

基于智能网卡卸载的高性能容器网络

刘梦馨

目录

- ✓ | 容器网络的趋势和挑战
- ✓ | 容器网络性能瓶颈
- ✓ | 通过 Kubernetes 管控智能网卡
- ✓ | 性能数据以及未来工作

容器网络应用场景日趋复杂

- 如何兼容传统网络架构，容器网络需和已有基础实施互通互联
- 容器多网络平面，多网卡的管理
- 多租户， VPC 类型容器网络需求
- 细粒度的防火墙控制
- 流量镜像， 流量加密， 多集群互联....
- 简单的L2或L3容器网络无法满足日益增长的复杂度



容器网络性能依然是关键难点

- 微服务化导致网络迅速膨胀
- 网络策略，规则增加额外的资源消耗
- Kube-Proxy 在大规模下的性能瓶颈
- 运营商，边缘计算等场景下对性能的极限需求



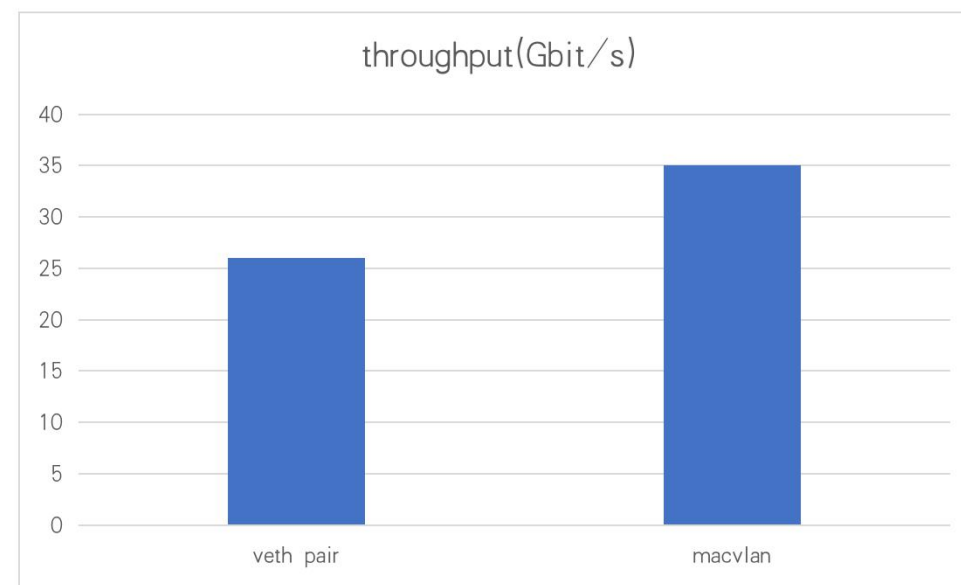
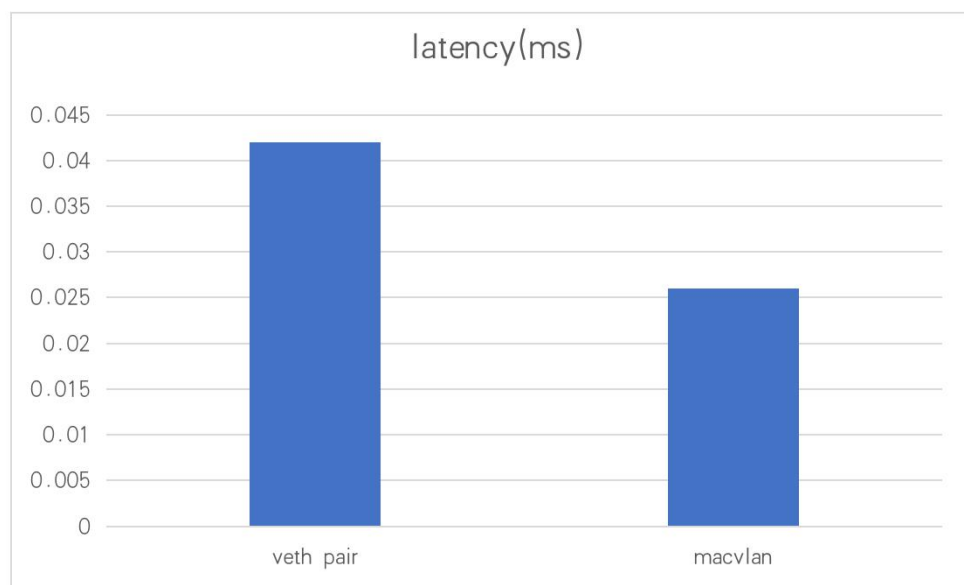
功能 VS 性能

- 功能: overlay, tunnel, openflow, conntrack, mirror, QoS...
- 性能: macvlan, SRIOV, DPDK



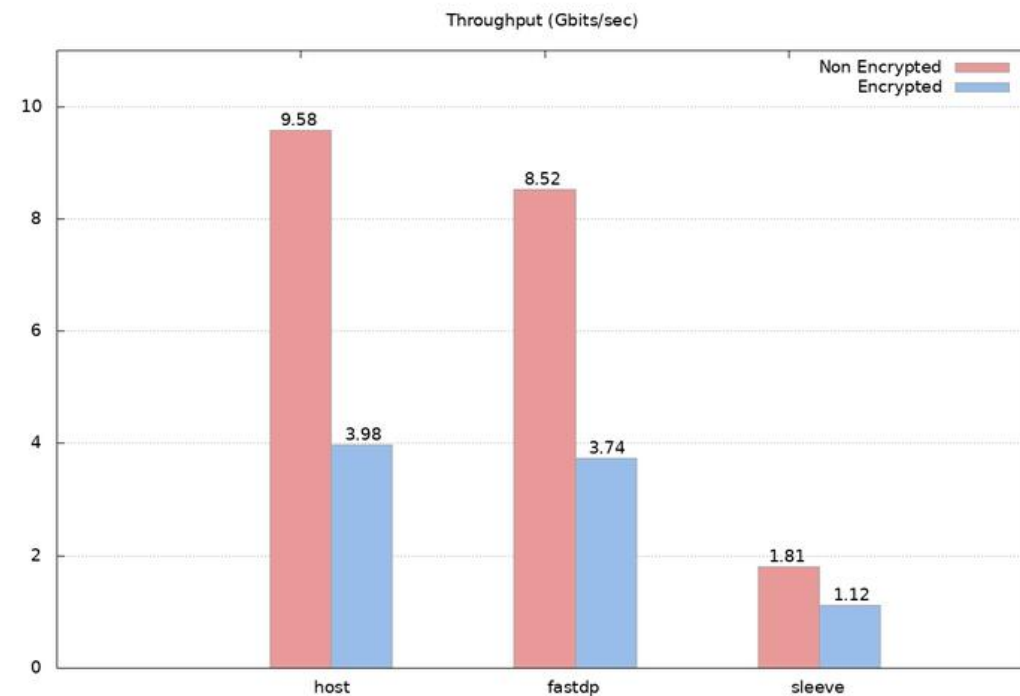
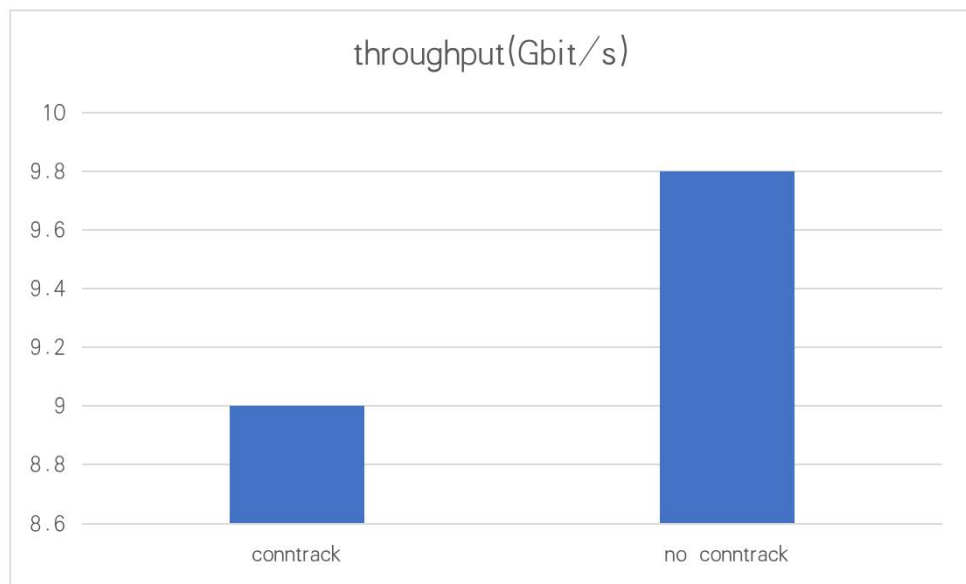
性能瓶颈分析

- Veth Pair 带来的多次 IO 复制



性能瓶颈分析

- Conntrack, IPsec 带来性能下降

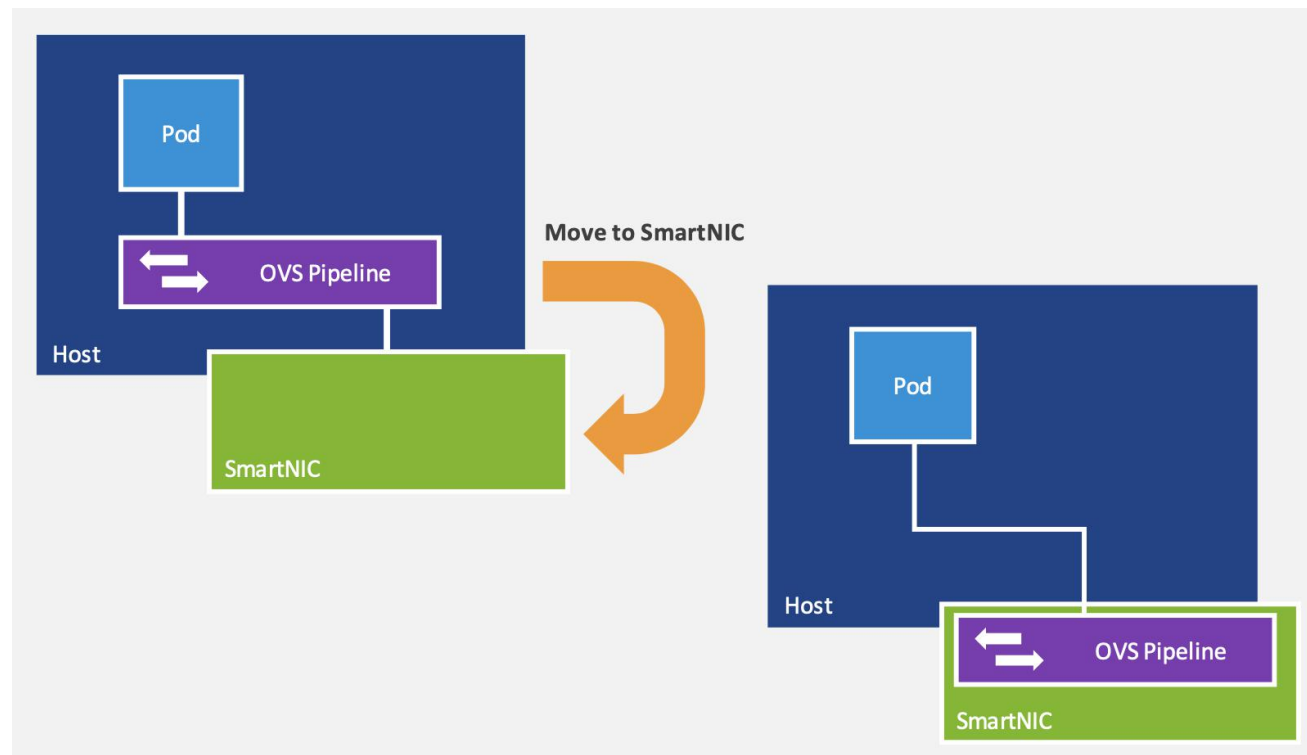


性能瓶颈分析

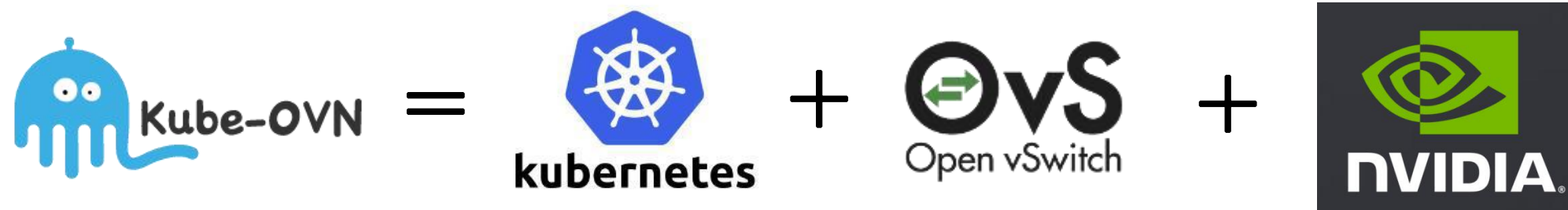
- Overlay 封装性能开销
- Openflow 流表匹配性能开销
- 流量镜像性能开销
- QoS 性能开销

智能网卡卸载能力

- Veth Pair -> VF/Rep
- Conntrack Offload
- Geneve Offload
- Openflow Offload
- Mirror Offload
- IPSec Offload
- QoS Offload



Kube-OVN



- OVS 实现丰富的 SDN 功能
- 利用智能网卡实现高性能网络
- 基于Kubernetes云原生架构
- 结合灵雀云多年的生产实践

Kubernetes + 智能网卡

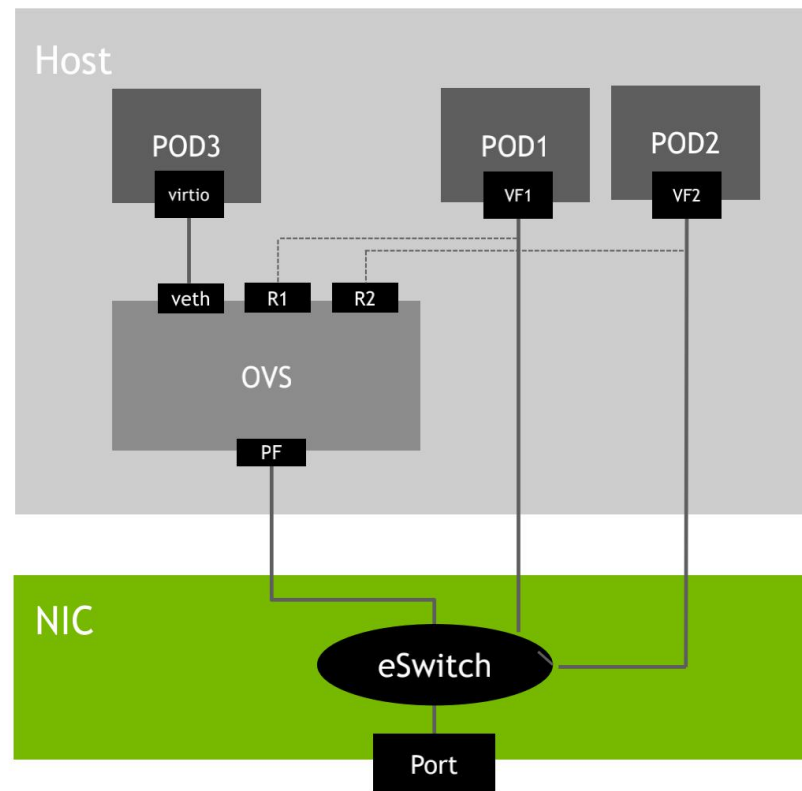
- SR-IOV Device Plugin 纳管智能网卡 VF 资源

```
apiVersion: v1
kind: ConfigMap
metadata:
  name: sriovdp-config
  namespace: kube-system
data:
  config.json: |
    {
      "resourceList": [{
        "resourcePrefix": "mellanox.com",
        "resourceName": "cx5_sriov_switchdev",
        "selectors": {
          "vendors": ["15b3"],
          "devices": ["1018"],
          "drivers": ["mlx5_core"]
        }
      }]
    }
  }
```

```
apiVersion: v1
kind: Pod
metadata:
  name: perf
  annotations:
    v1.multus-cni.io/default-network: default
spec:
  containers:
    - name: perf
      image: kubeovn/perf
      resources:
        requests:
          mellanox.com/cx5_sriov_switchdev: '1'
        limits:
          mellanox.com/cx5_sriov_switchdev: '1'
```

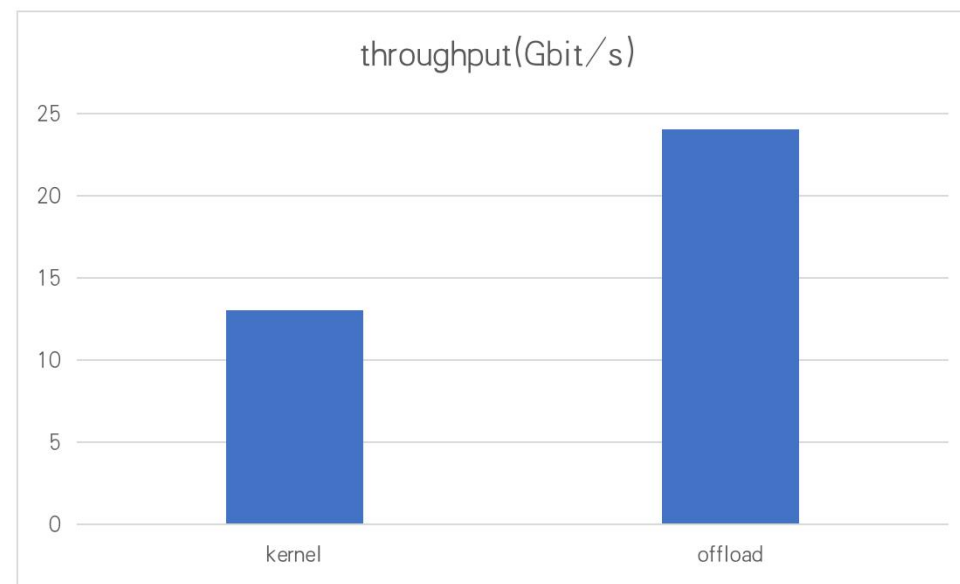
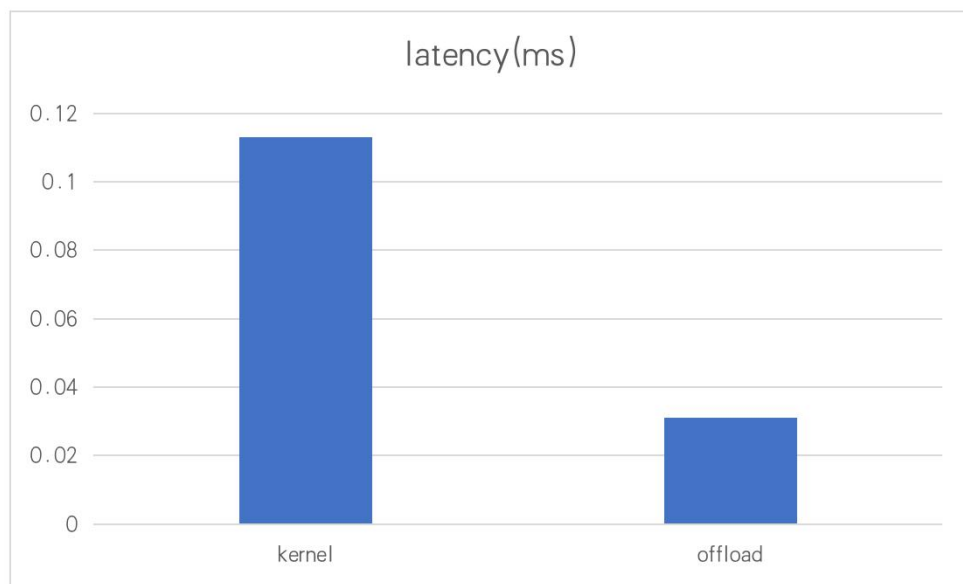
Kubernetes + 智能网卡

- VF 接入容器, Representor 接入 OVS
- 通过TC下发流表到硬件网卡
- Geneve, CT, NAT, QoS ...



测试数据

- Testpmd 21Mpps



未来计划

- CX5 -> Bluefield
- 去除 Kernel 依赖
- OVN/OVS 组件直接部署到网卡中
- 更多 Offload 功能集成
- 期待社区参与!



+



CLOUD NATIVE
COMPUTING FOUNDATION



灵雀云

数字化转型引领者

助力企业获得持续创新的核心能力



marketing@alauda.io



4006-252-832