

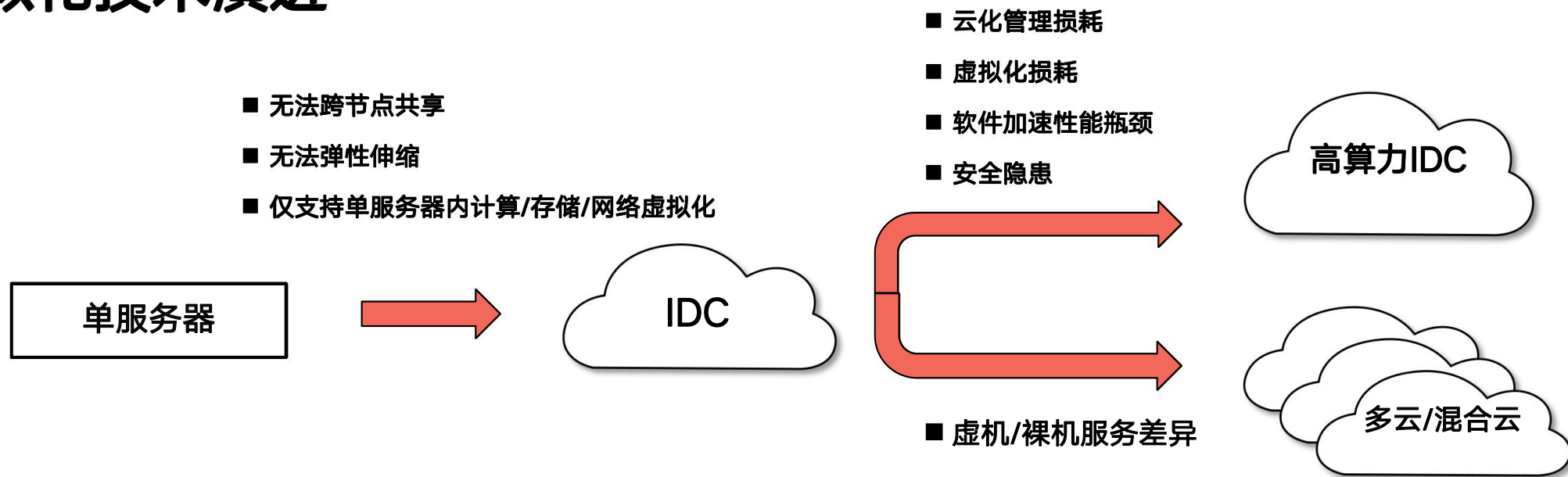
2021 中国智能网卡研讨会

CHINA SMARTNIC WORKSHOP

运营商智能网卡部署场景探索及思考

王瑞雪

虚拟化技术演进



虚拟化

- 单服务器物理资源共享
- 虚拟资源相互隔离

资源池云化

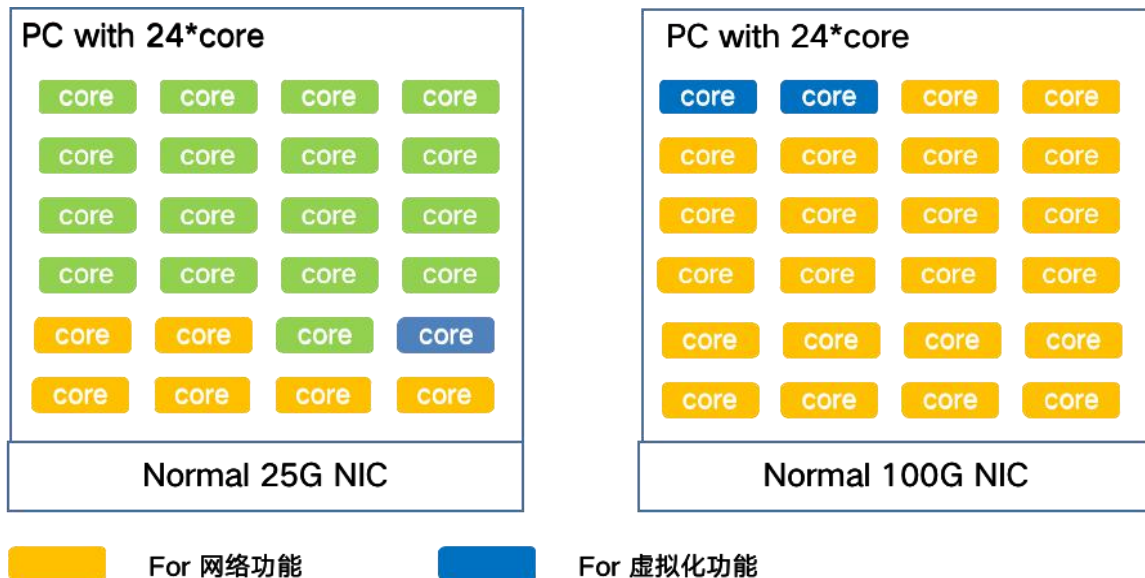
- 资源池内计算/存储/网络资源全面云化
 - 资源共享，弹性伸缩
 - 统一调度和接口

软硬协同虚拟化

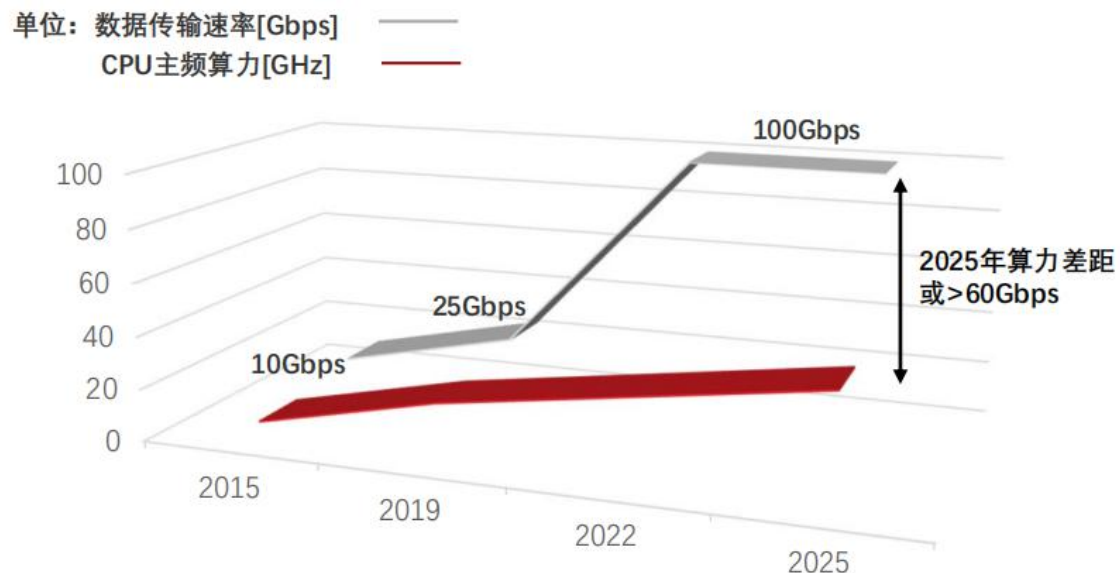
- 无管理消耗，虚拟化损耗微乎其微
 - 硬件加速提升性能
 - 高安全保障
 - 虚机/裸机无差异服务

虚拟化演进面临的挑战

应用激增使得数据中心流量以每年25%速度增长，网络向高带宽和新型传输体系发展，网络堆栈处理越发复杂；后摩尔定律时代，**CPU 计算能力增速低于网络传输速率增速**，且差距持续增大，将网络功能卸载到可编程硬件的需求愈发急迫



- 以24核计算型服务器为例，网络功能占用6个core，虚拟化功能占用1个core，可用于VM的core数量为17个，可用CPU资源比例为**70%**
- 当网卡升级到100G时，**CPU资源基本都被占用**，算力资源基本不可用



CPU算力增长速度与数据中心网络传输速率增长速度差距逐渐拉大

智能网卡能做什么

受市场需求的强驱动，在服务器侧引入智能网卡，将网络、存储、操作系统中**不合适CPU处理的高性能数据处理功能**卸载到硬件芯片执行，提升数据处理能力，释放CPU算力

网络、存储等功能卸载

➤ 释放可用CPU资源，提高对外服务能力，增产增收

DPDK、SPDK和RDMA等技术集成

针对特定业务逻辑进行硬件加速

- 通过各种KernelBypass技术，在用户态完成业务处理，提高服务器性能
- 硬件替代CPU完成专用设备NFV后处理逻辑，实现硬件加速

解决裸金属存储网络的安全隐患

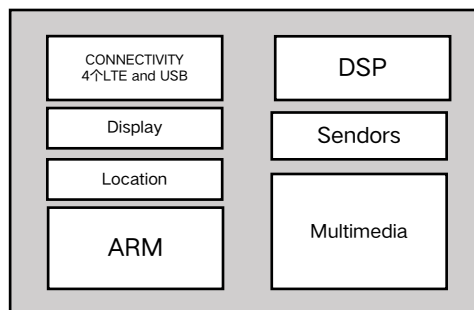
➤ 存储网络信息不再暴露给到用户权限界面之内，从而解决原有安全隐患

业务端到端网络可视化

➤ 将观察点从硬件交换机延伸到主机侧，实现网络端到端网络运维可视化

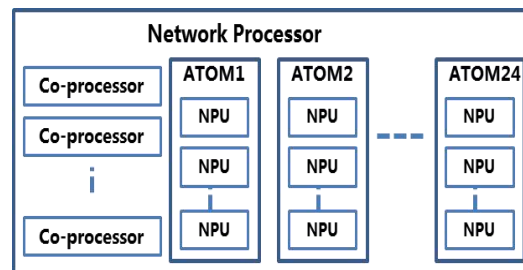
业界主流智能网卡形态

市场上加速芯片有多种选择，**单一加速芯片难以满足复杂多样的加速场景**，因此需要综合考虑业务需求、芯片能力以及加速芯片的产业成熟度、功耗等，选择适合的加速芯片，以达到更好的加速效果



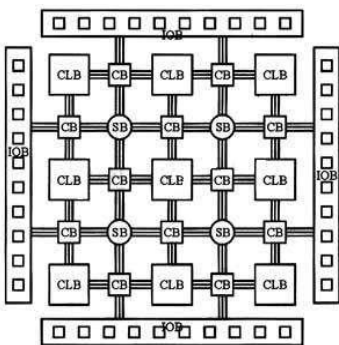
SoC

- ✓ 在终端市场应用较成熟
- ✓ 硬件需根据客户需求定制，周期较长
- ✓ 计算效率高
- ✓ 适合成熟算法及应用
- ✓ 功耗较低



NP

- ✓ 生态封闭，主流厂商已无路标
- ✓ 不支持重编程，难以解耦
- ✓ 功耗较低，成本高于FPGA



FPGA

- ✓ 生态开放，数据中心广泛应用
- ✓ 可重编程，适合演进中的算法及应用
- ✓ 适用于网络转发等并行计算场景，处理时延低
- ✓ 支持虚拟化
- ✓ 功耗适中



ASIC

- ✓ 硬件根据客户需求定制，开发成本昂贵，生产周期长
- ✓ 计算效率高
- ✓ 不具备灵活性，适合大规模使用的成熟算法及应用
- ✓ 功耗较低

5G时代运营商网络演进趋势

运营商网络正在从**自动化**向**智能化**，从聚焦**业务快速发放**向**数据高效处理**、**网络高效运维**转变

大带宽

数据驱动网络带宽以几何级数增长，
接口速率向100G接入400G汇聚演进

低时延

算力向边缘移动，降低成本
业务向边缘移动，降低时延

可视化运维

业务端到端网络流量可视化，实现智能化、精细化运维

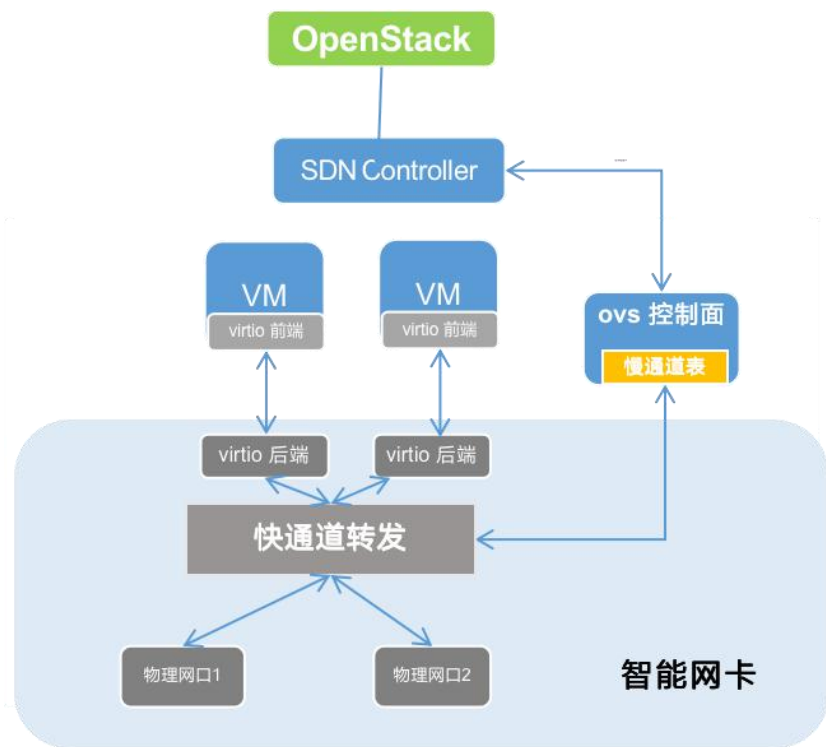
NFV

4G、5G控制面网元已全面NFV化，
UPF、DPI、BRAS-U等网元逐步NFV部署



网络功能卸载，持续向大规模、大带宽演进

中国移动IT云和网络云均采用混合SDN方案，面向不同业务提供虚拟机或裸机部署能力，面向虚拟化场景，引入智能网卡突破提升vSwitch转发性能和数据处理能力；面向裸机场景，引入智能网卡构建弹性裸金属服务



➤ 面向虚拟化场景，将vSwitch转发面卸载至智能网卡中，提升转发性能及表项规格

- 软件vSwitch依靠CPU转发，通过将转发面卸载到硬件网卡，实现主机侧CPU零消耗
- 满足25G、100G网卡及后续更大带宽网卡线速转发要求
- 提升快表规格到百万级甚至千万级，满足SBC等大规模网元部署需求

➤ 面向裸金属场景，实现网络功能及云管理组件等卸载，简化裸机发放流程，提升网络性能

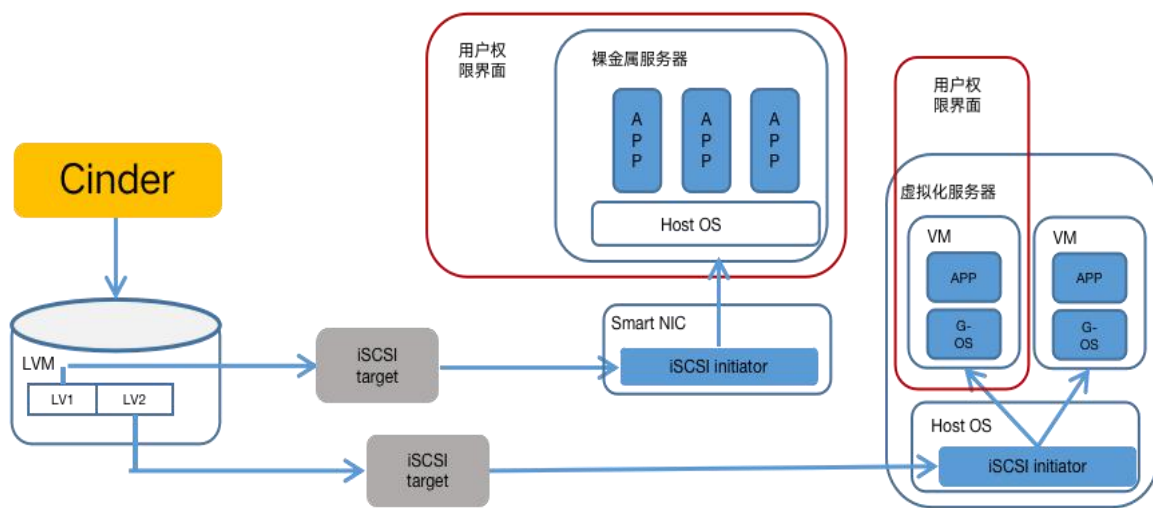
- 实现对VxLan、Flow Table、ACL等网络功能转发卸载和加速，实现网络虚拟化
- 通过快速通道直接对外通信，缩短了网络处理路径耗时，减少了CPU开销，降低时延
- 将云管理组件卸载至智能网卡中，实现裸机与虚拟机同样的发放与管理流程，实现裸机快速发放、迁移、云盘启动等功能，简化客户配置

存储功能卸载，提升存储性能，构建低时延网络

存储功能卸载包括**云盘挂载卸载**和**高性能存储协议卸载**，前者通过支持virtio-blk，提高存储访问灵活性和安全性；后者面向边缘计算视频加速、CDN等场景，进一步提升存储协议处理性能，构建端到端低时延网络

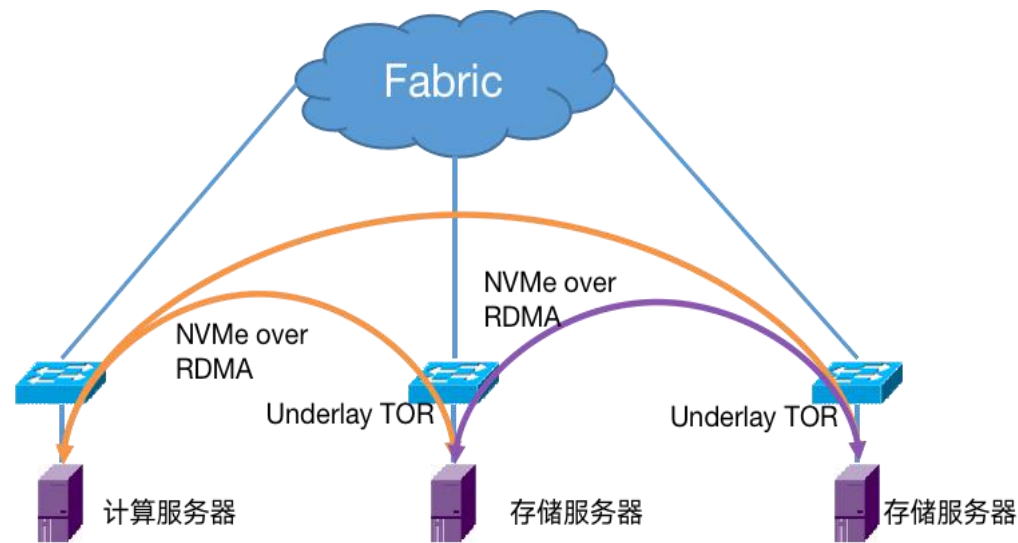
场景一：裸金属云盘挂载

- 减少本地挂盘，利用存储池化资源池，降低成本，提升可靠性
- 模拟virtio-blk，存储网络不再直接暴露给用户操，减少安全隐患



场景二：高性能存储协议卸载

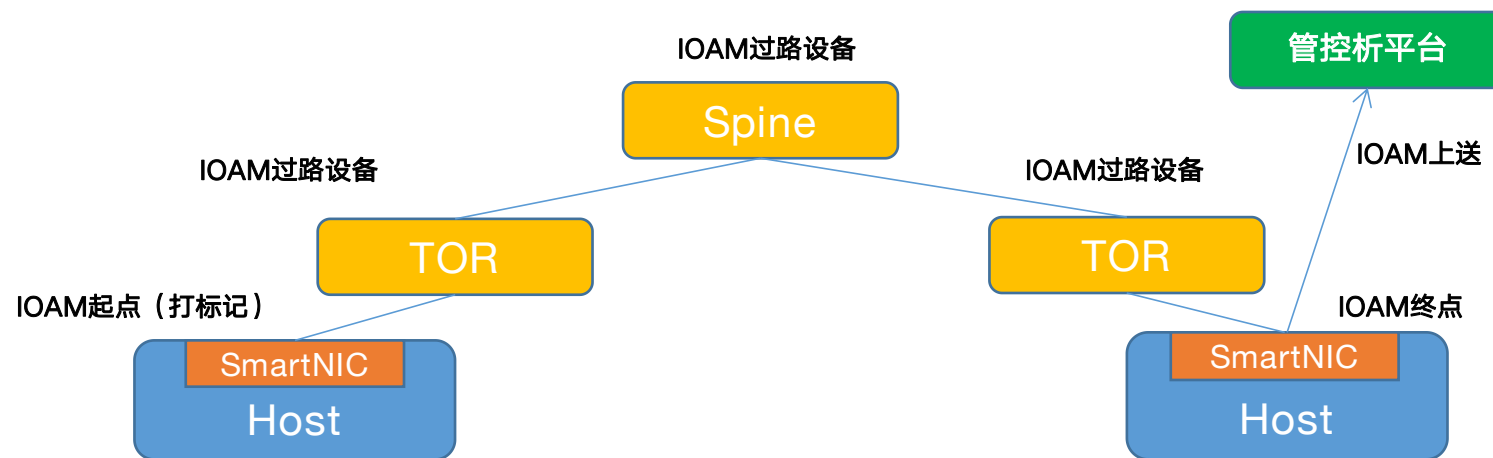
面向边缘视频加速、CDN等场景，引入智能网卡实现NVMe-oF、ROCE等协议卸载，提升协议处理性能，释放CPU算力



运维能力卸载，满足业务网络端到端可视化

传统SNMP协议，无法反映路径、丢包、时延等真实网络状态；随着VxLAN、RDMA等协议部署，网络转发行为愈发黑盒；当前硬件交换机及vSwitch实现仍存在限制，采样性能及精细化程度受限

- TOR作为观察报文采样的起点和终点，对业务报文进行IOAM报文解封装及染色，存在以下问题：
 - TOR采样匹配表项规格不足，协议解封装和采样性能受限，采样报文数量受限
 - TOR作为最后一跳，无法上送真实出接口时间戳，时延计算不准确
- vSwitch作为报文采样的起点和终点，但会大量消耗Host端CPU资源，但裸机侧无法实现



引入**智能网卡**实现真正实现业务端到端
网络可视化，降低CPU消耗

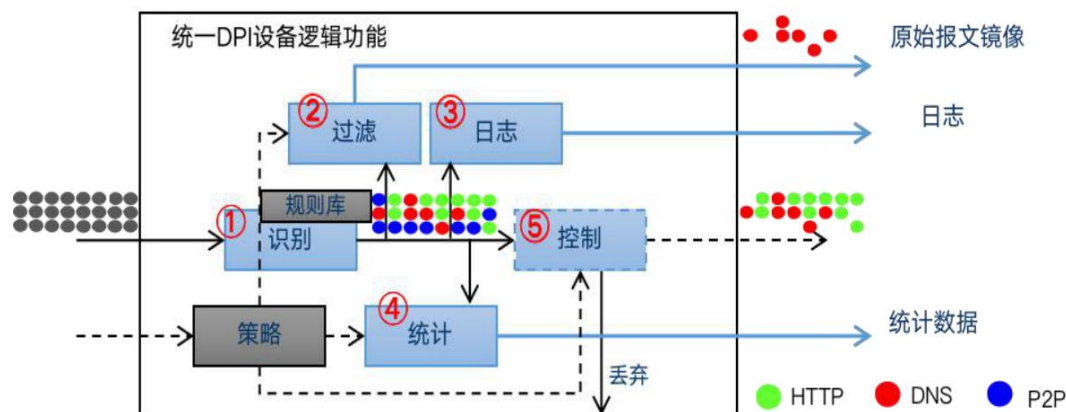
业务功能卸载，满足高性能网元虚拟化要求

5G时代用户面网元大量下沉到边缘，以满足低时延、大带宽的业务需求，同时通过NFV技术，提升网元部署的灵活性，实现软硬件解耦；相对于传统专用硬件，虚拟化用户面网元**采用CPU处理业务逻辑，导致网络功能受限、处理性能低**

UPF虚拟化采用智能网卡加速业务处理已成为业内主流方案，UPF位置下沉及虚拟化对周边网元部署位置及形态产生影响；DPI作为重要配套设备，其高新建、高并发、高吞吐的特点使得性能成为vDPI虚拟化部署的瓶颈

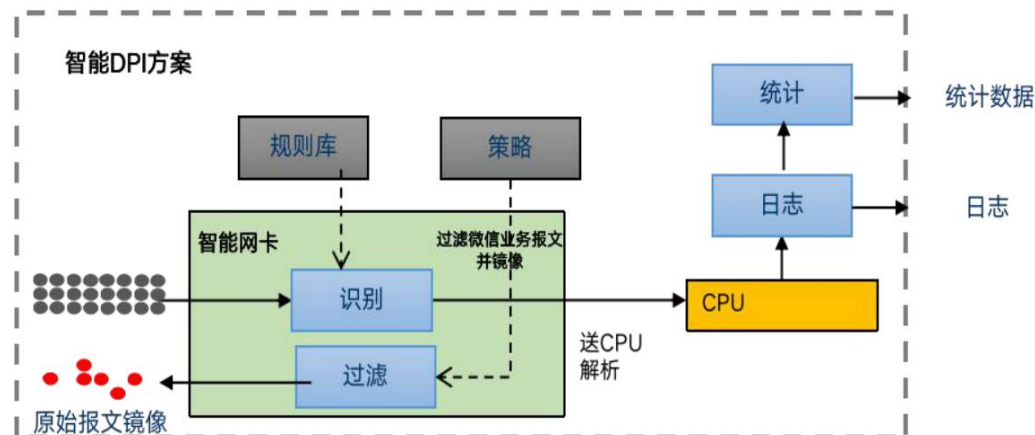
虚拟化DPI

- 所有流程均经CPU，DPI设备虚拟化后性能下降
- 镜像能力受限，无法满足上层需求



智能化DPI

- 识别：预先加载业务识别规则库，移植CPU识别能力
- 镜像：基于网卡的识别功能，实现业务级百G流量镜像
- 性能：智能网卡实现流量卸载，性能提升20~30%

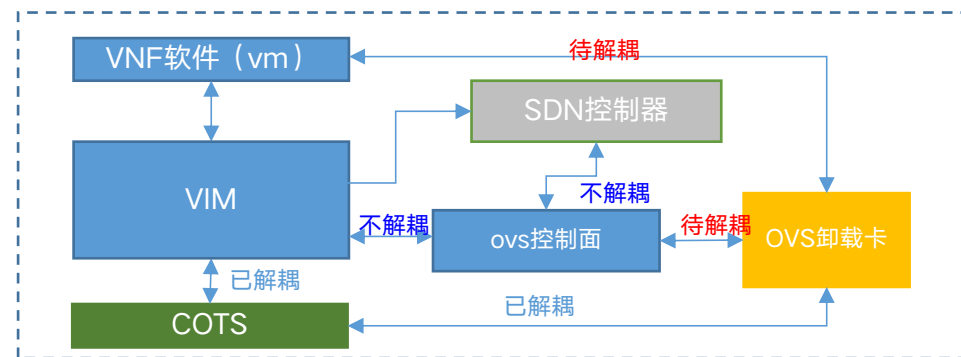


运营商网络引入智能网卡面临的挑战

➤ 标准化待成熟，引入面临解耦压力

● 仅OVS转发面卸载，涉及两个接口解耦

- 标准化vDPA框架实现virtio 1.0标准接口的前后端协商
- 标准化网络功能及对应rte_flow标准接口



- 运营商**多云多业务场景需适配不用类型智能网卡**，随着逐步引入存储功能卸载、业务功能卸载、运维功能卸载后，会迎来越来越多的异厂家、异构智能网卡及解耦场景

➤ 集成度、灵活性及可靠性取舍



THANKS