

2020 CHINA CLOUD
NETWORKING SUMMIT

中国云网络峰会

📅 2020.12.19 📍 北京

软硬件一体化在滴滴云网络落地实践

滴滴出行 技术专家 张同浩

- 面临的问题
- 硬件选型
- 转发模型
- 网关节点探索
- 计算节点探索
- 开源社区贡献



面临的问题

数据中心流量激增

纵向扩展、横向扩展

集群成本

硬件芯片选择

标卡offload/ASIC 专用芯片: 1. 性能 2. 灵活性(很难适应性云上复杂业务场景)

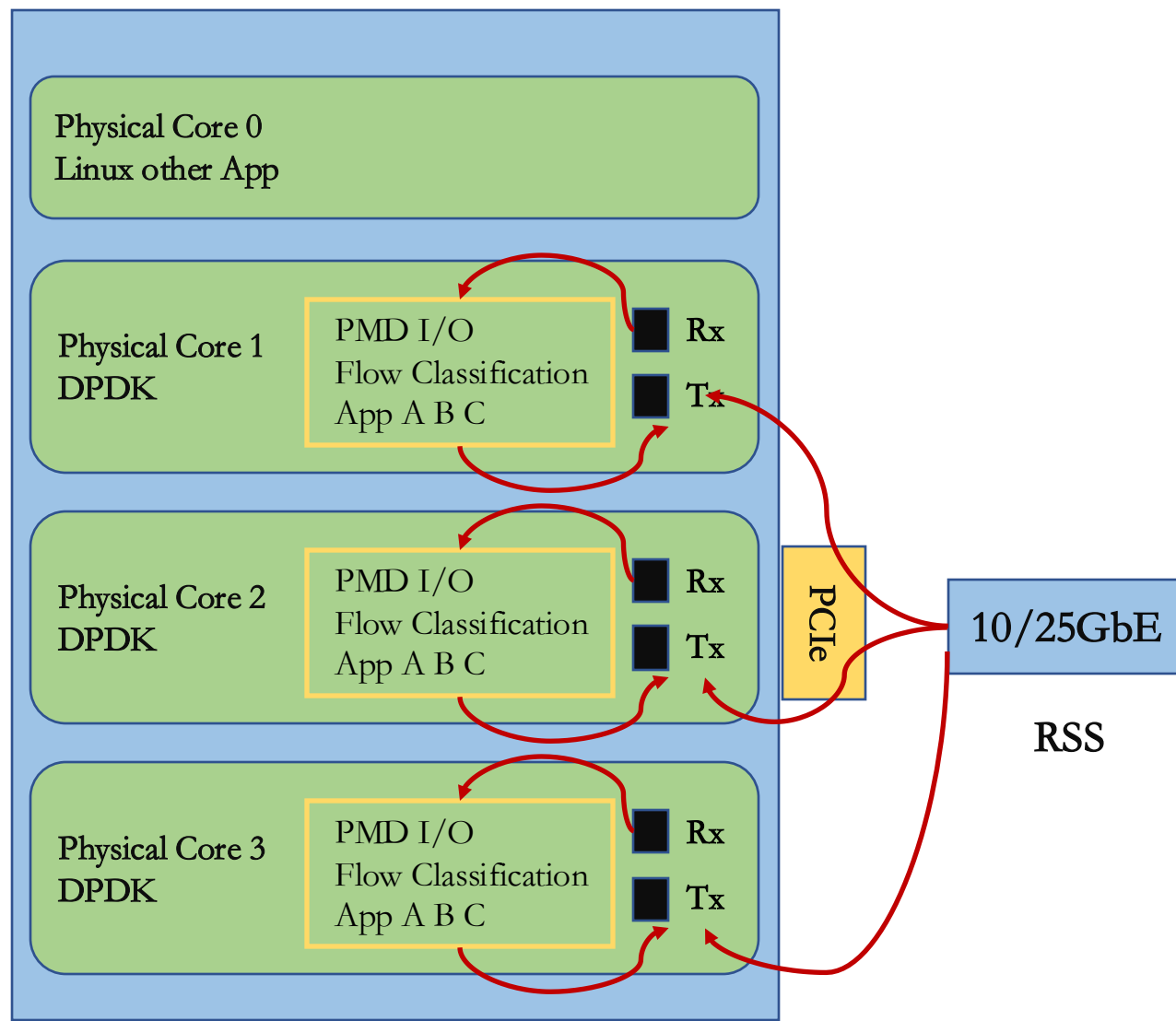
FPGA: 1. 学习成本 2. 价格

P4: 1. 软转发能力(状态) 2. 适用场景 3. 基础网络

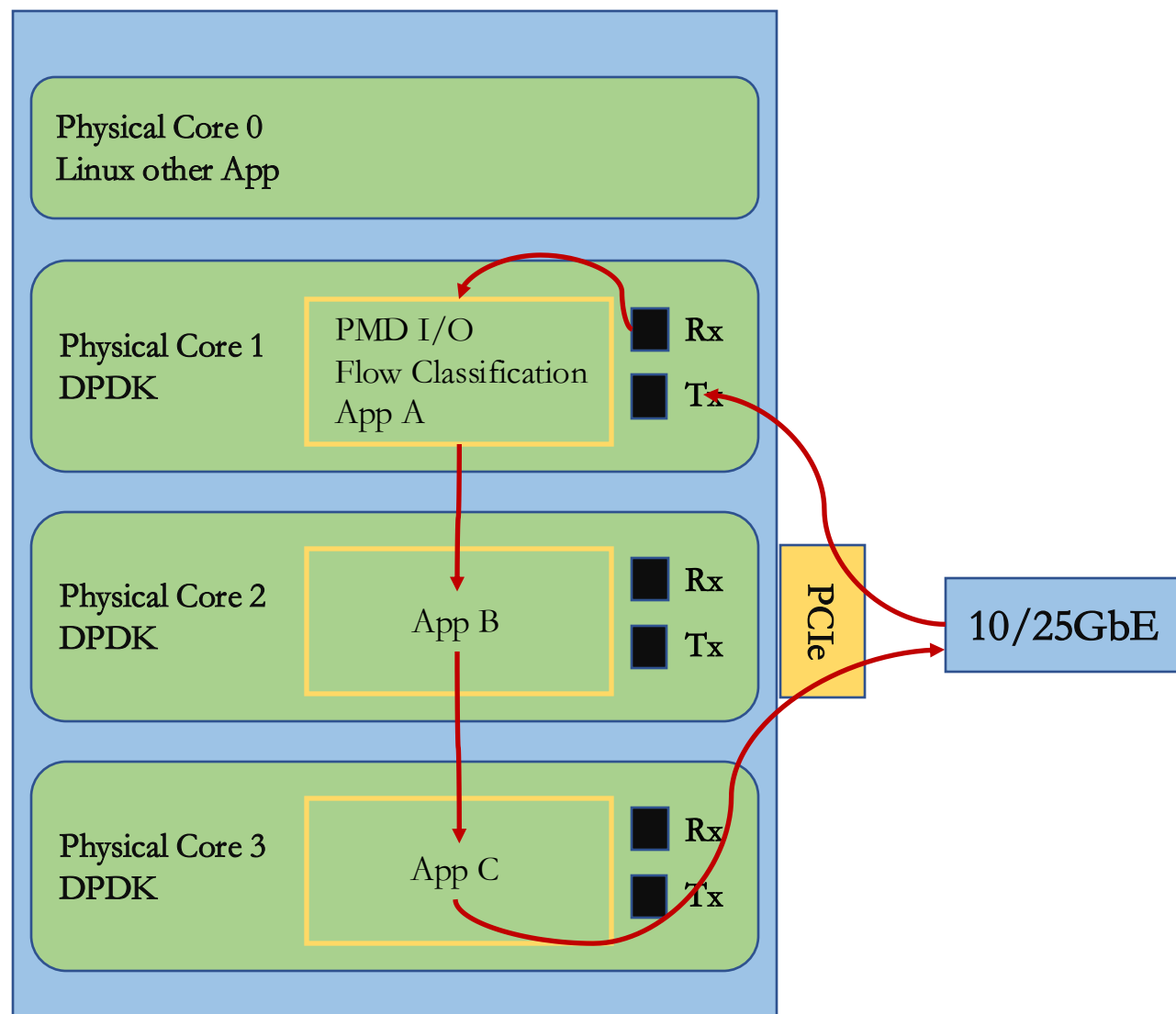
可编程网卡芯片:

1. 性能
2. 规则匹配: 模糊匹配, 精确匹配, 全匹配, metadata
3. 丰富的action: tunnel/rewrite/CPU/jump/meter ...
4. 上送CPU能力 - 状态, 采集
5. hairpin

转发模型的选择



转发模型的选择



转发模型的选择

Match Fields	Priority	Counters	Instructions	Timeouts	Cookie
--------------	----------	----------	--------------	----------	--------

业务解耦

高度灵活性

高性能转发

网络平台化

网关节点加速探索



VPN



SNAT GW



防火墙



弹性公网



连接数据中心



172.16.0.0

172.16.1.0

172.16.2.0

DPDK 软转发
纵向扩展能力
横向扩展能力



DC2

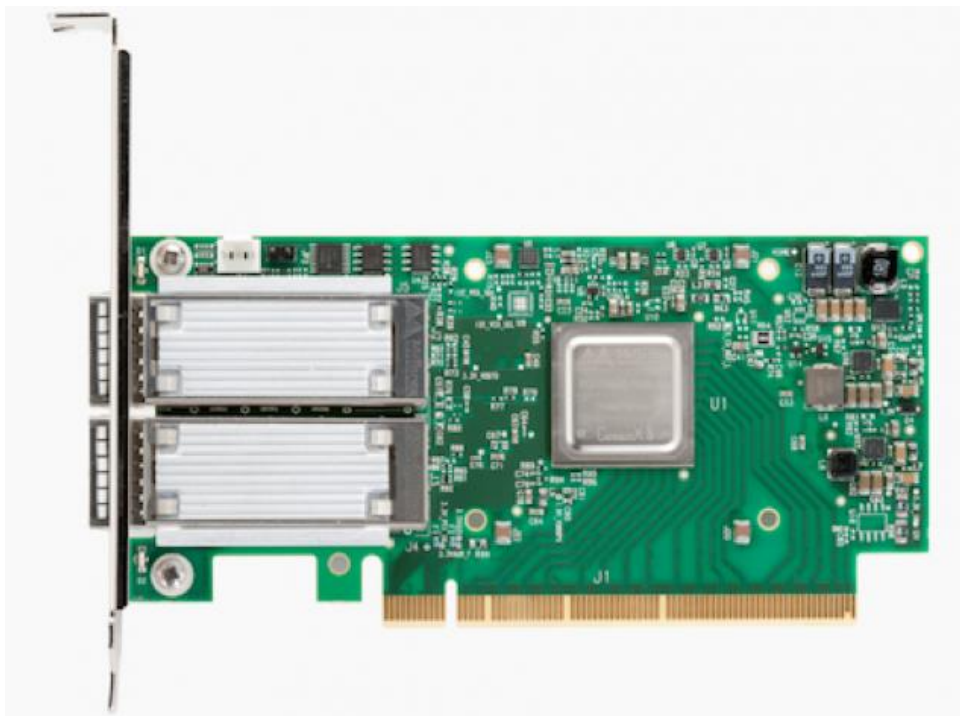


DC2



DC2

网关节点加速探索



1. Hairpin
2. `ovs-appctl dpctl/add-flow match+actions`
3. DPDK `rte_flow offload(meter/decap/encap/hairpin)`
4. 无状态, full-offload
5. 网关平台化

计算节点加速探索

计算节点 1.0

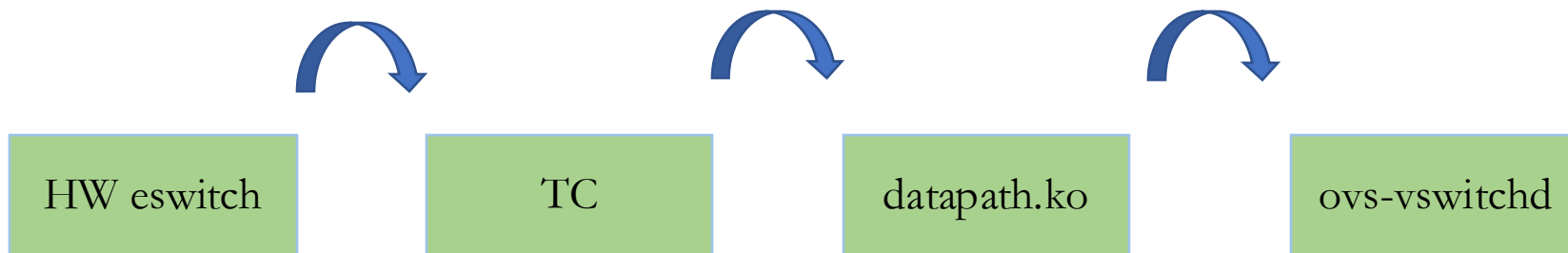
1. Linux Kernel vhost-net
2. DPDK vhost-user

计算节点加速探索

计算节点 2.0

1. TC Flower
2. vDPA (vHost DataPath Acceleration)

计算节点加速探索



数据面

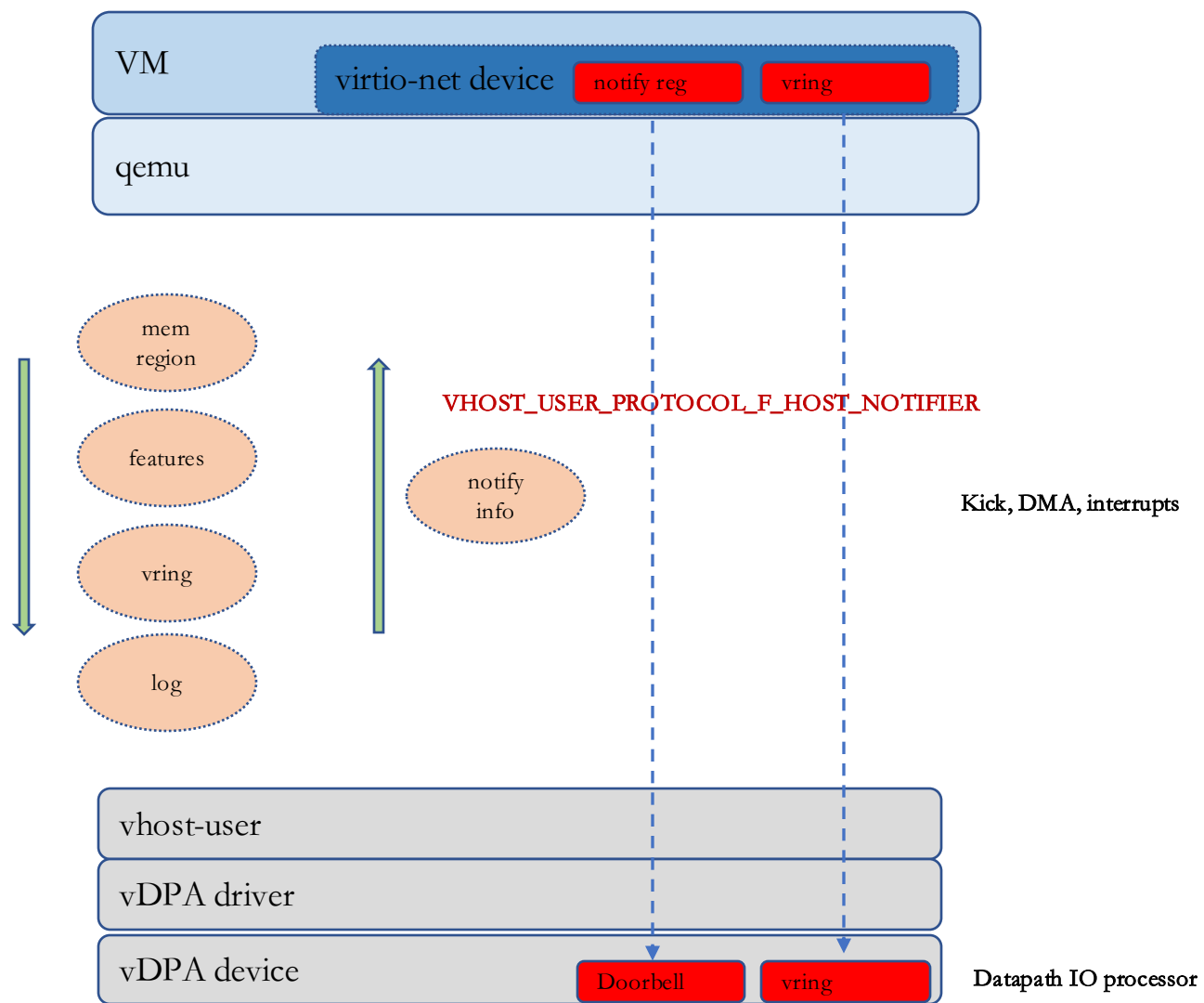
TC Flow mask 匹配

TC 优先级

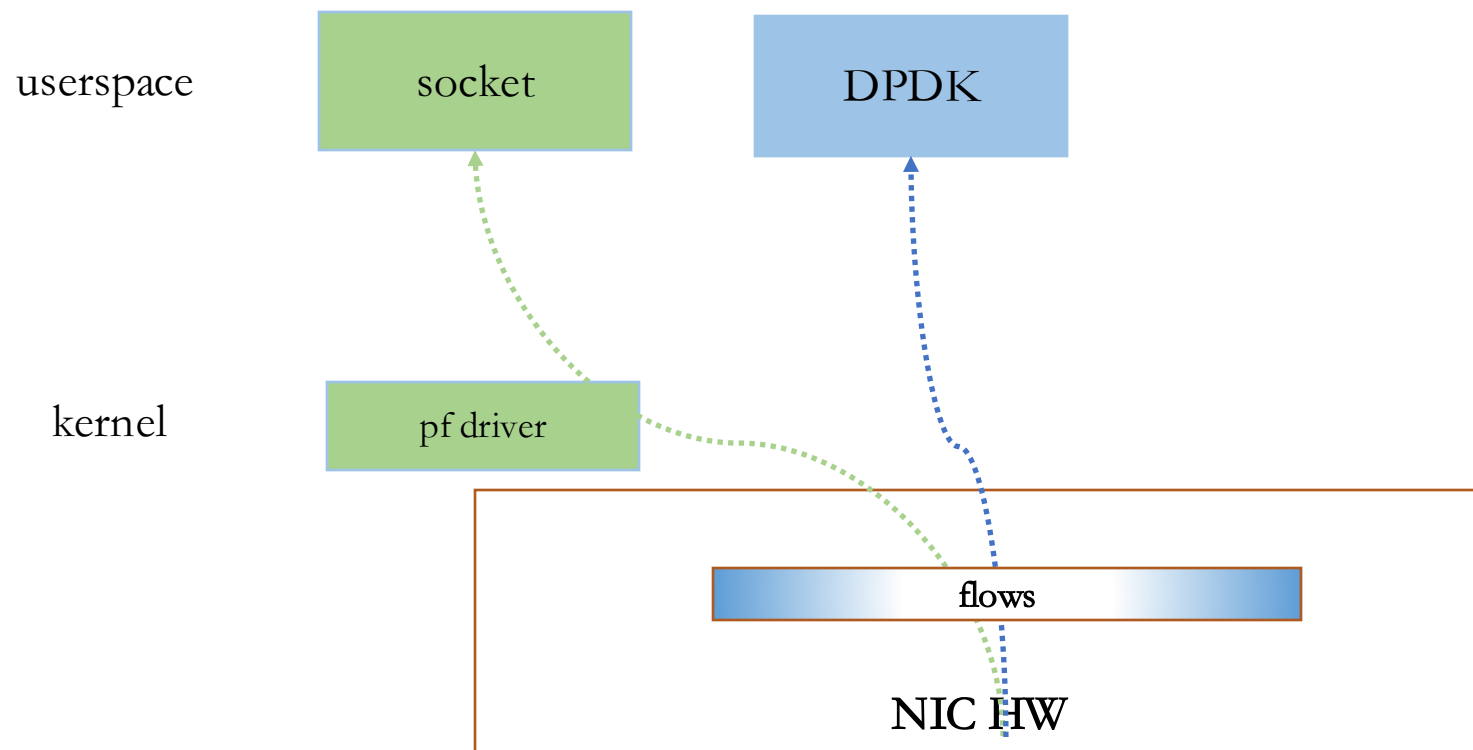
TC recirc_id

高内核版本? 稳定性?

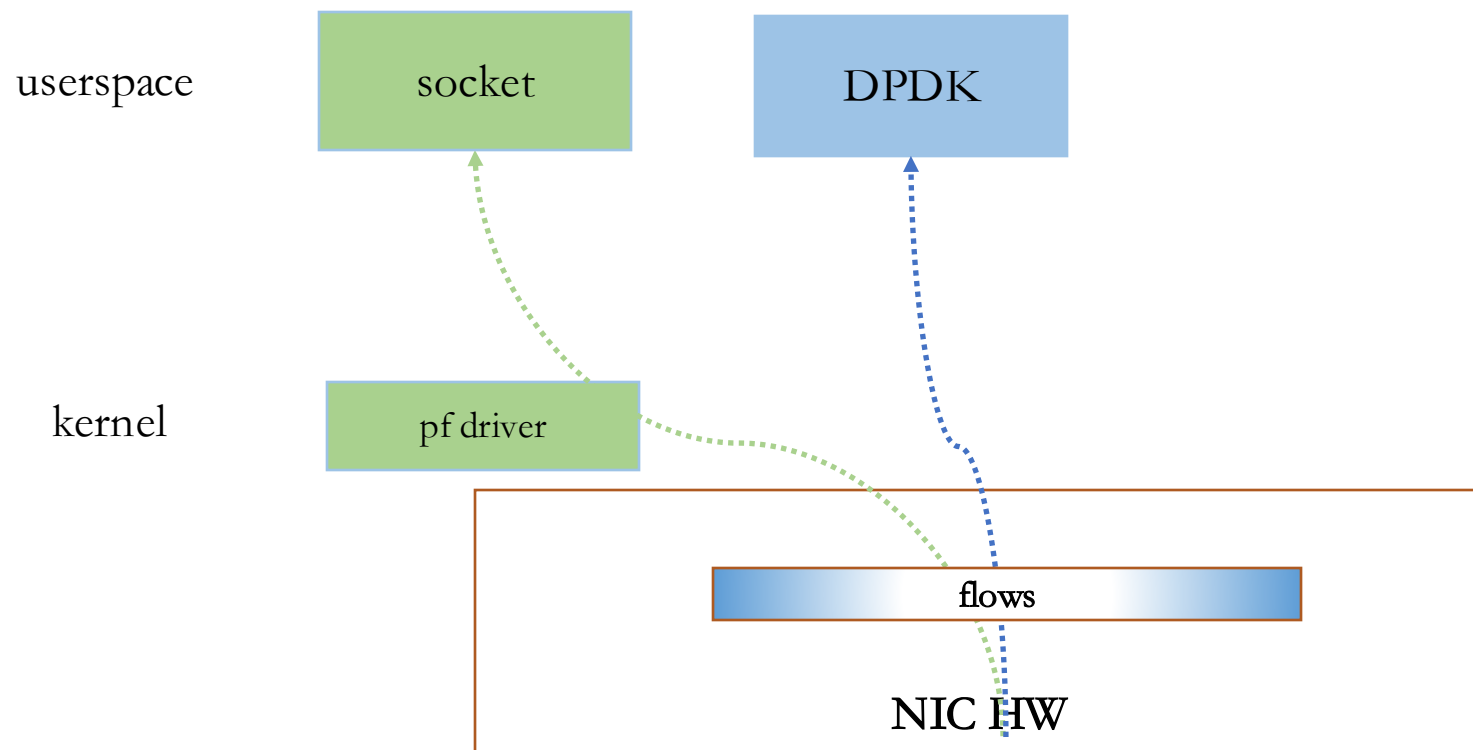
计算节点加速探索



计算节点加速探索



计算节点加速探索



计算节点加速探索

br-int

ARP ?

路由表 ?

优先级/重计算 ?

流表管理

计算节点加速探索

br-int

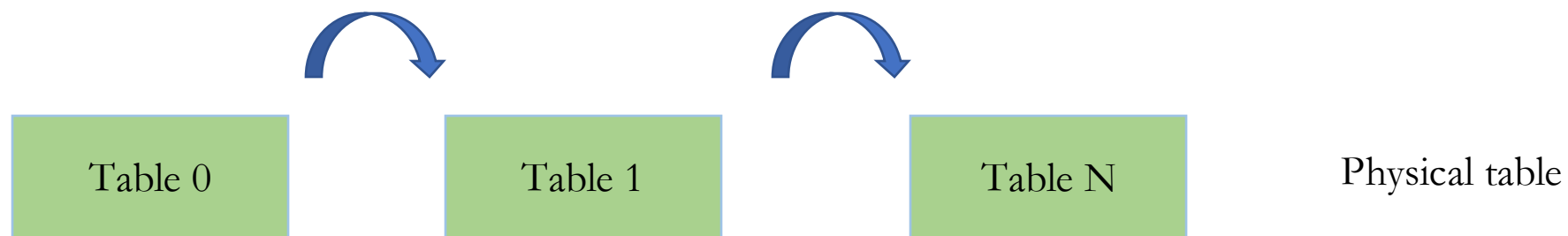
ARP ?

路由表 ?

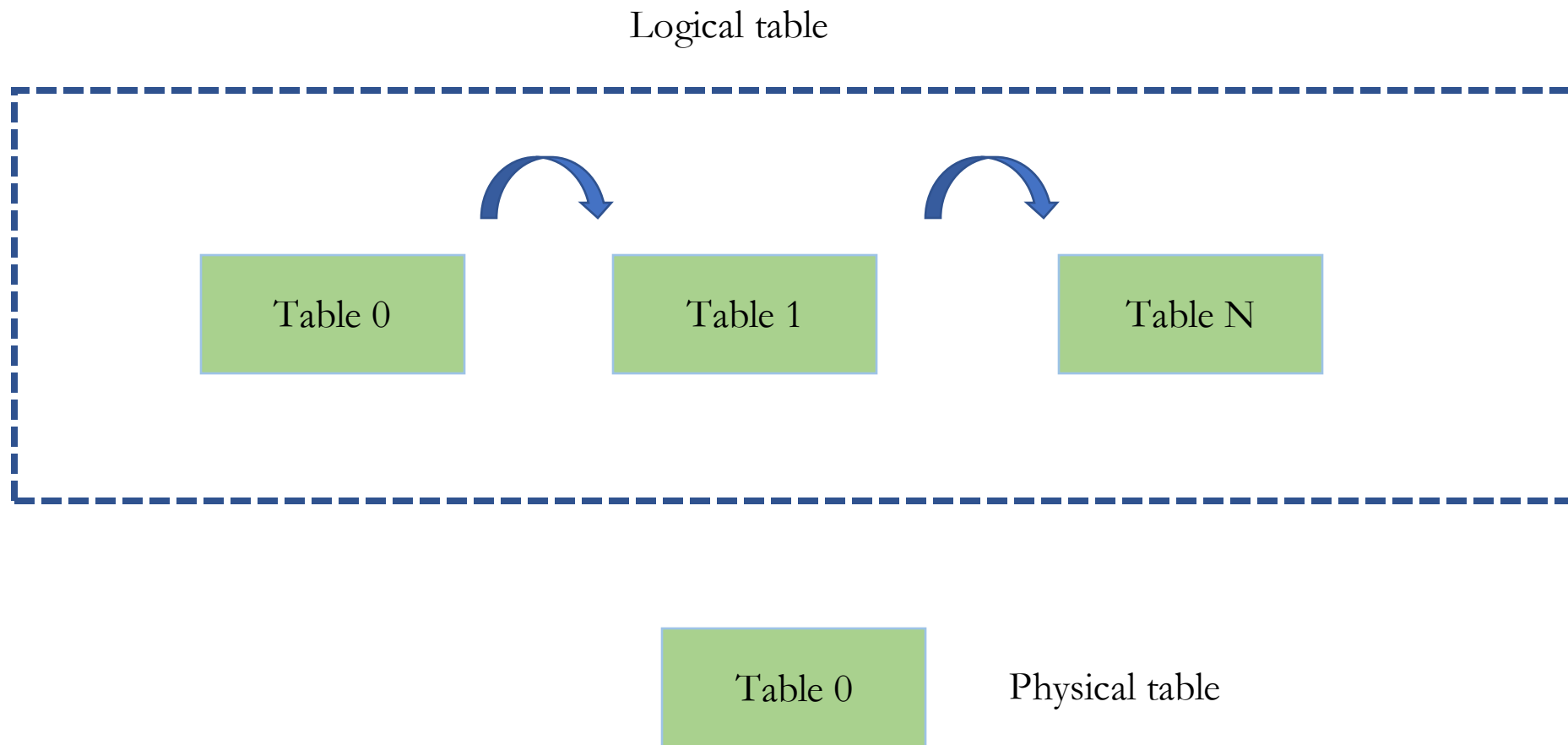
优先级/重计算 ?

流表管理

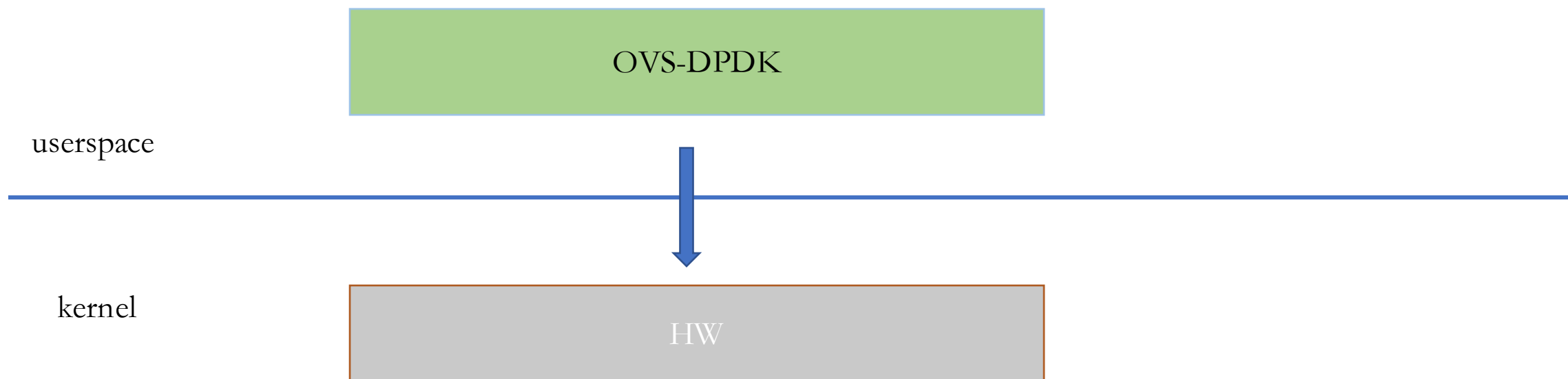
计算节点加速探索



计算节点加速探索



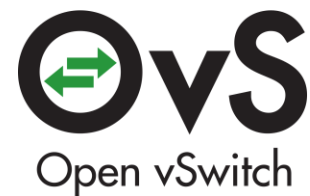
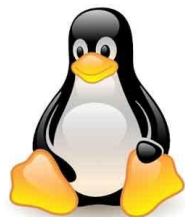
计算节点加速探索



计算节点加速探索

方案	吞吐性能64B	网络延迟
vDPA offload	25Mpps	100us

开源社区贡献



1. 社区接收共计： 150+
2. 进行时？ 哪些特性



THANKS!

