服务可感知的零信任容器网络及其DPU卸载

向阳朝
神州数码
上海, 中国

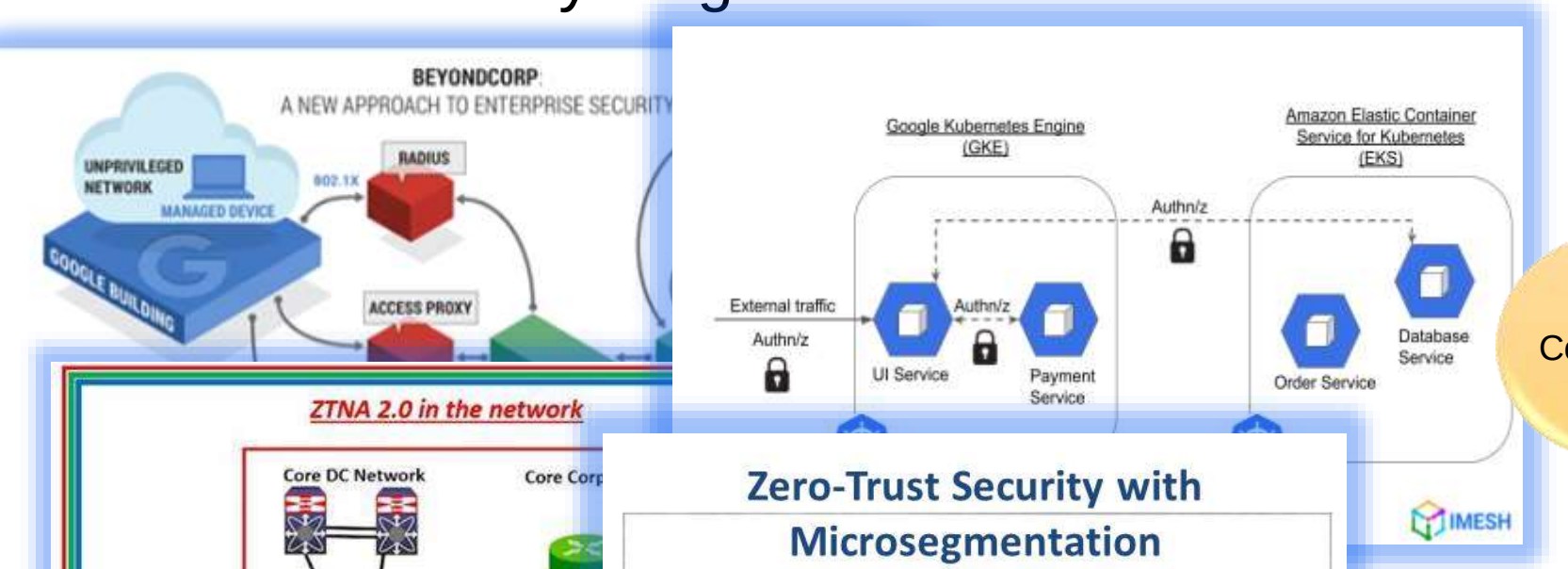
KubeCon + CloudNativeCon + Open Source Summit China 2023

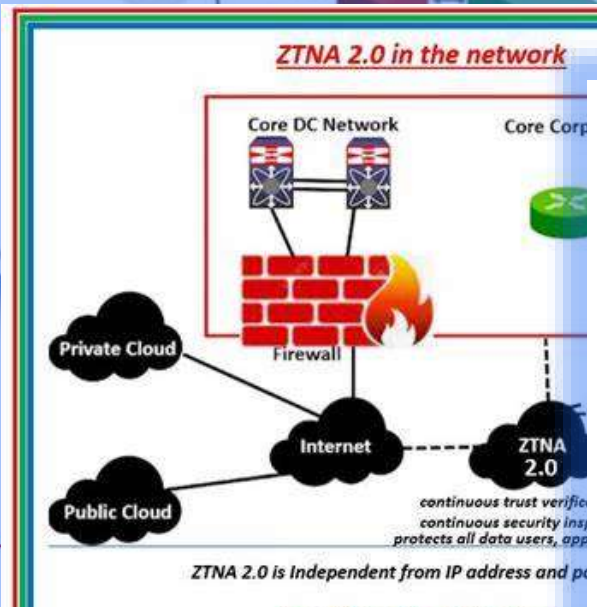
神州数码 Digital China

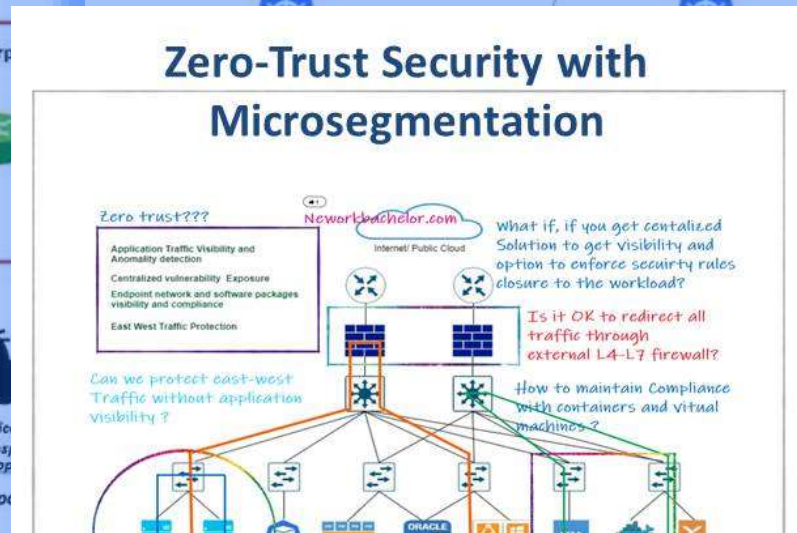
Security Level:

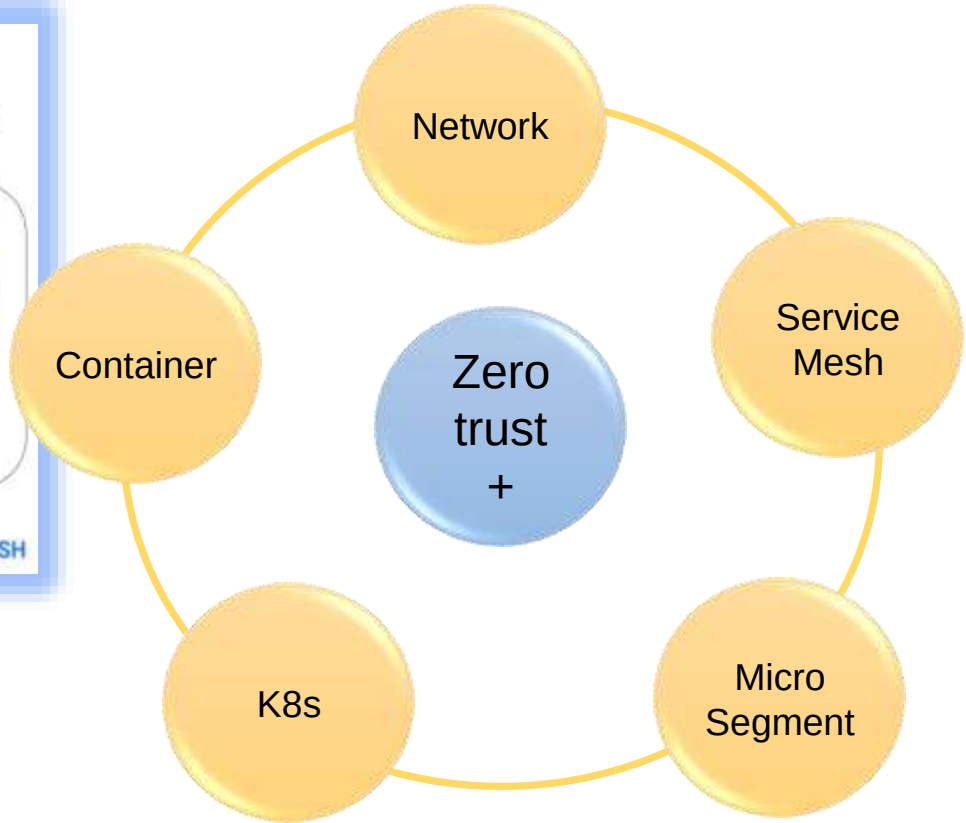
什么是零信任

- 不直接授权信任任何用户、设备、服务
- 口令、证书是不够的
- zero + everything









如何达到零信任

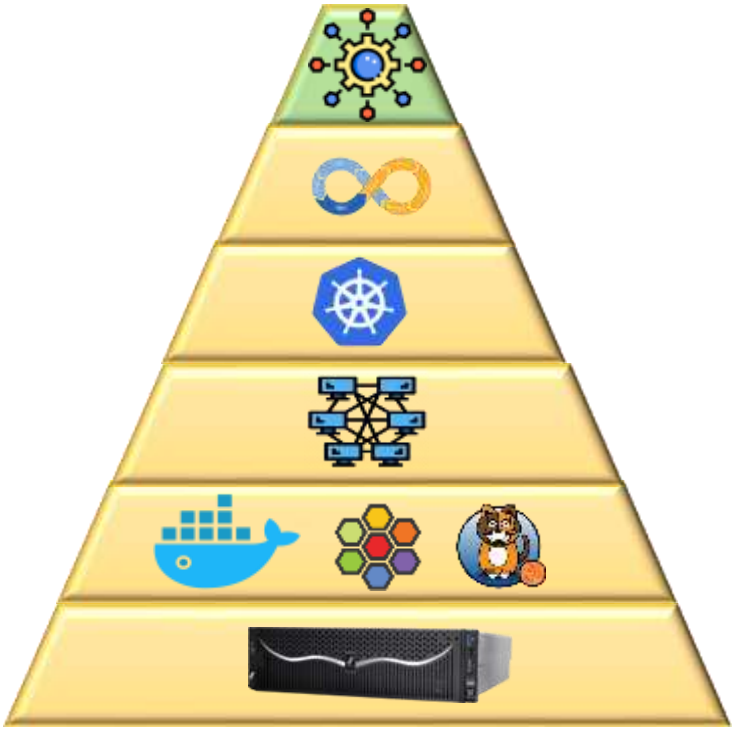
神州数码 Digital China

- 挑战
  - 身份泄露非常普遍
  - 东西横向移动攻击非常普遍
- 目标
  - 限制身份泄露的影响Limit the effect of identity theft
  - 限制横向移动攻击Limit the lateral movement attack
  - 可立即阻止基于任何实体的访问 (user, device, service, location)
- 行动—任何一项都不足以满足零信任架构
  - 持续验证: 行文，环境, AI/ML
  - 所有通信都需加密
  - 通信双方互验证
  - 证书和密钥轮换
  - 微分段 : 网络, 服务
- 完全的零信任也许无法达到
  - 性能，可用性，费用

云原生零信任架构

容器网络和服务网格提供了零信任的通信解决方案

- 云原生系统
  - 弹性、可管理、可观测的松耦合系统
- 云原生应用
  - 一组小的、独立的、松耦合服务
- 云原生基础架构
  - 容器、网络、调度、CI/CD
  - 服务网格



- 行动
  - 持续验证
  - 通信加密
  - 互认证
  - 证书和密钥轮换
  - 微分段

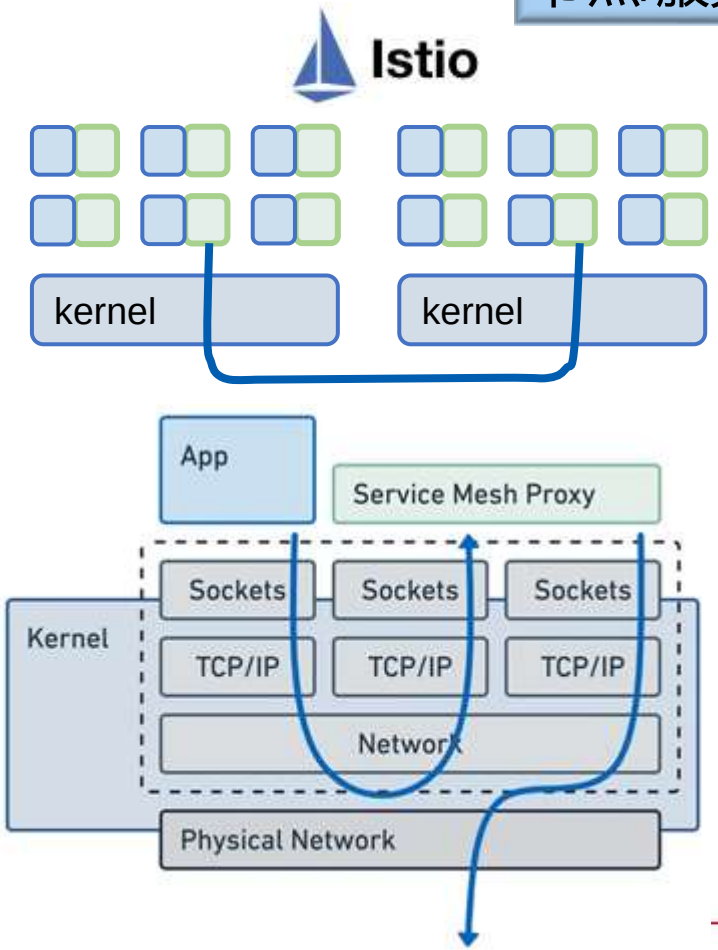
→

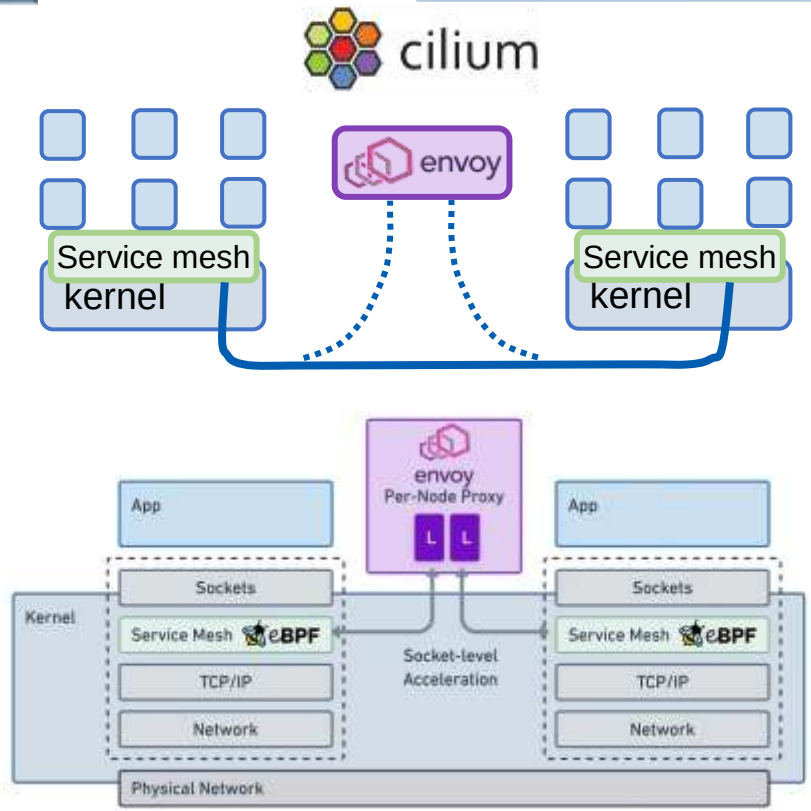
  - 遥测: 容器、网络、服务
  - 会话加密
  - 会话互认证
  - 证书和会话密钥轮换
  - 微服务分段

边车 & 无边车服务网格

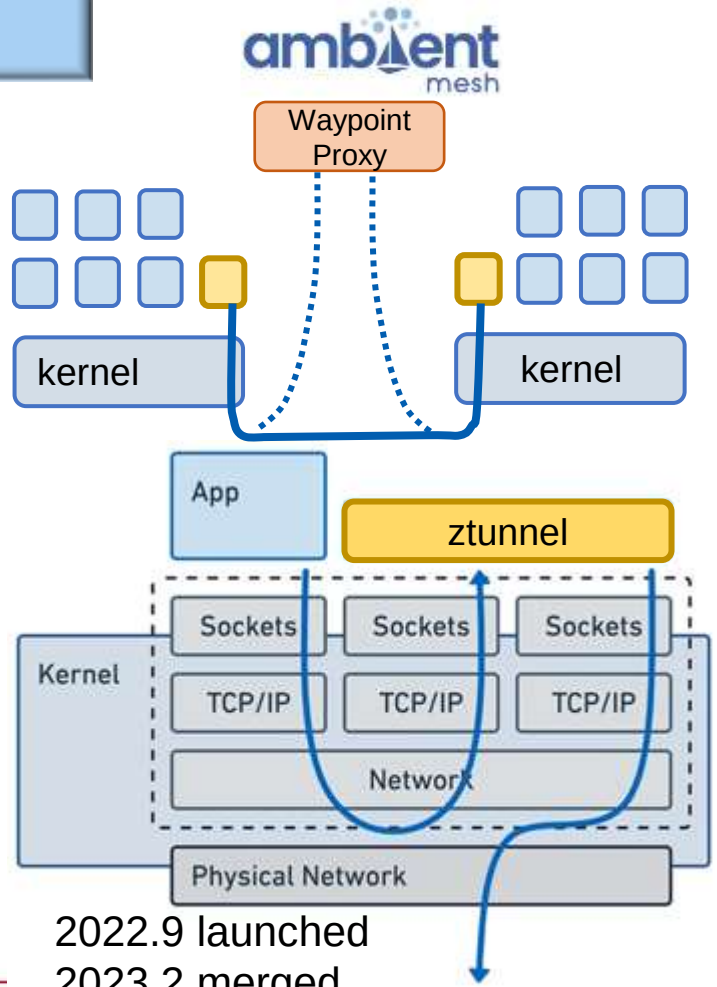
下称L4到CNI
减少会话代理
eBPF
节点/服务共享

减少侵入性
POD和服务基础架构结构
减少延迟
减少host资源占用





2022.7 v1.12



2022.9 launched
2023.2 merged

没有完美的解决方案

神州数码 Digital China

在性能、安全性、水平扩展性、安全性等方面平衡

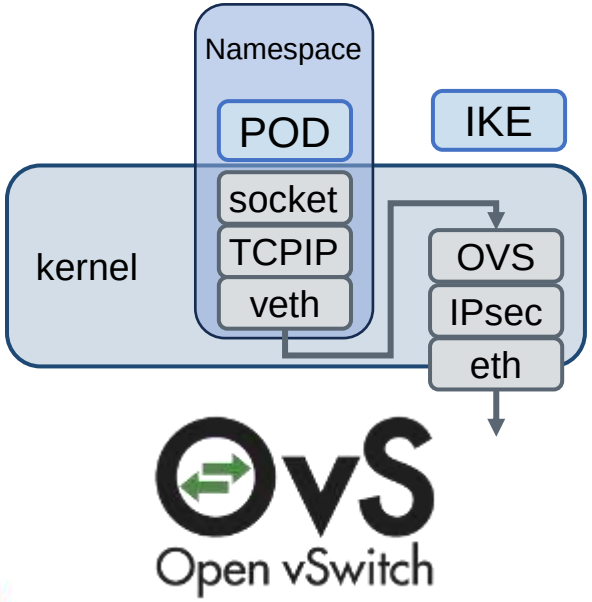
	Istio	cilium	ambient mesh
Latency	★★	★★★	★★
Host CPU usage	★★	★★★★	★★★★★
Host memory usage	★★	★★★★	★★★★★
POD invasion	★★	★★★★★	★★★★★
Kernel Req.	★★★★★	★	★★★★★
Istio compatibility	★★★★★	★★	★★★★
scalability	★★★★★	★★	★★★★
Service awareness	★★★★★	★★★	★★★
Breach radius	★★★★★ POD	★★ Node	★★ Node
vulnerability	★★ App, Envoy	★★★★ Envoy	★★★★ Ambient

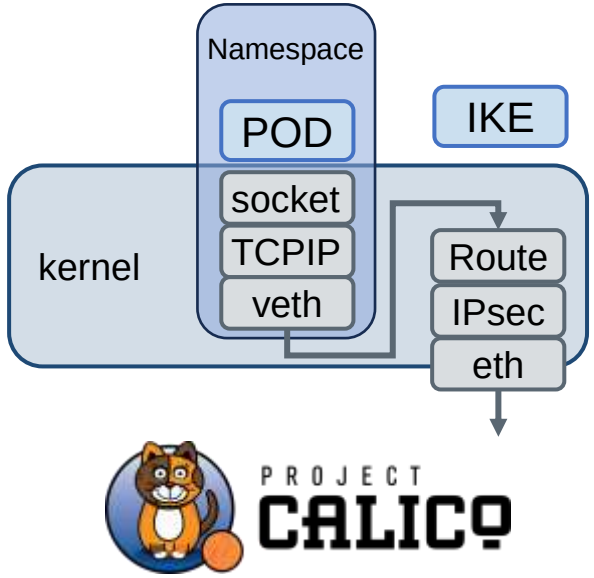
星星越多越好

零信任容器网络

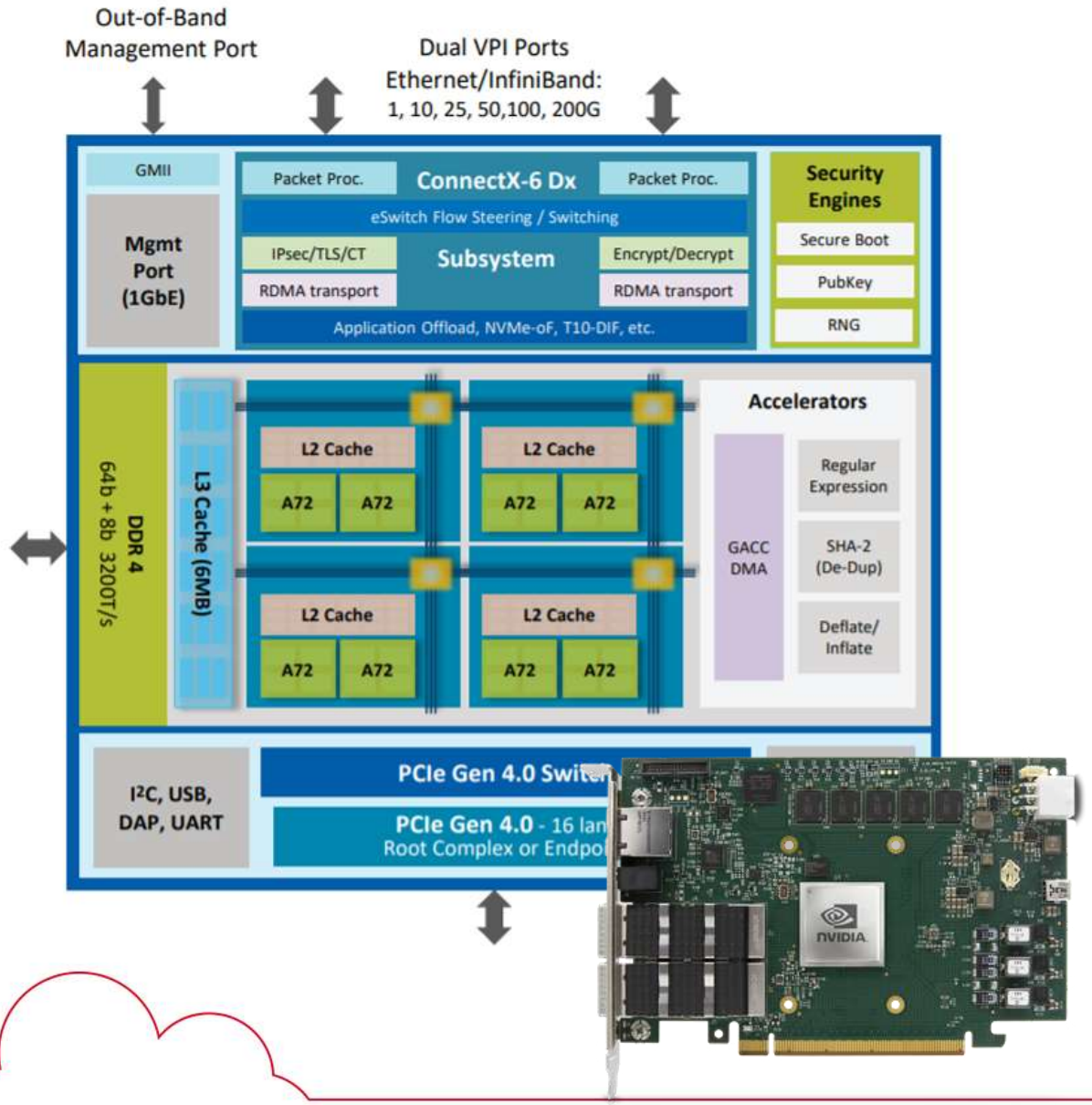
很多场景并没有(必要/可能)部署服务网格，但在加密的容器网络隧道中仍然需要感知、验证服务双方身份。

- 在节点间通过Ipsec/Wireguard加密隧道传输容器数据包
- 目前的标准开源实现中，在host的出接口IPsec隧道中不提供对容器数据的识别和加密
- 基于POD ip识别工作负载身份





DPU能提供什么能力



- Overlay网络加速
  - VXLAN, GENEVE, NVGRE
- 连接跟踪
- 数据流采样和统计
- IPsec/TLS数据传输加密
  - AES-GCM 128/256-bit key
- 公钥加速 (PKA)
  - RSA, Diffie-Hellman, DSA, ECC,EC-DSA, EC-DH

基于DPU的零信任云原生服务网格架构

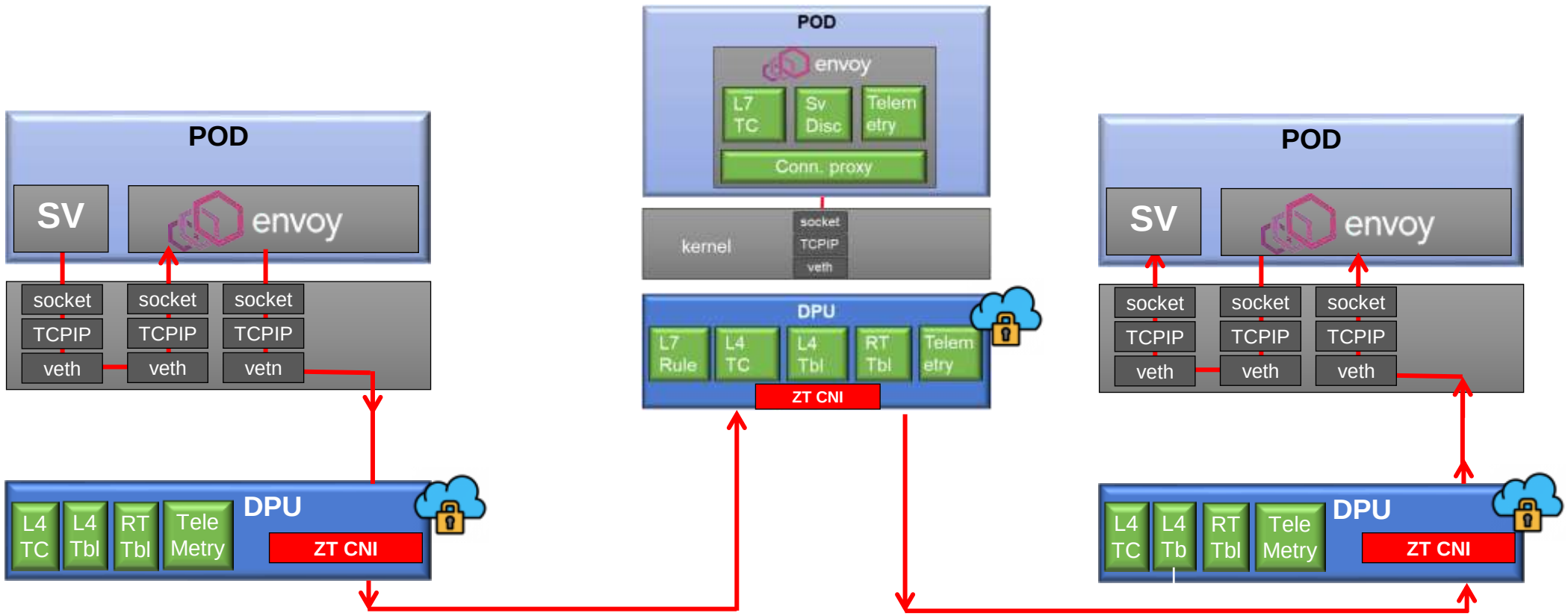
神州数码 Digital China

- 目标
  - 提高吞吐，降低延时
  - 与istio生态兼容
  - 保持或增强安全性
  - 支持无边车的场景
  - 服务与云原生基础架构解耦
- 挑战
  - 服务可感知
  - 安全破防影响半径

基于DPU的零信任云原生服务网格架构

神州数码 Digital China

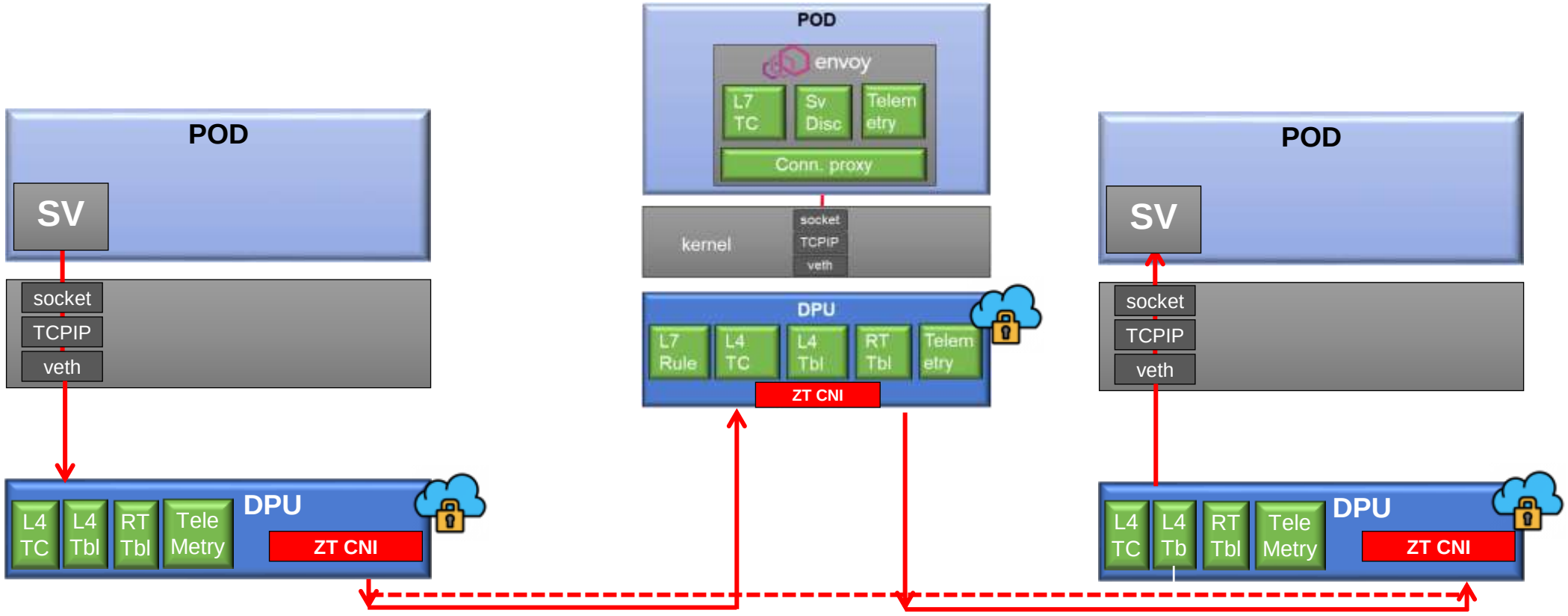
- 目标
  - 提高吞吐，降低延时
  - 与istio生态兼容
  - 保持或增强安全性
  - 支持无边车的场景
  - 服务与云原生基础架构解耦
- 挑战
  - 服务可感知
  - 安全破防影响半径



基于DPU的零信任云原生服务网格架构

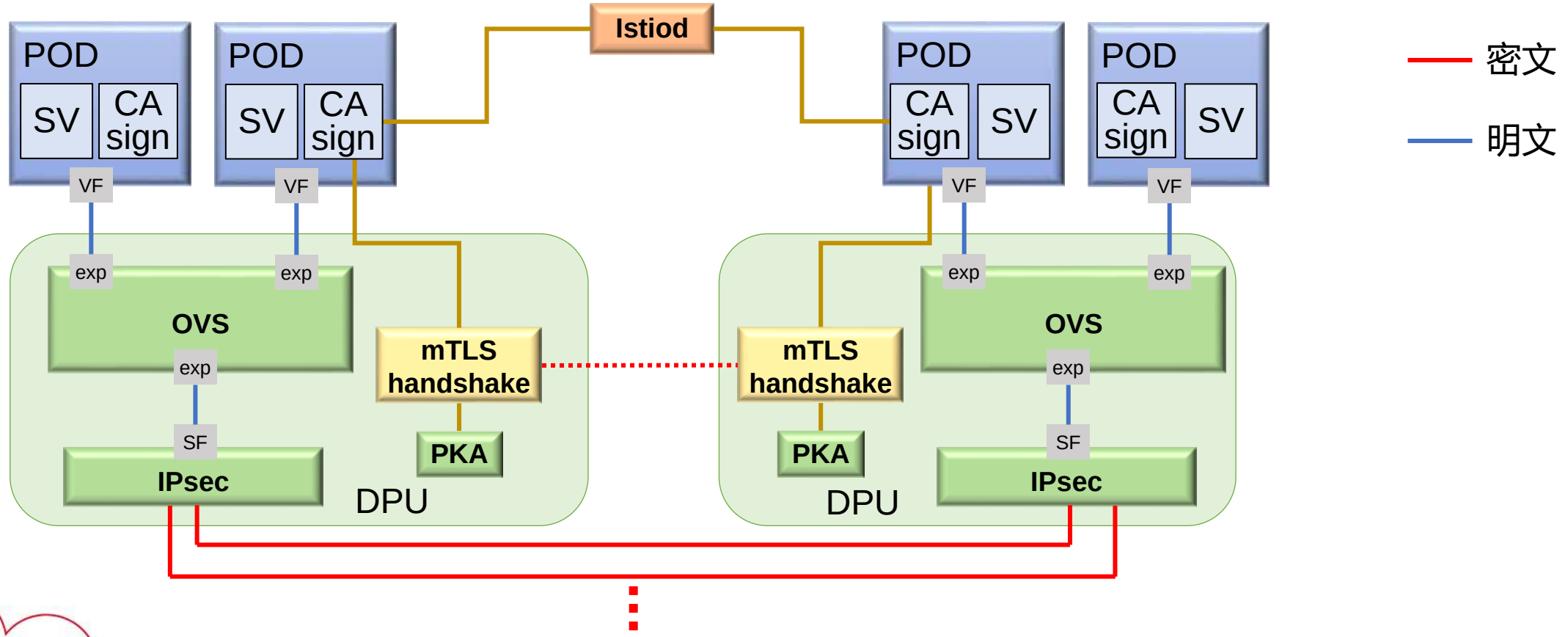
神州数码 Digital China

- 目标
  - 提高吞吐，降低延时
  - 与istio生态兼容
  - 保持或增强安全性
  - 支持无边车的场景
  - 服务与云原生基础架构解耦
- 挑战
  - 服务可感知
  - 安全破防影响半径



基于DPU的零信任容器网络架构

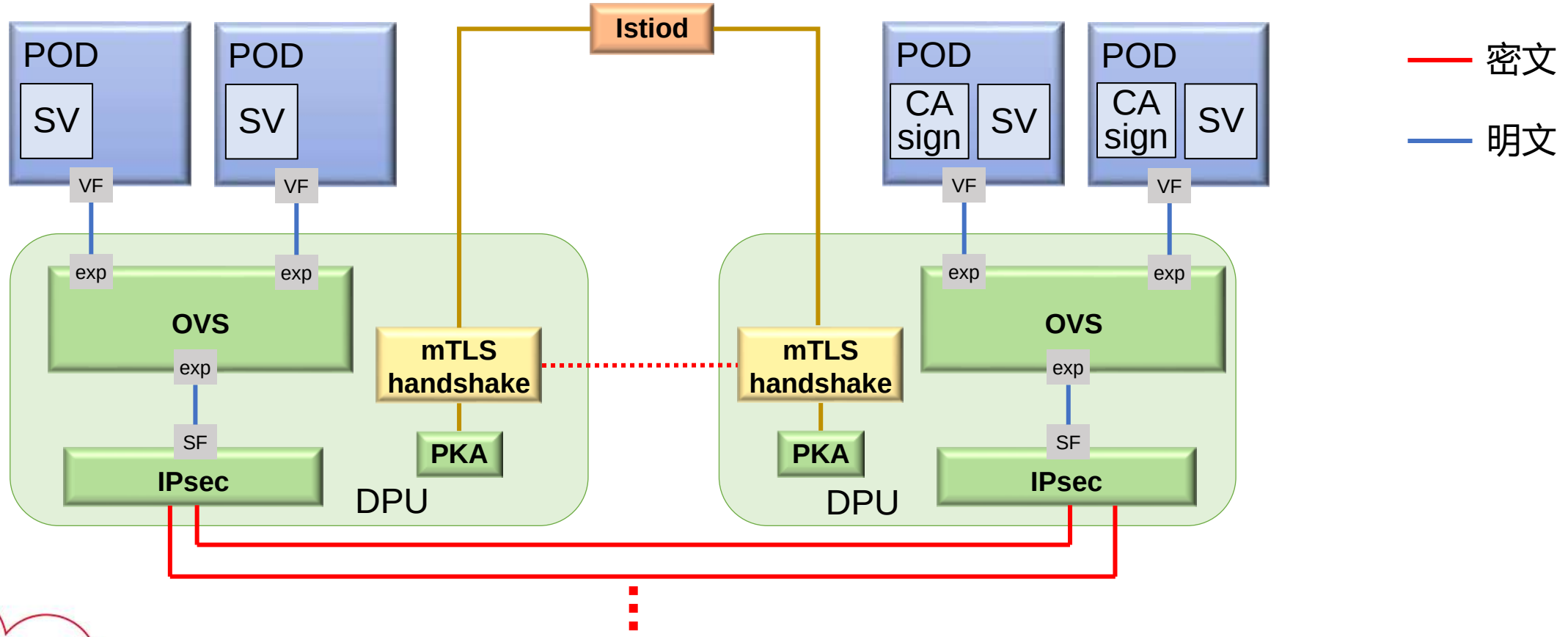
- 基于POD的IPsec隧道
- 基于mTLS的互认证和密钥协商
- 使用DPU metadata中的vfid作为POD身份标识
- 若扩大安全破防半径到node, 可极大加速mTLS性能



基于DPU的零信任容器网络架构

- 基于POD的IPsec隧道
- 基于mTLS的互认证和密钥协商

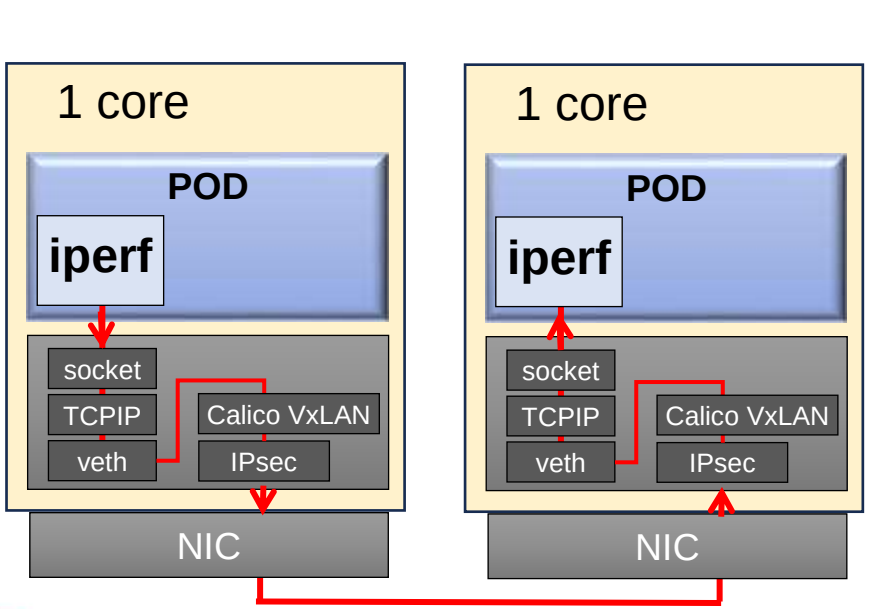
- 使用DPU metadata中的vfid作为POD身份标识
- 若扩大安全破防半径到node, 可极大加速mTLS性能



零信任容器网络测试床 有/无DPU

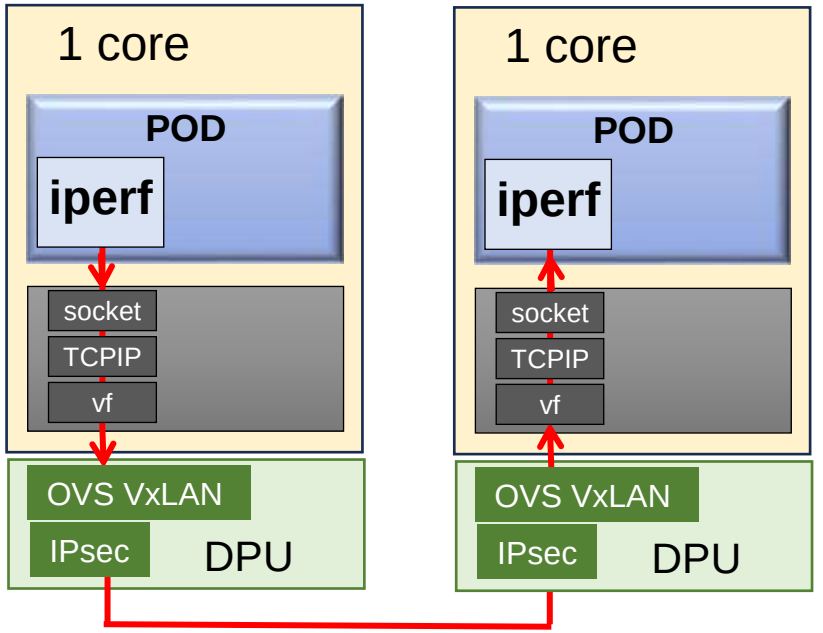
神州数码 Digital China

- 测试床
  - 服务器: 神州数码 2280，CPU: 鲲鹏920, 2*48 cores @2.6G
  - DPU: nVidia bluefield 2
  - TSO and Armv8 Cryptographic Extension缺省使能
  - Iperf 发送4~8 tcp 流，绑定在1个核上
- 1.8x~23.5x 加密TCP流吞吐提升 @ 1个核



Calico+VxLAN+ipsec

VS

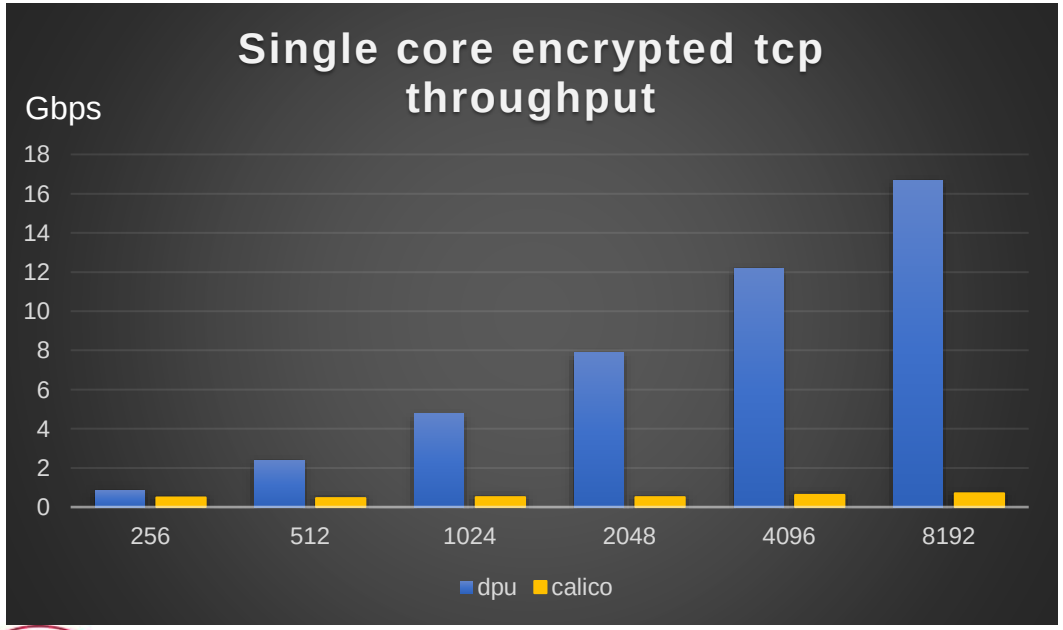


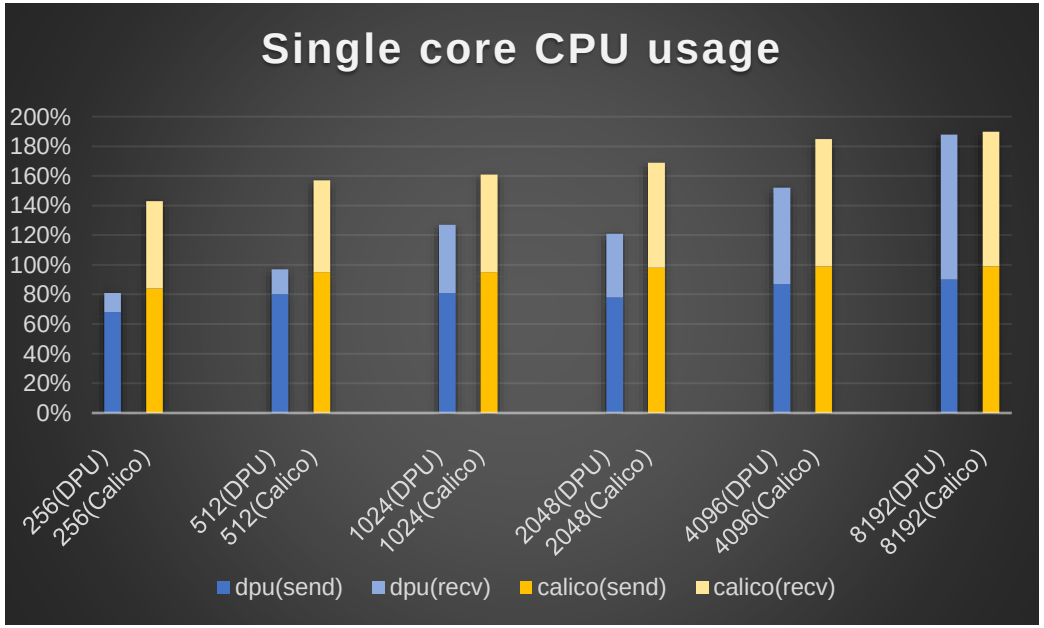
DPU ovs+VxLAN+ipsec

零信任容器网络性能 有/无DPU

神州数码 Digital China

- 1.8x~23.5x 加密TCP流吞吐提升
- 40~1% CPU利用率降低
  - Iperf 使用了 10~30% CPU
  - 3x~24x 吞吐提升@每核

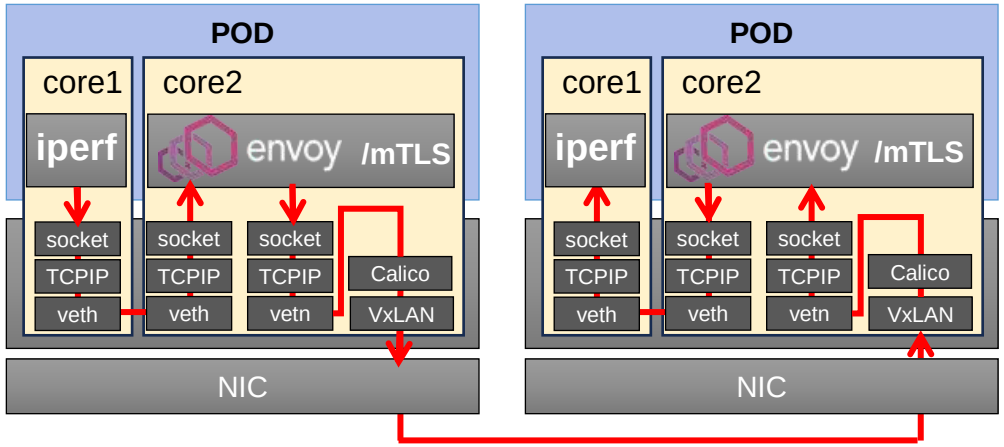




边车性能测试床 有/无零信任DPU容器网络

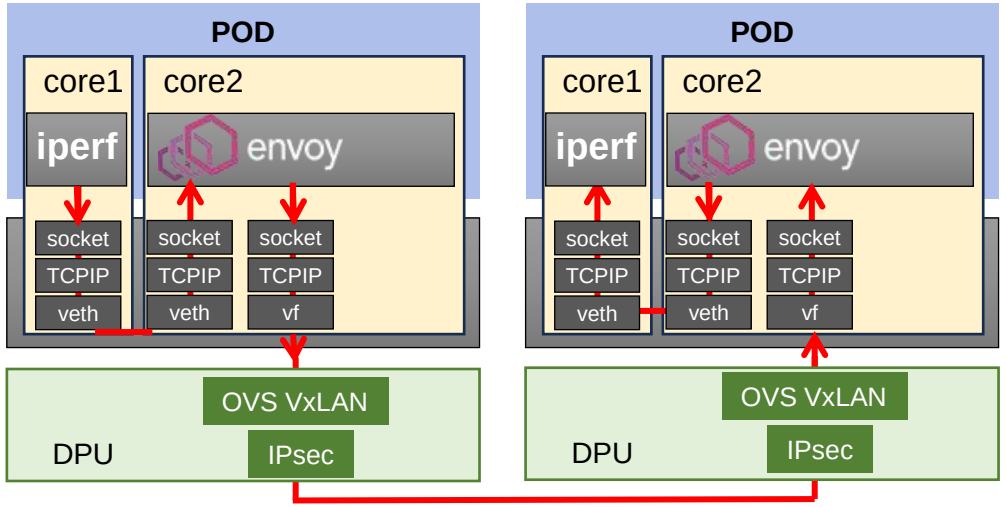
神州数码 Digital China

- 测试床
  - 服务器: 神州数码昆泰2280，CPU: 鲲鹏920, 2*48 cores @2.6G
  - DPU: nVidia bluefield 2
  - TSO and Armv8 Cryptographic Extension缺省使能
  - Iperf 发送4条tcp流，绑定在1个核上
- 3.2x~4.9x 加密tcp吞吐提升 @ 1个核
- 1.5x~4.1x 加密tcp吞吐提升 @ 4个核



Envoy sidecar+mTLS+Calico+VxLAN

VS



Envoy sidecar+DPU (ovs+VxLAN+ipsec)

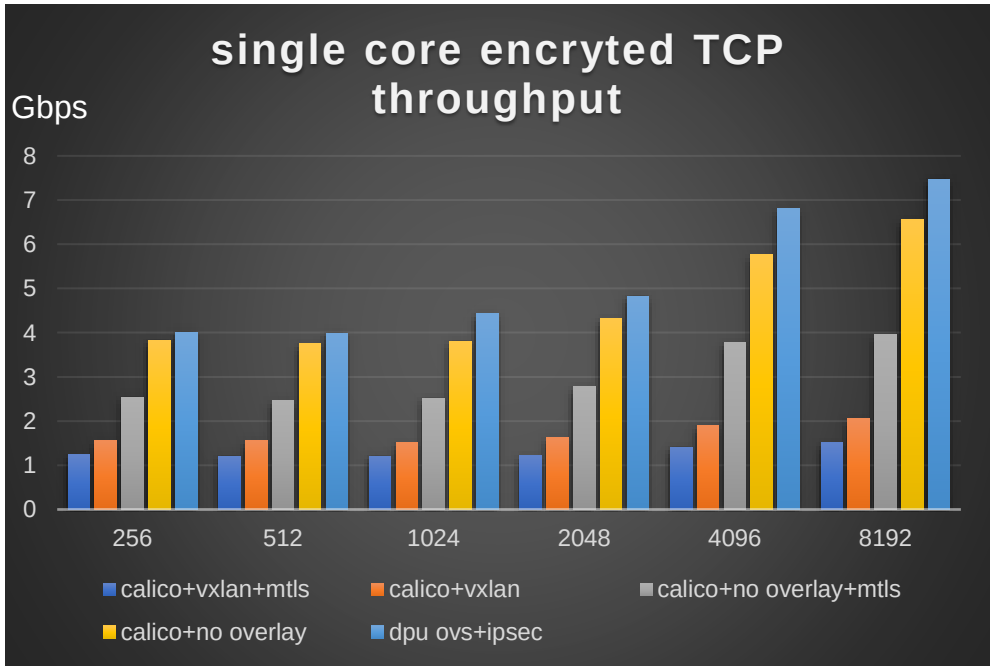
边车吞吐 有/无零信任DPU容器网络

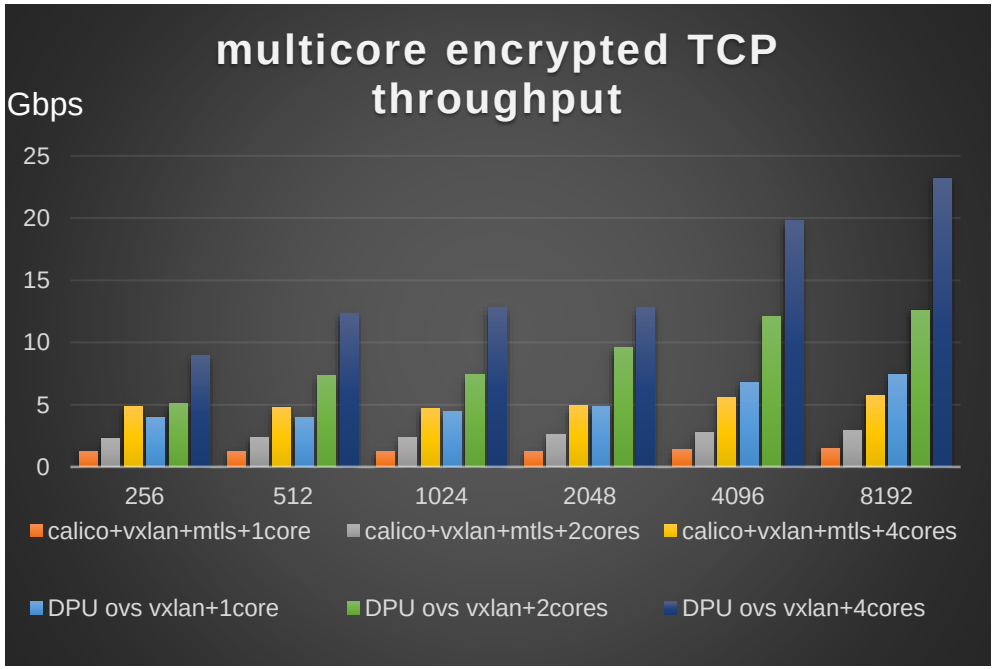
神州数码 Digital China

- vxlan 降低吞吐 60~70%
- Mtls 降低吞吐 30~40%

- 3.2x~4.9x 提升 @ 1 core
- 1.5x~4.1x 提升 @ 4 cores

TSO(vxlan) 比mtls对吞吐形象更大
多核时，内核协议栈锁对性能有明显影响

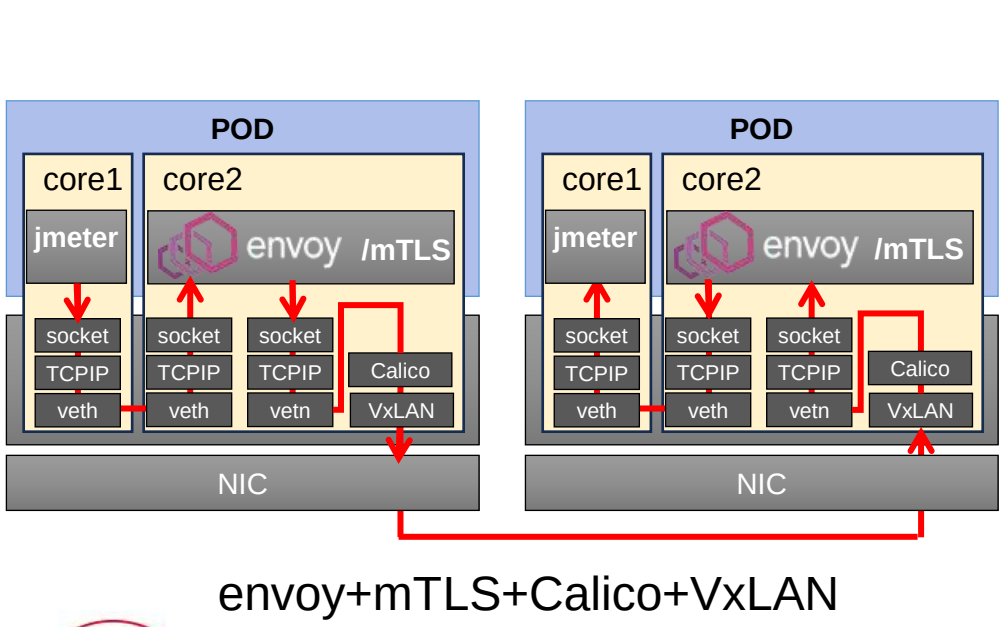




越高越好

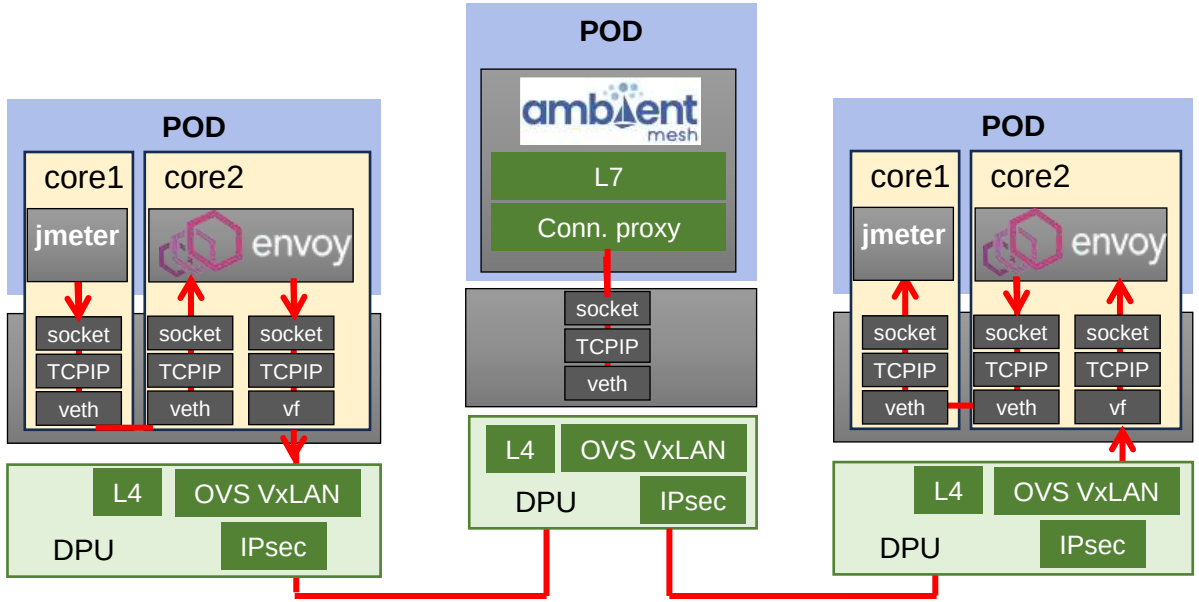
服务网格测试床 有/无边车

- 测试床
  - 服务器: 神州数码昆泰2280，CPU: 鲲鹏920, 2*48 cores @2.6G
  - DPU: nVidia bluefield 2
  - TSO and Armv8 Cryptographic Extension缺省使能
  - Jmeter 发送 50个 并发 http流
- 4x 加密QPS提升 @ 1 core
- 50% 加密http延迟降低



envoy+mTLS+Calico+VxLAN

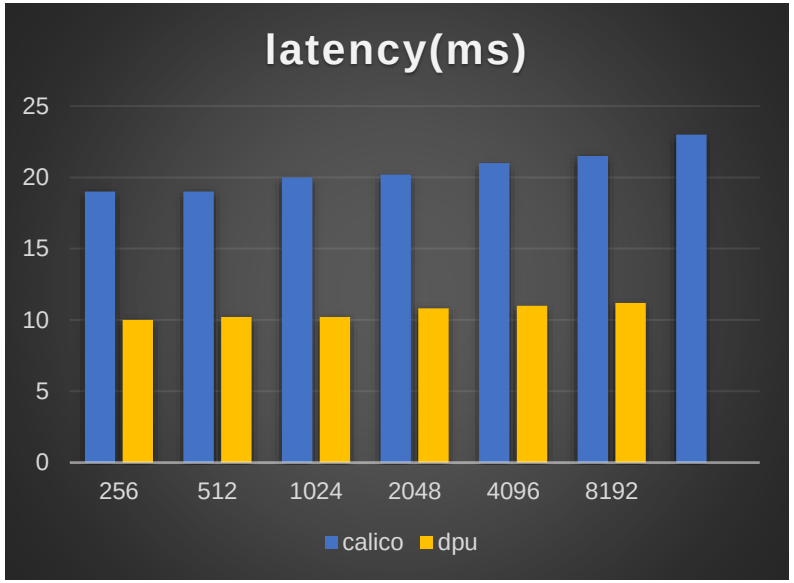
VS

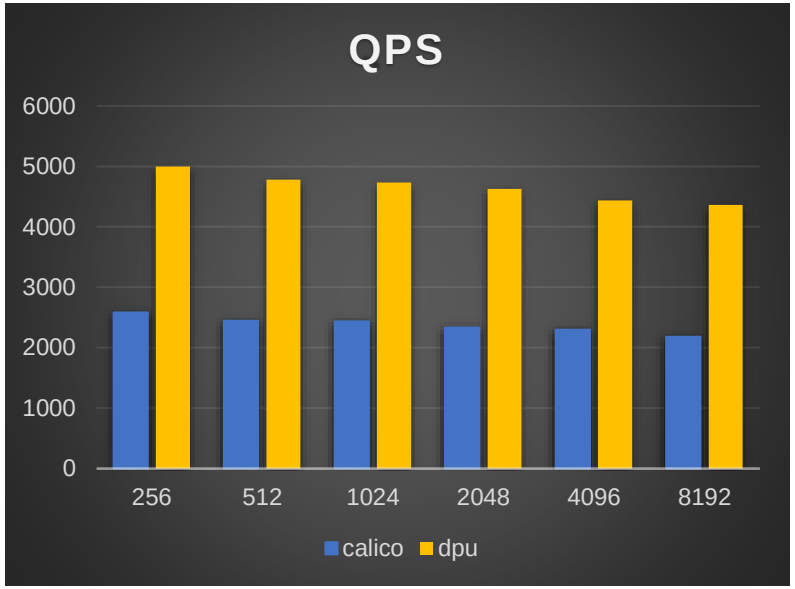


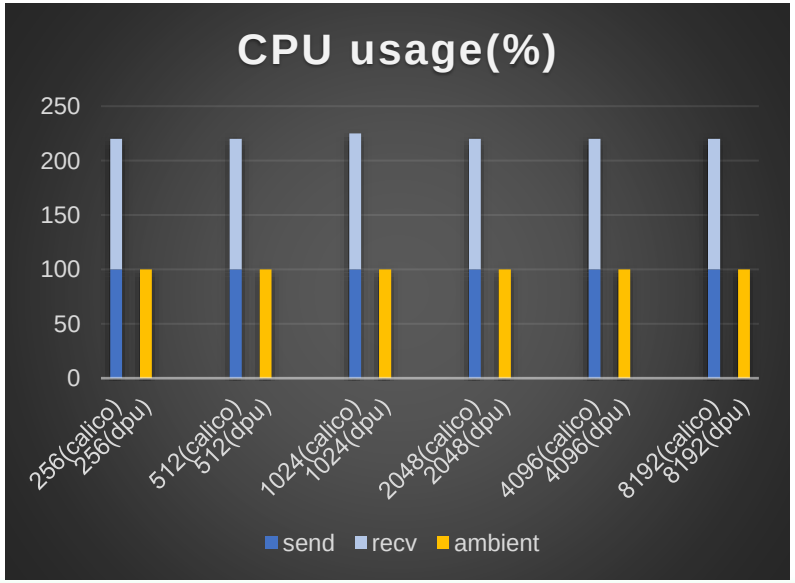
ambient+DPU ovs+VxLAN+ipsec

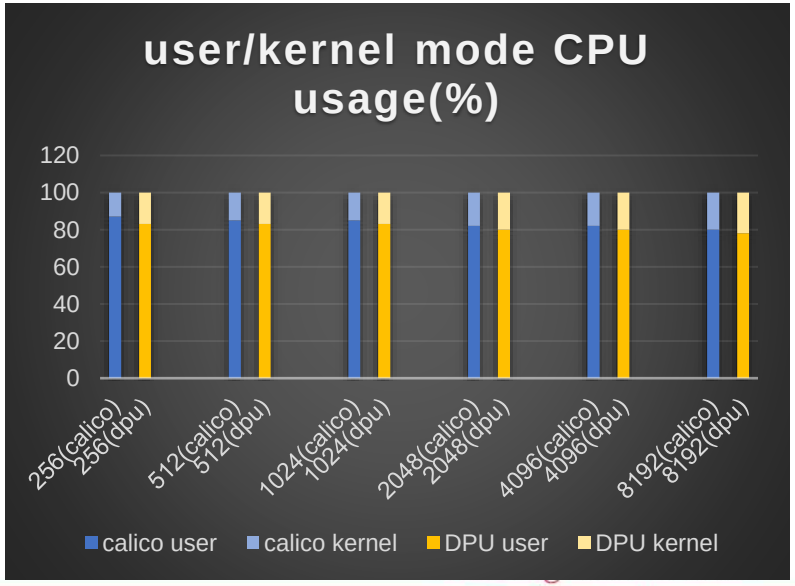
服务网格性能 有/无边车

- 2x 加密 http QPS提升
- 54% CPU占用率降低
  - 4x QPS提升@1 core
- 50% 延迟降低
- 80% CPU 被envoy L7消耗
  - L7 解析
  - 规则匹配
  - 下一步性能提升方向









KUNTAI 神州鲲泰

智 算 神 州
鲲 泰 领 航

神州数码
Digital China