contributor
Apache Committer

嘉宾

Agenda

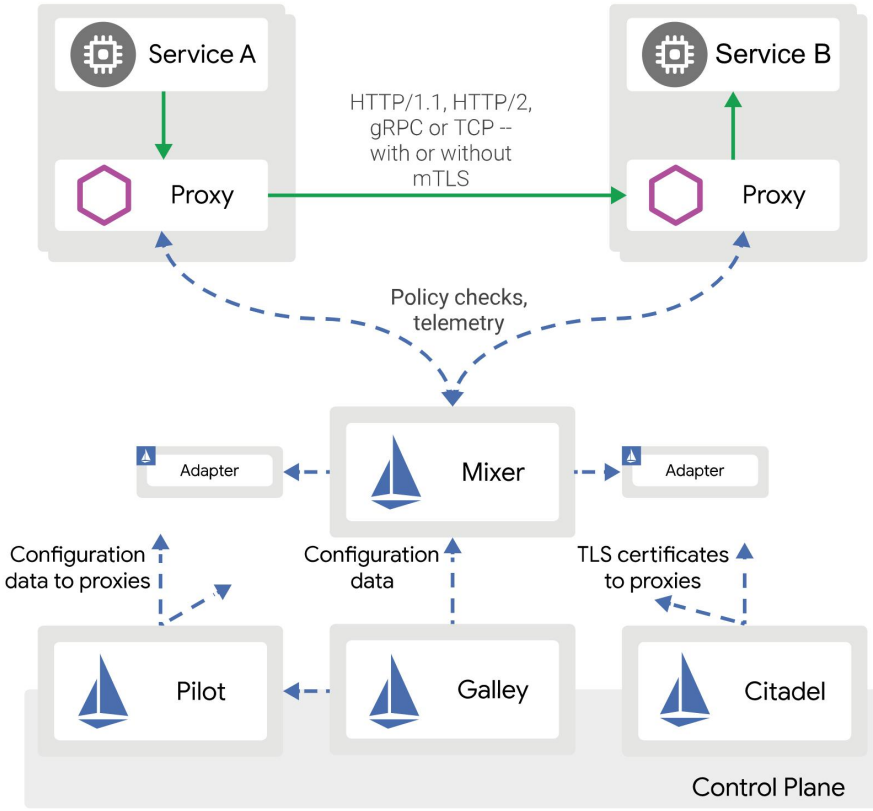
Istio现状

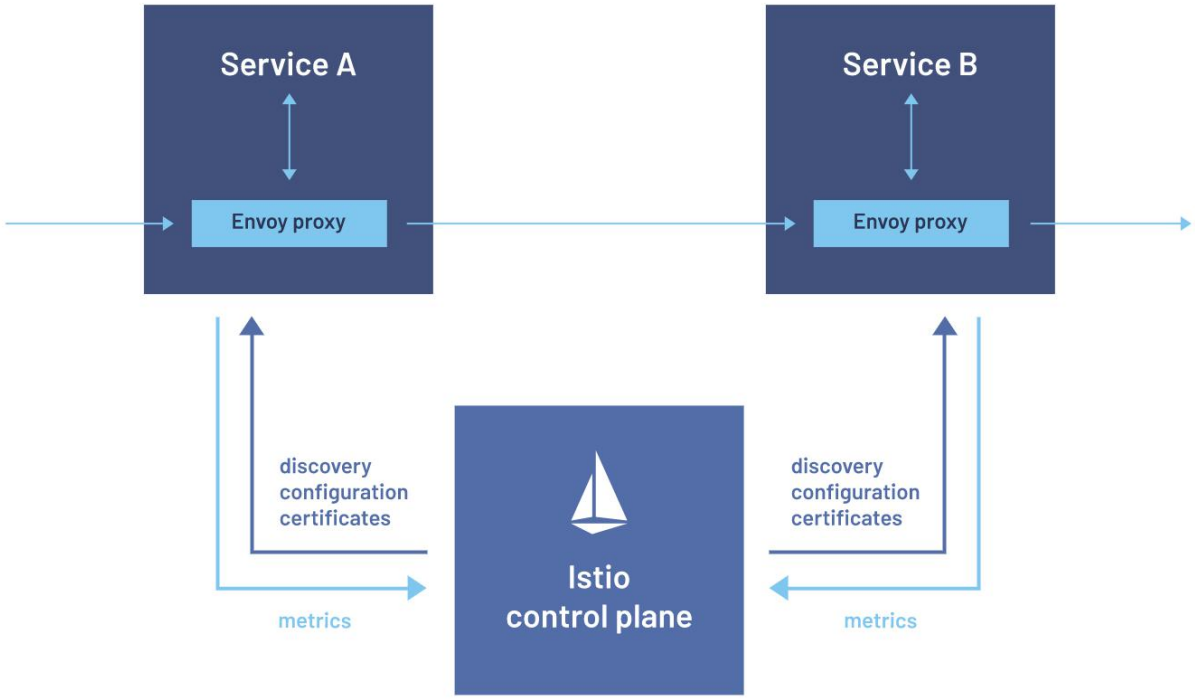
落地Istio何时利大于弊

小电落地Istio完整实践

痛点issue分析和改进

Istio现状


Service A
Service B
Proxy
HTTP/1.1, HTTP/2,
gRPC or TCP --
with or without
mTLS
Proxy
Policy checks,
telemetry
Adapter
Mixer
Adapter
Configuration
data to proxies
Configuration
data
TLS certificates
to proxies
Pilot
Galley
Citadel
Control Plane


Service A
Envoy proxy
Service B
Envoy proxy
discovery
configuration
certificates
metrics
Istio
control plane
discovery
configuration
certificates
metrics

Istio现状

tetrate

Performance summary for Istio 1.10

The Istio load tests mesh consists of **1000** services and **2000** sidecars with 70,000 mesh-wide requests per second. After running the tests using Istio 1.10, we get the following results:

- The Envoy proxy uses **0.35 vCPU** and **40 MB memory** per 1000 requests per second going through the proxy.
- Istiod uses **1 vCPU** and 1.5 GB of memory.
- The Envoy proxy adds 2.65 ms to the 90th percentile latency.

Agenda

tetrate

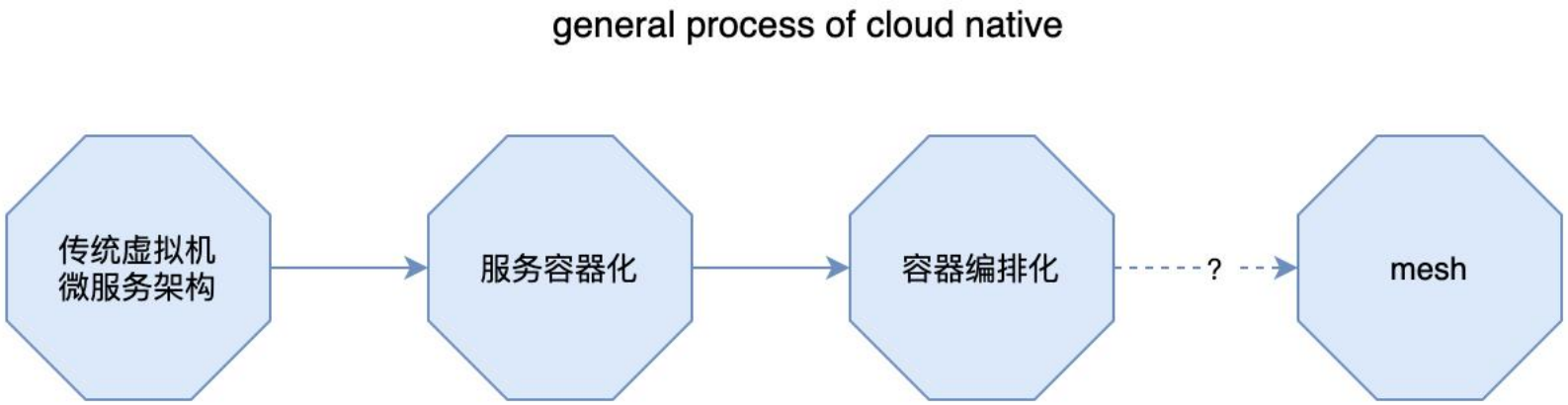
Istio 现状

落地Istio何时利大于弊

小电Istio完整落地实践

痛点Issue分析与改进

落地Istio何时利大于弊



decoupling ≠ inductionless

落地Istio何时利大于弊

tetrate

- 云原生基础（k8s)
- rpc协议
- 需求明确
- 性能评估
- 专业维护团队

Agenda

Istio 现状

落地Istio何时利大于弊

小电Istio完整落地实践

痛点Issue分析与改进

实践

tetrate

背景

- 90%+ 服务运行在自建k8s上，少量服务部署在虚拟机上。
- 服务类型以java服务为主，存在其他语言编码的服务。
- rpc体系主要由java为主的springcloud体系构成，协议为http 1.1。其他语言服务使用自开发的简易sdk加入rpc集群。
- 注册中心使用eureka。

tetrate

背景问题

- eureka的维护困难。

  社区难以继续支持Eureka的维护，Eureka官方宣布2.x不再开源。原架构下迁移注册中心风险过高。

- 服务健康感知延迟。

  eureka是客户端循环拉去活跃服务列表的机制，这让服务健康状态的感知延时过大，面对服务下线、node宕机都有很大的风险

- 新增调度功能时，微服务sdk版本升级管理困难。

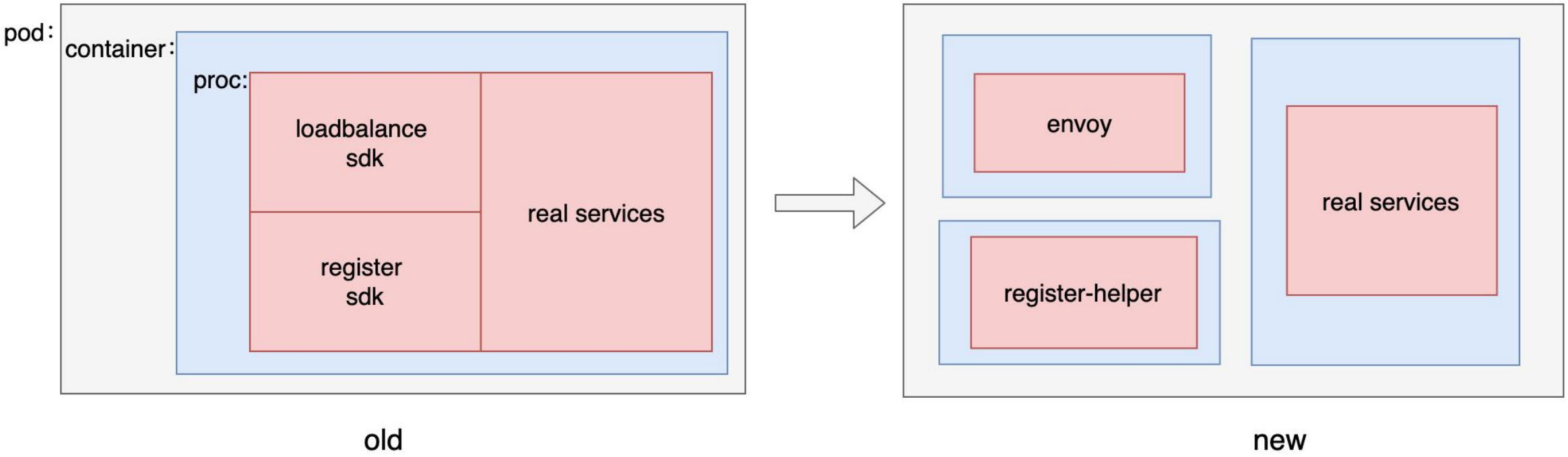
  由于历史原因，各开发使用的微服务sdk版本不一致，甚至其他语言中存在各种自实现的sdk，这让管理十分困难。微服务sdk和应用的耦合，让升级推广困难。

- 测试环境流控复杂，资源浪费严重。

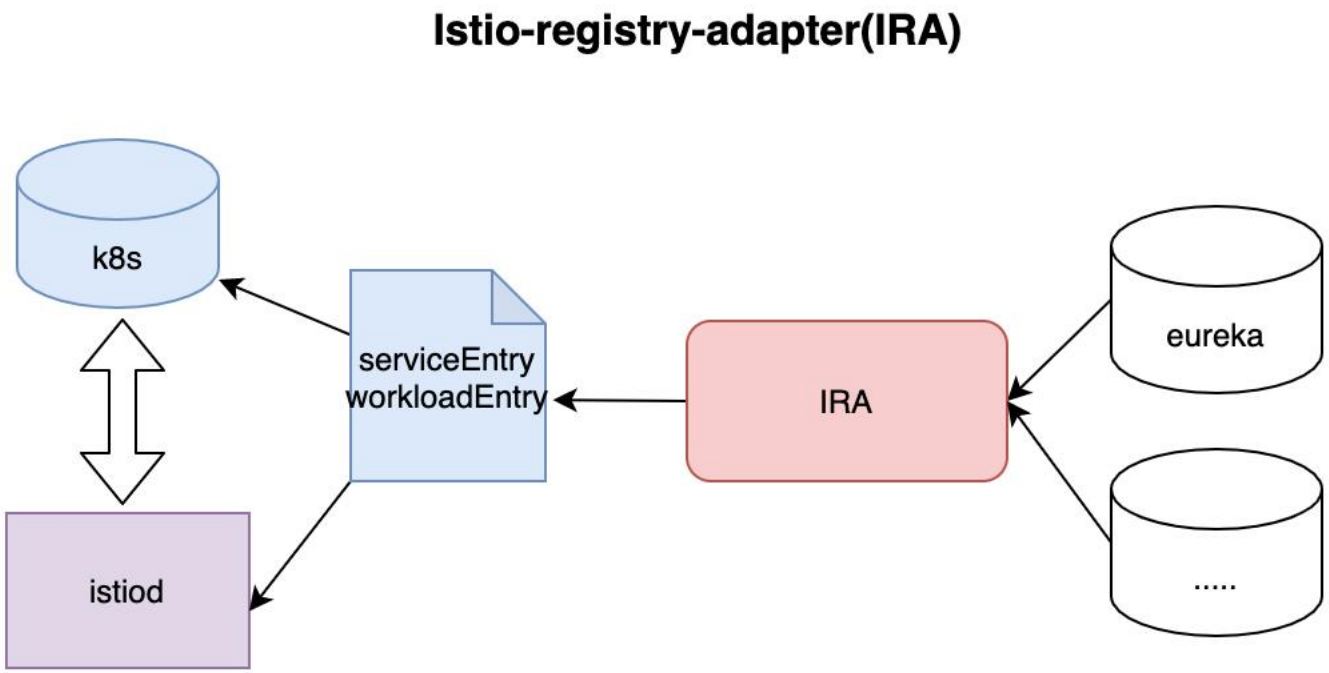
  由于当前微服务架构以及微服务sdk功能的限制，测试环境经常为了某个项目调试而起一堆关联pod，造成资源浪费。

实践

方案-应用改造

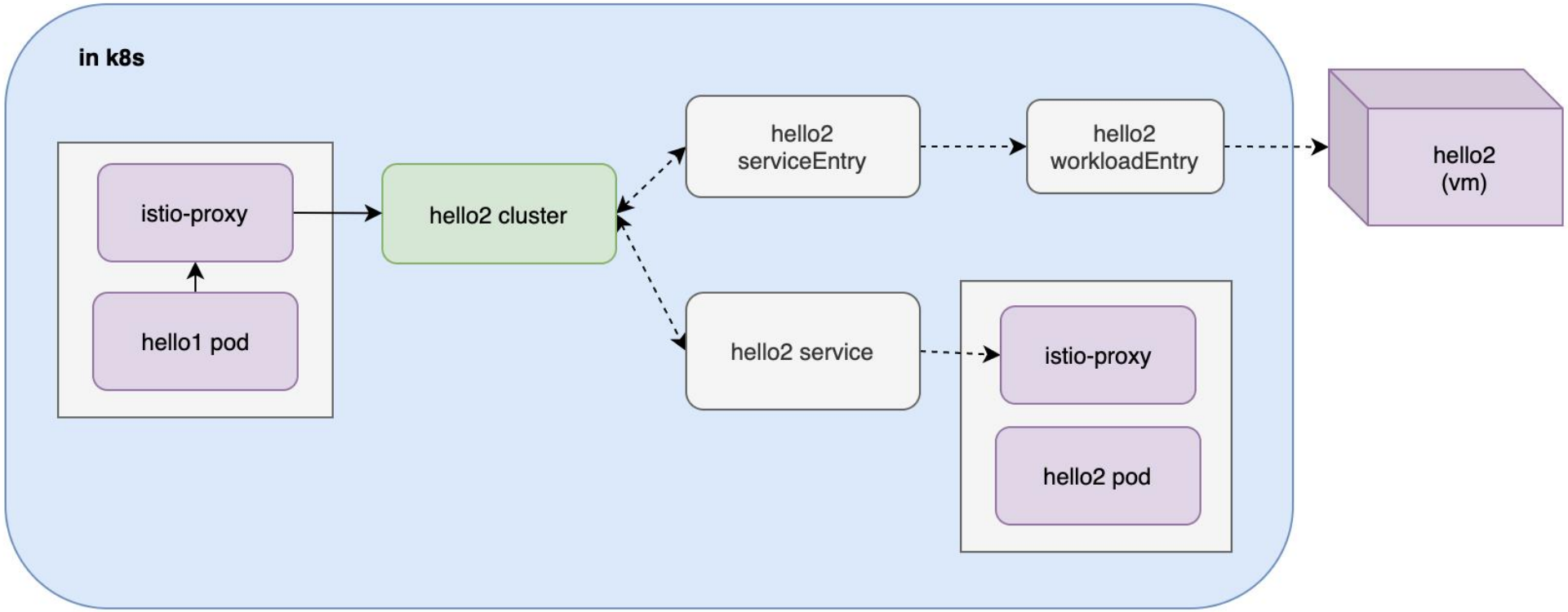

pod:
container:
proc:
loadbalance
sdk
register
sdk
real services
old
envoy
register-helper
real services
new

方案-过渡


Istio-registry-adapter(IRA)
k8s
serviceEntry
workloadEntry
istiod
IRA
eureka
.....

tetrate

方案-过渡



hello1 request hello2

实践

tetrate

方案-Istio改进

- 单集群多租户加强
- access log采集优化
- wasm插件、Istio各种配置的运维管理
- 机器权限收拢

实践

方案-Istio改进

- **机器权限收拢:　采用Istio cni插件保持业务服务严格的主机权限**

```
*nat
:PREROUTING ACCEPT [0:0]
:INPUT ACCEPT [0:0]
:POSTROUTING ACCEPT [0:0]
:OUTPUT ACCEPT [0:0]
:ISTIO_REDIRECT - [0:0]
:ISTIO_IN_REDIRECT - [0:0]
:ISTIO_INBOUND - [0:0]
:ISTIO_OUTPUT - [0:0]
-A PREROUTING -p tcp -j ISTIO_INBOUND
-A OUTPUT -p tcp -j ISTIO_OUTPUT
-A ISTIO_REDIRECT -p tcp -j REDIRECT --to-ports 15001
-A ISTIO_IN_REDIRECT -p tcp -j REDIRECT --to-ports 15006
-A ISTIO_INBOUND -p tcp -m tcp --dport 22 -j RETURN
-A ISTIO_INBOUND -p tcp -m tcp --dport 15090 -j RETURN
-A ISTIO_INBOUND -p tcp -m tcp --dport 15021 -j RETURN
-A ISTIO_INBOUND -p tcp -j ISTIO_IN_REDIRECT
-A ISTIO_OUTPUT -s 127.0.0.6/32 -o lo -j RETURN
-A ISTIO_OUTPUT ! -d 127.0.0.1/32 -o lo -m owner --uid-owner 1337 -j ISTIO_IN_REDIRECT
-A ISTIO_OUTPUT -o lo -m owner ! --uid-owner 1337 -j RETURN
-A ISTIO_OUTPUT -m owner --uid-owner 1337 -j RETURN
-A ISTIO_OUTPUT ! -d 127.0.0.1/32 -o lo -m owner --gid-owner 1337 -j ISTIO_IN_REDIRECT
-A ISTIO_OUTPUT -o lo -m owner ! --gid-owner 1337 -j RETURN
-A ISTIO_OUTPUT -m owner --gid-owner 1337 -j RETURN
-A ISTIO_OUTPUT -d 127.0.0.1/32 -j RETURN
-A ISTIO_OUTPUT -d 100.0.0.0/8 -j RETURN
-A ISTIO_OUTPUT -j ISTIO_REDIRECT
COMMIT
```

```
*nat
:PREROUTING ACCEPT [0:0]
:INPUT ACCEPT [0:0]
:POSTROUTING ACCEPT [0:0]
:OUTPUT ACCEPT [0:0]
:ISTIO_REDIRECT - [0:0]
:ISTIO_IN_REDIRECT - [0:0]
:ISTIO_INBOUND - [0:0]
:ISTIO_OUTPUT - [0:0]
-A PREROUTING -p tcp -j ISTIO_INBOUND
-A OUTPUT -p tcp -j ISTIO_OUTPUT
-A ISTIO_REDIRECT -p tcp -j REDIRECT --to-ports 15001
-A ISTIO_IN_REDIRECT -p tcp -j REDIRECT --to-ports 15006
-A ISTIO_INBOUND -p tcp -m tcp --dport 22 -j RETURN
-A ISTIO_INBOUND -p tcp -m tcp --dport 15090 -j RETURN
-A ISTIO_INBOUND -p tcp -m tcp --dport 80 -j RETURN
-A ISTIO_INBOUND -p tcp -m tcp --dport 15021 -j RETURN
-A ISTIO_INBOUND -p tcp -j ISTIO_IN_REDIRECT
-A ISTIO_OUTPUT -p tcp -m tcp --dport 80 -j RETURN
-A ISTIO_OUTPUT -s 127.0.0.6/32 -o lo -j RETURN
-A ISTIO_OUTPUT ! -d 127.0.0.1/32 -o lo -m owner --uid-owner 1000610001 -j
ISTIO_IN_REDIRECT
-A ISTIO_OUTPUT -o lo -m owner ! --uid-owner 1000610001 -j RETURN
-A ISTIO_OUTPUT -m owner --uid-owner 1000610001 -j RETURN
-A ISTIO_OUTPUT ! -d 127.0.0.1/32 -o lo -m owner --gid-owner 1000610001 -j
ISTIO_IN_REDIRECT
-A ISTIO_OUTPUT -o lo -m owner ! --gid-owner 1000610001 -j RETURN
-A ISTIO_OUTPUT -m owner --gid-owner 1000610001 -j RETURN
-A ISTIO_OUTPUT -d 127.0.0.1/32 -j RETURN
-A ISTIO_OUTPUT -j ISTIO_REDIRECT
COMMIT
```

Agenda

Istio 现状

落地Istio何时利大于弊

小电Istio完整落地实践

痛点issue分析与改进

实践issue

tetrate

memory of envoy

主要原因：

- 配置全量配置
- envoy worker thread

解决手段：

- delta xds （短期内可能无法实现）
- Sidecar配置
- 合理规划业务组namespace布局

实践issue

protocol sniffing


```
"dynamic_listeners": [
 {
    "name": "0.0.0.0_8080",
    "active_state": {
     "listener": {
      "@type": "type.googleapis.com/envoy.api.v2.Listener",
      "name": "0.0.0.0_8080",
      "filter_chains": [
       {
        "filters": [
         {
          "name": "envoy.http_connection_manager",
          "typed_config": {
           "@type":
"type.googleapis.com/envoy.config.filter.network.http_connection_manager.v2.HttpConnectionManager",
            ...
          }
         }
        ]
       }
      ],
      "deprecated_v1": {
       "bind_to_port": false
      },
      "traffic_direction": "OUTBOUND"
     },
     "last_updated": "2021-05-26T11:36:14.746Z"
    }
   }
]
```



```
{
  "name": "0.0.0.0_8080",
  "active_state": {
   "version_info": "2021-05-26T12:18:06Z/6",
   "listener": {
    "@type": "type.googleapis.com/envoy.api.v2.Listener",
    "name": "0.0.0.0_8080",
    "filter_chains": [
     {
      "filter_chain_match": {
       "application_protocols": [
        "http/1.0",
        "http/1.1",
        "h2c"
       ]
      },
      "filters": [
       {
        "name": "envoy.http_connection_manager",
        "typed_config": {
         "@type": "type.googleapis.com/envoy.config.filter.network.http_connection_manager.v2.HttpConnectionManager",
          ...
       }...
      ]
     },
     {
      "filter_chain_match": {},
      "filters": [{
       "name": "envoy.tcp_proxy",
       "typed_config": {
        "@type": "type.googleapis.com/envoy.config.filter.network.tcp_proxy.v2.TcpProxy",
        ...
       },...],
      "metadata": {"name": "PassthroughFilterChain"}
     ],
     "listener_filters": [
      {"name": "envoy.listener.tls_inspector",...},
      {"name": "envoy.listener.http_inspector",...}
     ],
     "listener_filters_timeout": "5s",
     "traffic_direction": "OUTBOUND",
     "continue_on_listener_filters_timeout": true
    },
    "last_updated": "2021-05-26T12:18:07.047Z"
   }
  }
```

实践issue

tetrate

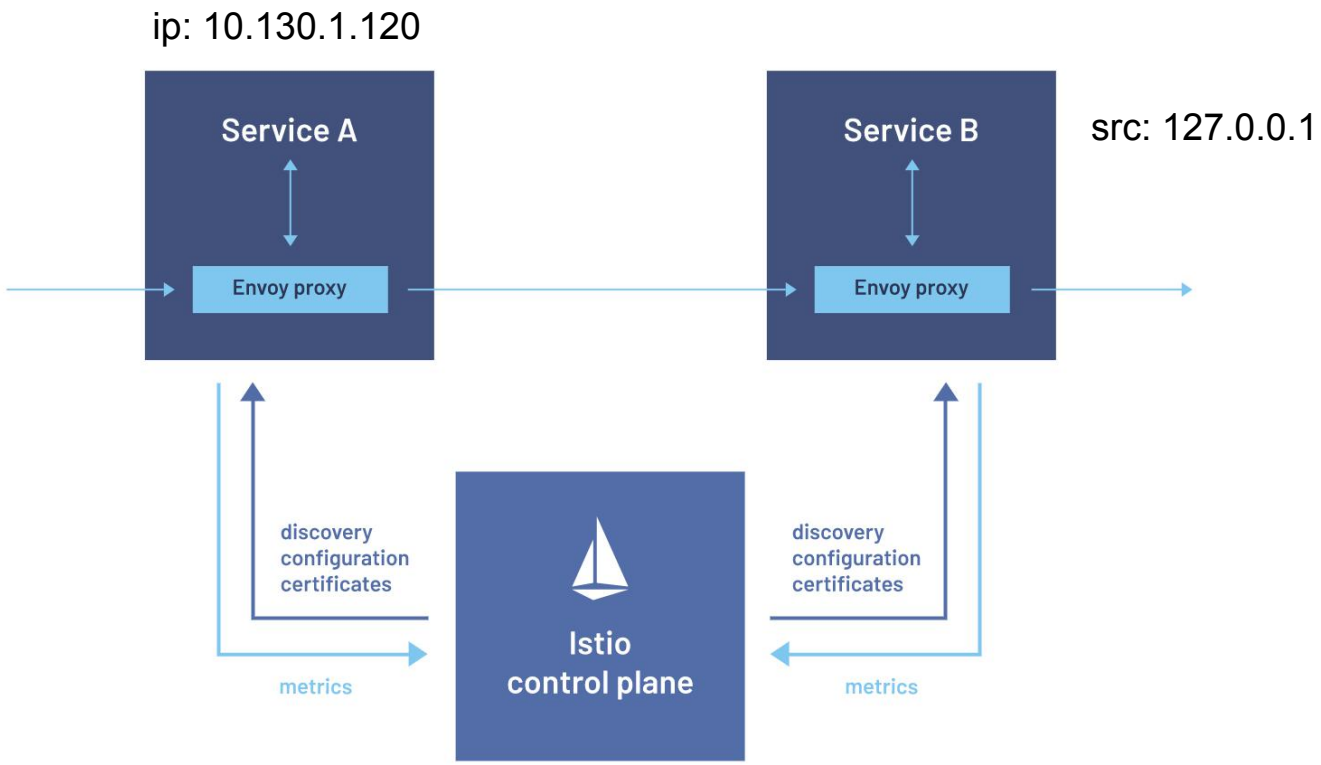
nonstandard protocol

- 不规范的bug流量
- 错用端口
- 明确不需要istio劫持的流量

iptable执行流量劫持时作区分。Istio提供
traffic.sidecar.istio.io/excludeOutboundPorts、
traffic.sidecar.istio.io/excludeOutboundIPRanges、
traffic.sidecar.istio.io/excludeInboundPorts等pod注解配置，用于指定某些特殊流量放行，不使用envoy拦截。同样适用于使用istio-cni代替initContainer的场景

实践issue

original source reservation



实践issue

tetrate

original source reservation

L4

interceptionMode : TPROXY （不建议）

通过增加original src filter 修改tcp src ip address为收到请求的ip

L7 (http)

```
apiVersion: networking.istio.io/v1alpha3
kind: EnvoyFilter
metadata:
  name: sidecar-add-xxf
spec:
  configPatches:
  - applyTo: NETWORK_FILTER
    match:
      listener:
        filterChain:
          filter:
            name: envoy.http_connection_manager
    patch:
      operation: MERGE
      value:
        typed_config:
          "@type": type.googleapis.com/envoy.config.filter.network.http_connection_manager.v2.HttpConnectionManager
          skip_xff_append: false
          use_remote_address: true
          xff_num_trusted_hops: 1
```

tetrate

Q&A

Tetrate 中国

云原生社区