

2021 中国智能网卡研讨会

CHINA SMARTNIC WORKSHOP

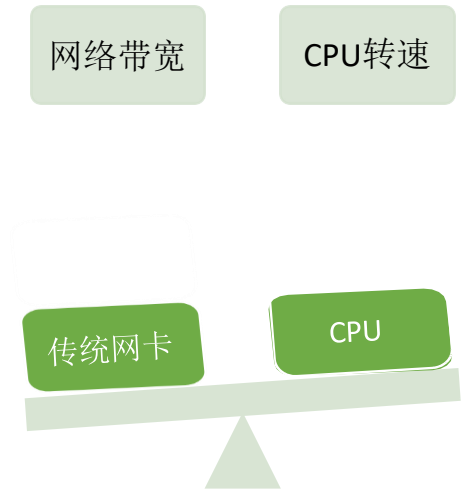
锐文科技 在智能网卡上的探索

阎 燕

网络控制器 -- 智能网络控制器

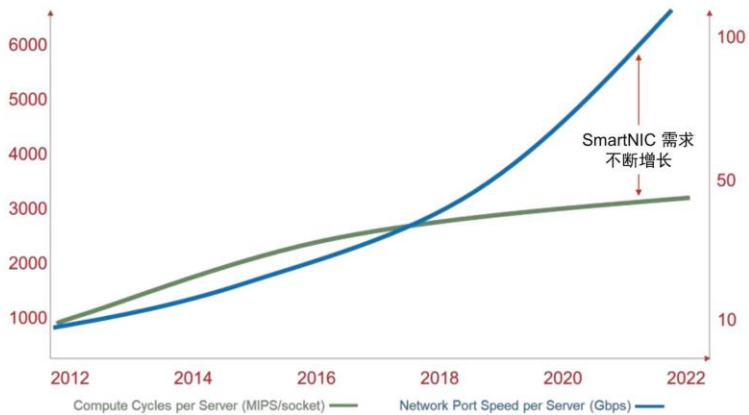
网络控制器

- 传统网络控制器及网卡
 - 外部-服务器/PC的数据通路
 - 网络数据的接收/发送
 - 端口速度在10Gbps及以下
- 传统网络环境，相对简单
 - 单机单系统



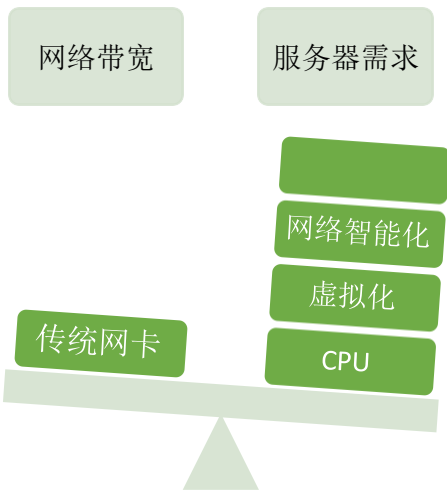
----->

端口速度超过摩尔定律

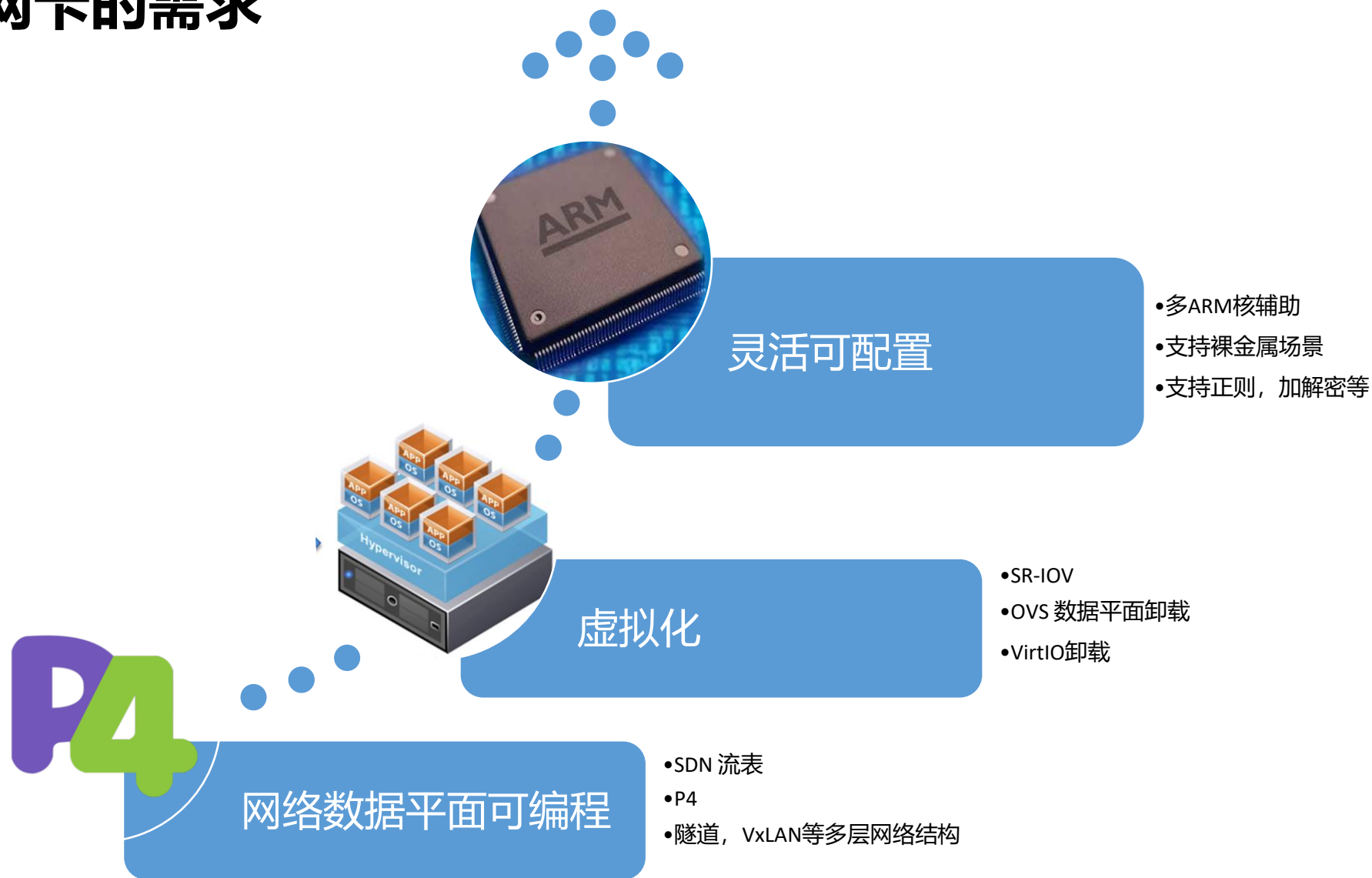


智能网络控制器

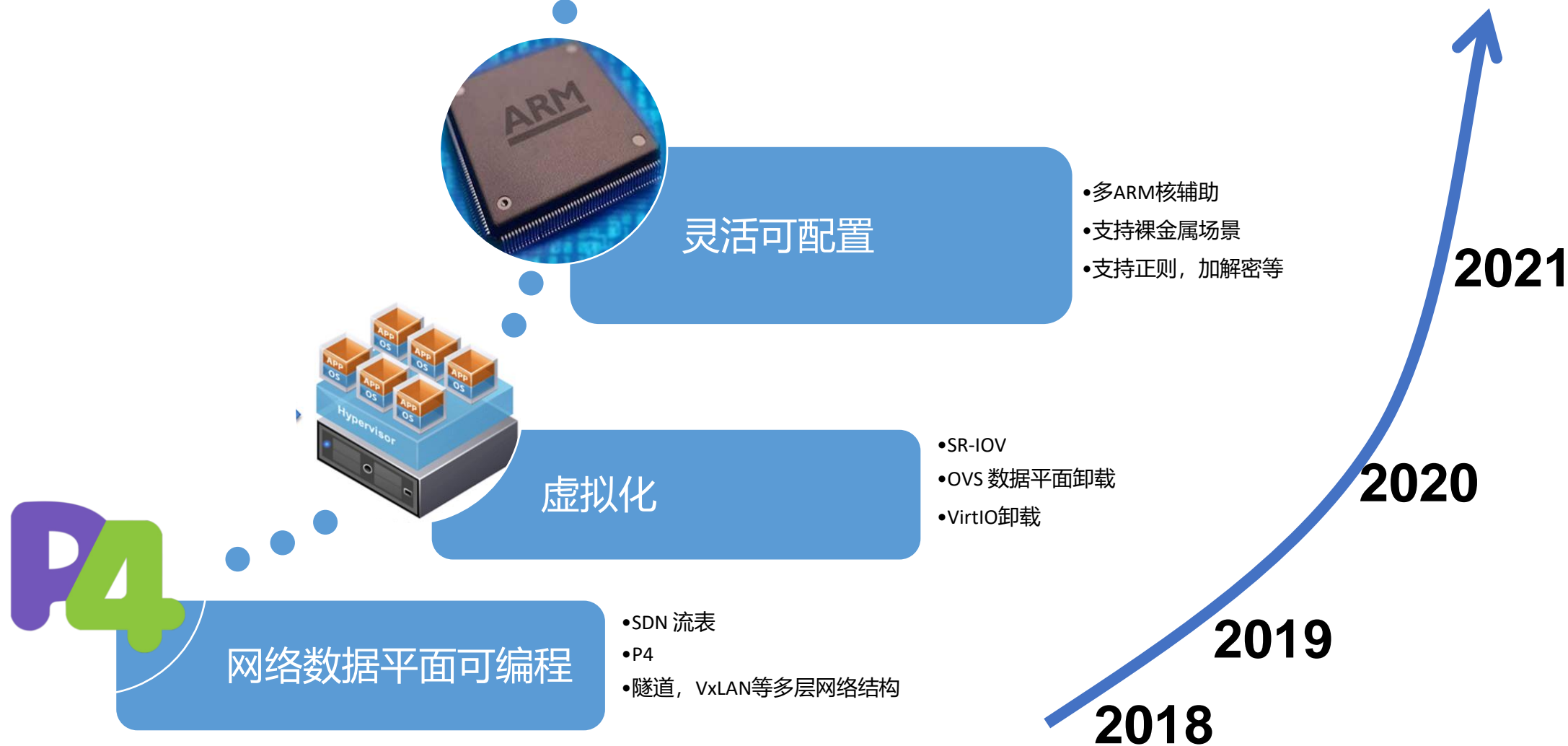
- 服务器虚拟化
 - 虚拟机，容器
- 网络智能化
 - 软件定义网络，网络功能虚拟化
- 数据量增长
 - 5G，物联网
 - 智慧城市，机器学习，AI



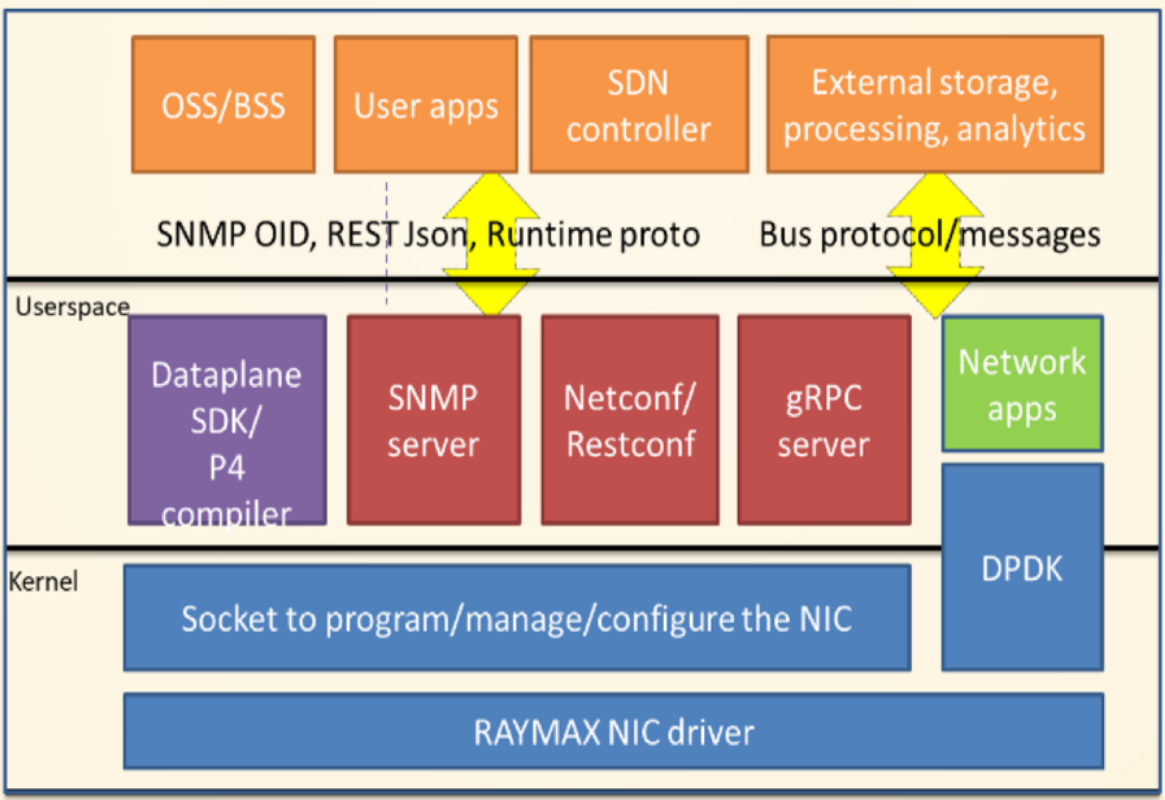
对智能网卡的需求



锐文科技在智能网卡上的探索



RAYMAX xSmartNIC 软件构架



Linux及DPDK驱动



多虚拟机的虚拟化平台

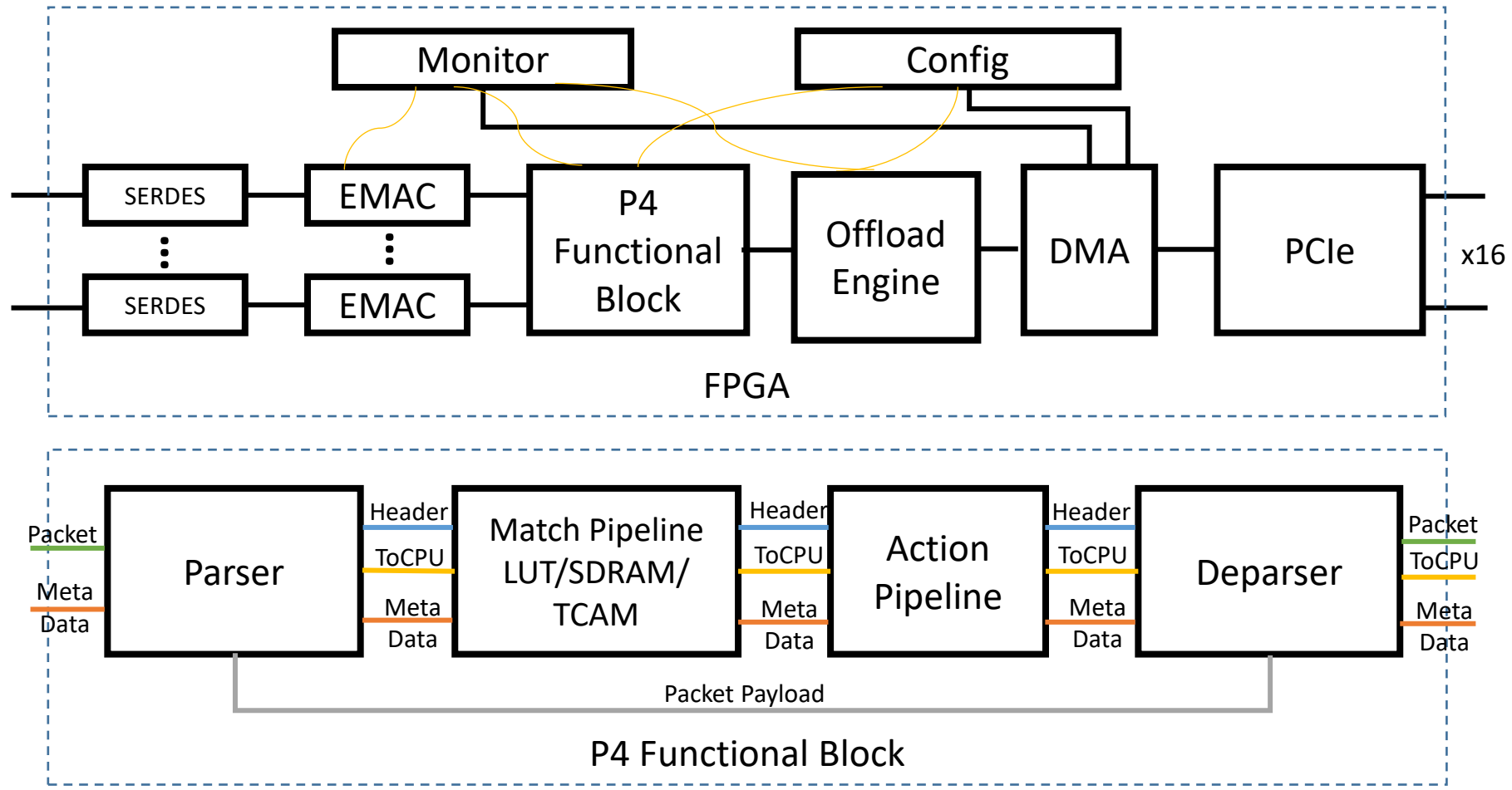


SDN, P4及传统网络接口



网路包/数据流监控

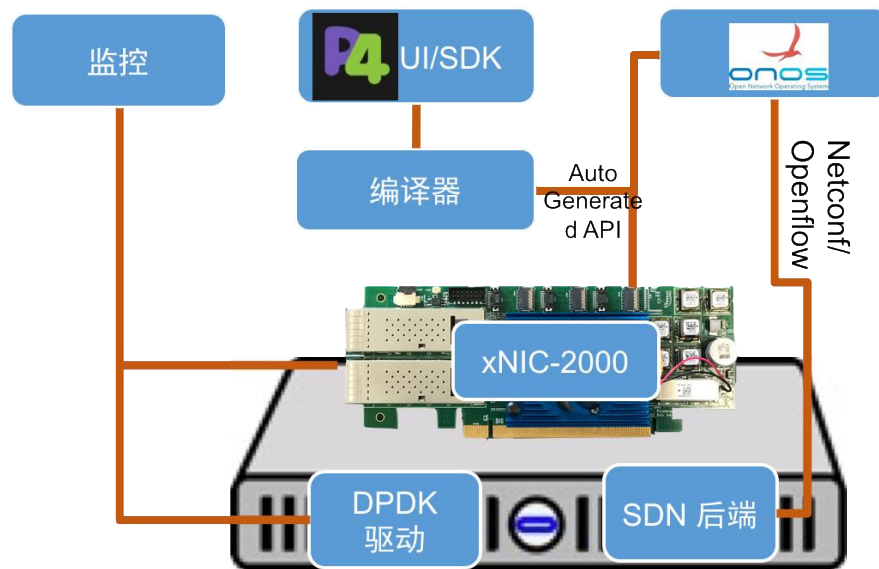
RAYMAX xSmartNIC 逻辑构架



支持网络数据平面可编程

➤ P4可编程性

P4提供了网络硬件的可编程性，用户可以根据需求灵活匹配网络包内容，并能对网络包进行修改。通过P4编译器，P4文件可以被编译成xSmartNIC支持的命令下发给驱动及网卡。



P4编译器和配置

RAYMAX

xNIC Management Center

admin

HEADER

xNIC Dashboard

xNIC Config

P4 Management

xNIC Config

Permission Management

Log Management

P4 Manage

New/Edit

Description : xnic_ipv6.p4

Content :

+ Upload P4 File

```
}  
  
header ipv6_t {  
    bit<4> version;  
    bit<6> dscp;  
    bit<2> ecn;  
    bit<20> flowlabel;  
    bit<16> totalLen;  
    bit<8> protocol;  
}
```

RAYMAX

xNIC Management Center

admin

HEADER

xNIC Dashboard

xNIC Config

P4 Management

xNIC Config

Permission Management

Log Management

P4 Manage

P4 Compiler

P4 Config

+ Save + Send xnic_ipv6_udp.p4

#					
1	hdr.ethernet.dstAddr :	xxxxxxxxxxxx	Mask:	11111111111111111111111111111111 (exact)	
	hdr.ethernet.srcAddr :	xxxxxxxxxxxx	Mask:	11111111111111111111111111111111 (exact)	Action : <div></div>
2	hdr.ethernet.dstAddr :	xxxxxxxxxxxx	Mask:	11111111111111111111111111111111 (exact)	
	hdr.ethernet.srcAddr :	xxxxxxxxxxxx	Mask:	11111111111111111111111111111111 (exact)	Action : <div></div>
3	hdr.ipv6.srcAddr :	xxxx:xxxx	Mask:	00000000000000000000000000000000 (lpm)	Action : <div></div>
4	hdr.ipv6.srcAddr :	xxxx:xxxx	Mask:	00000000000000000000000000000000 (lpm)	Action : <div></div>
5	hdr.ipv6.srcAddr :	xxxx:xxxx	Mask:	00000000000000000000000000000000 (lpm)	Action : <div></div>
6	hdr.ipv6.srcAddr :	xxxx:xxxx	Mask:	00000000000000000000000000000000 (lpm)	Action : <div></div>
7	hdr.ipv6.srcAddr :	xxxx:xxxx	Mask:	00000000000000000000000000000000 (lpm)	Action : <div></div>
8	hdr.ipv6.srcAddr :	xxxx:xxxx	Mask:	00000000000000000000000000000000 (lpm)	Action : <div></div>
9	hdr.ipv6.dstAddr :		Mask:		Action : <div></div>

xNIC Dashboard

Time : 2019-08-20 14:34:04 - 2019-08-20 14:49:04 Last 15 minutes 192.168.7.30-70:1d:08:01:e9:59

Latency 0.66us

TX-bps 75.03G

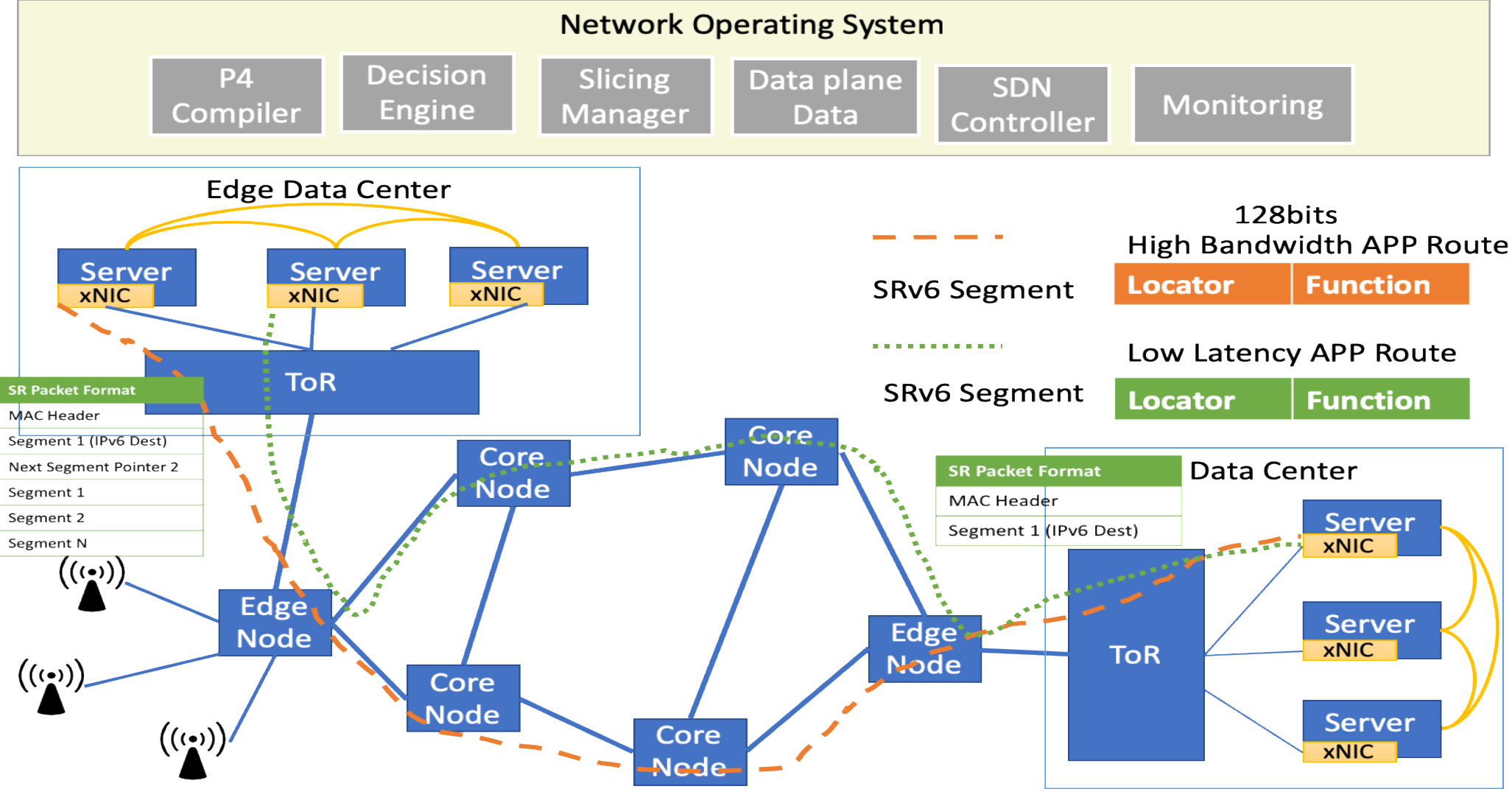
RX-bps 75.03G

Temp 70.6°C

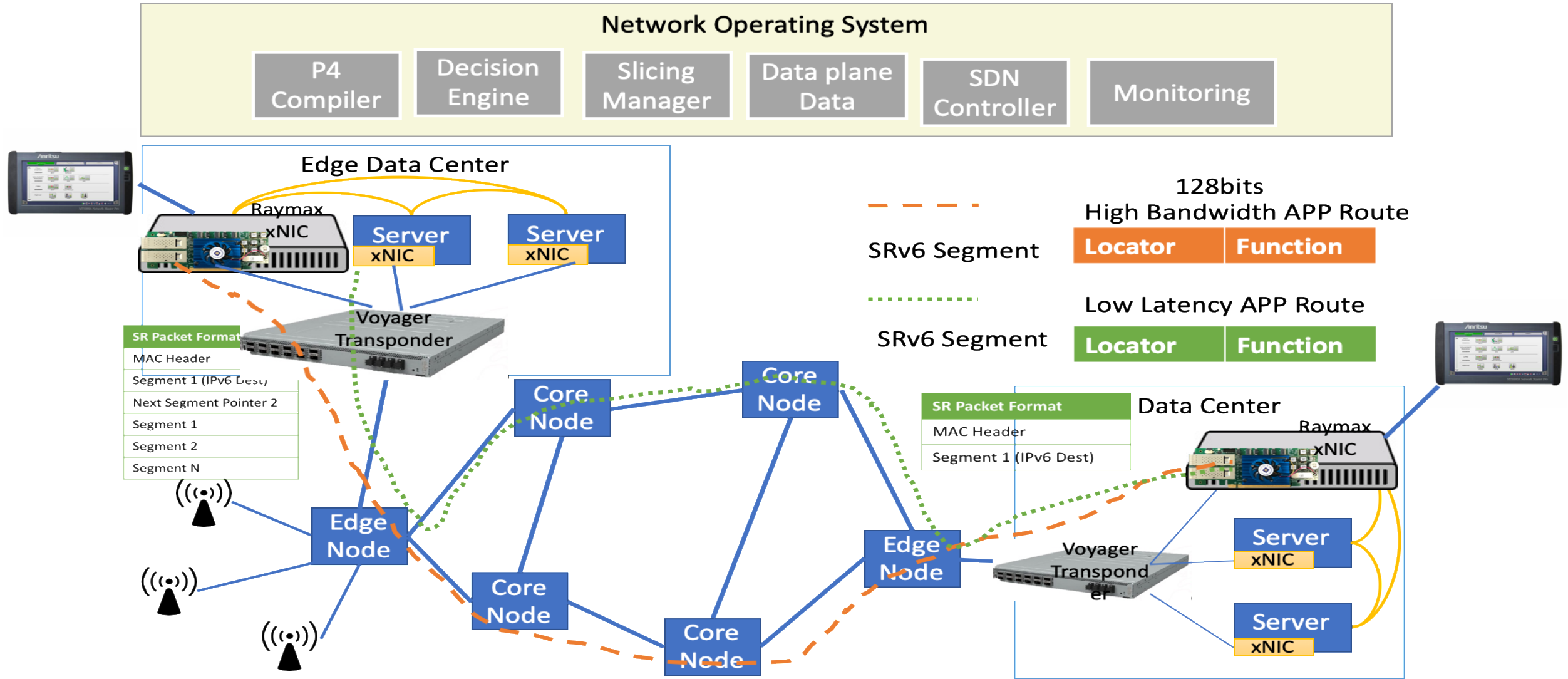
2021 CHINA

RAYMAX 锐文科技

应用案例：SRv6 Segment Routing



应用案例：SRv6 Segment Routing



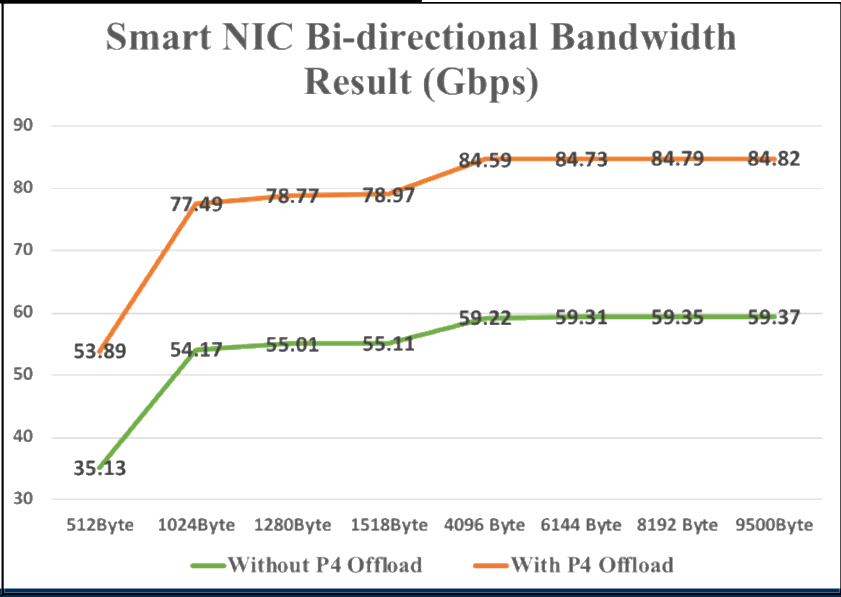
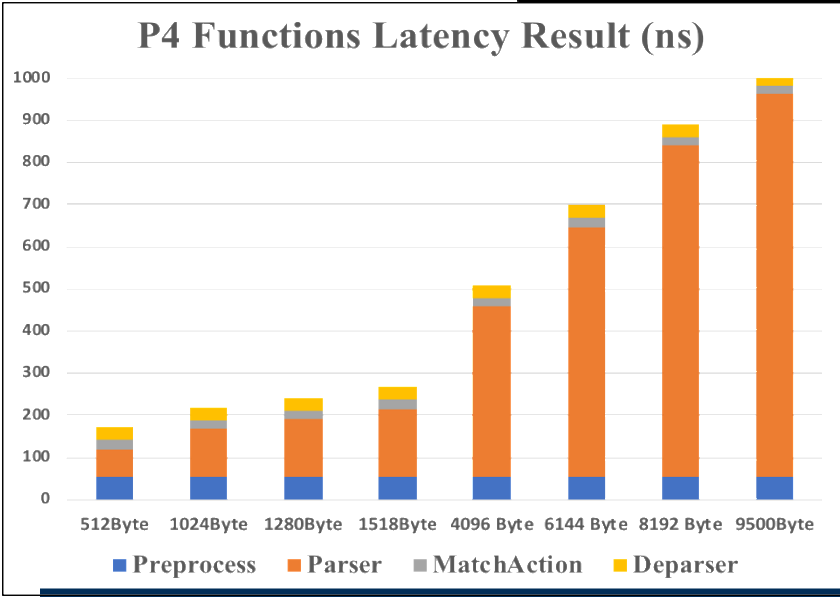
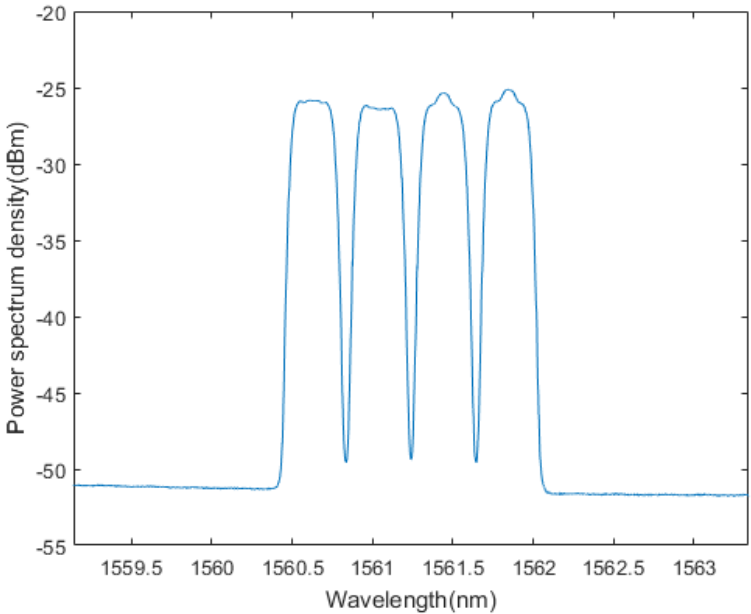
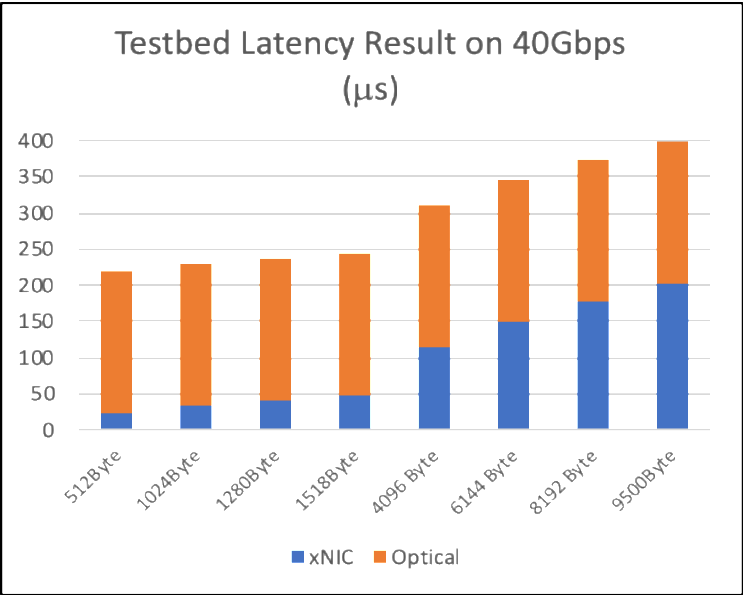
应用案例：测试结果

```
cumulus@cumulus212:mgmt-vrf:~$ net show transponder
Module: 1 ready Acacia Comm Inc. AC400-004-290 S/N:165354407 50.50C 11.84V
Laser: 191.15 THz - 196.10 THz, 6.00 GHz fine tune, independent lanes

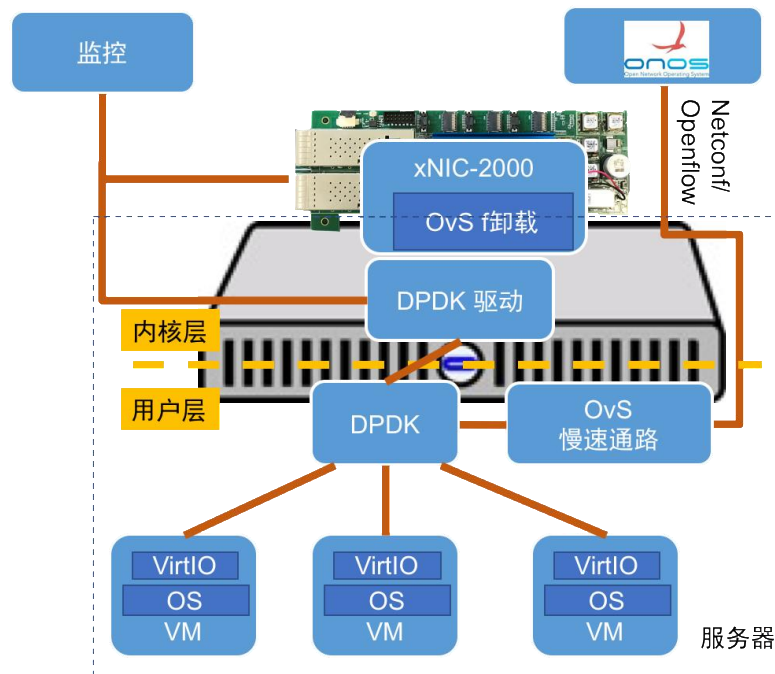
----- L3 -----
Network Interfaces
----- L4 -----
Modulation pm-qpsk pm-qpsk
Frequency 192.05 THz, Channel 19 192.10 THz, Channel 20
Current BER 3.249e-05 2.959e-05
TX/RX Power -4.00dBm/-21.06dBm -3.99dBm/-20.20dBm
Encoding differential differential
Alignment TX & RX TX & RX
Grid Spacing 50ghz 50ghz
FEC Mode 25% 25%
Uncorrectable FEC Errs 178226034398 171059678437
TX/RX Turn-up power_adjusted/locked power_adjusted/locked

Module: 2 ready Acacia Comm Inc. AC400-004-290 S/N:165156077 52.12C 11.84V
Laser: 191.15 THz - 196.10 THz, 6.00 GHz fine tune, independent lanes

----- L1 -----
Network Interfaces
----- L2 -----
Modulation pm-qpsk pm-qpsk
Frequency 191.95 THz, Channel 17 192.00 THz, Channel 18
Current BER 1.213e-05 3.057e-05
TX/RX Power -3.99dBm/-21.11dBm -3.98dBm/-21.17dBm
Encoding differential differential
Alignment TX & RX TX & RX
Grid Spacing 50ghz 50ghz
FEC Mode 25% 25%
Uncorrectable FEC Errs 122546779655 166530921285
TX/RX Turn-up power_adjusted/locked power_adjusted/locked
```



RAYMAX xSmartNIC 对虚拟化的支持



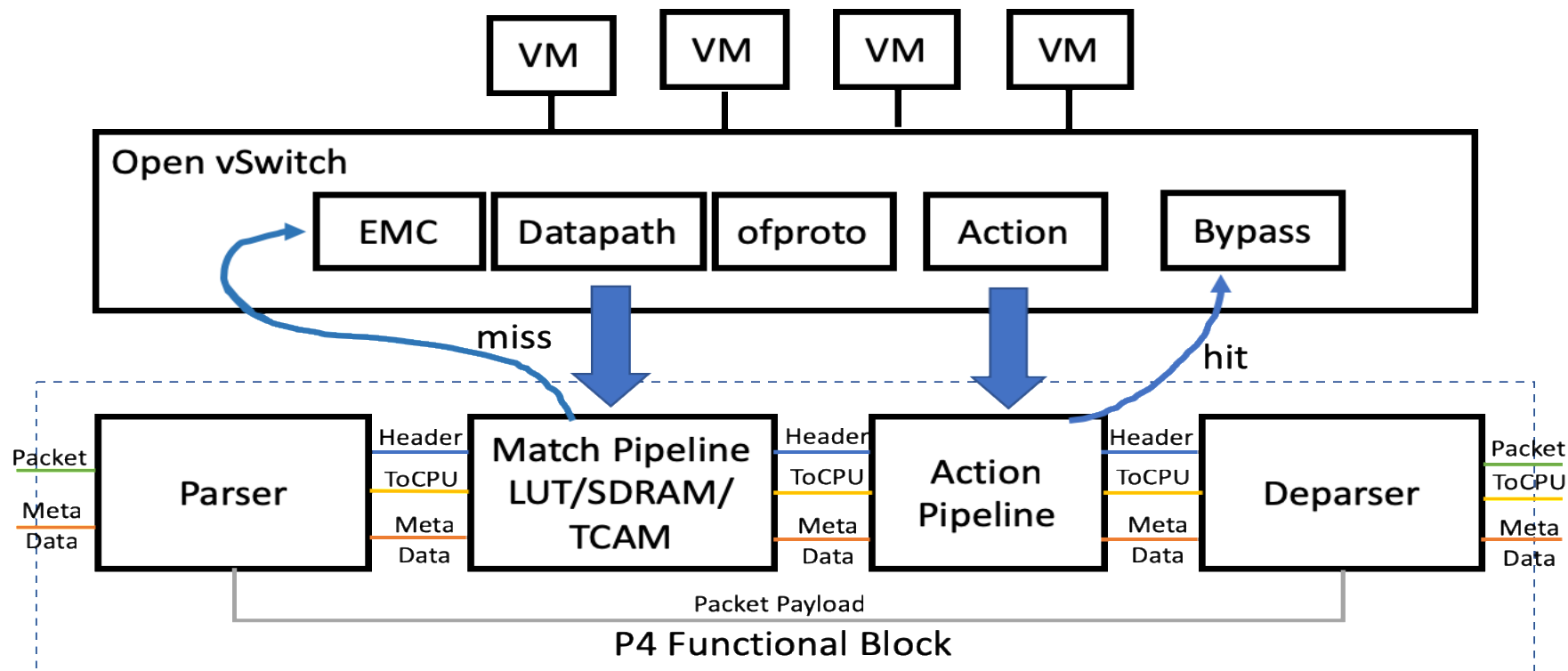
➤ SR-IoV

➤ Open vSwitch 卸载

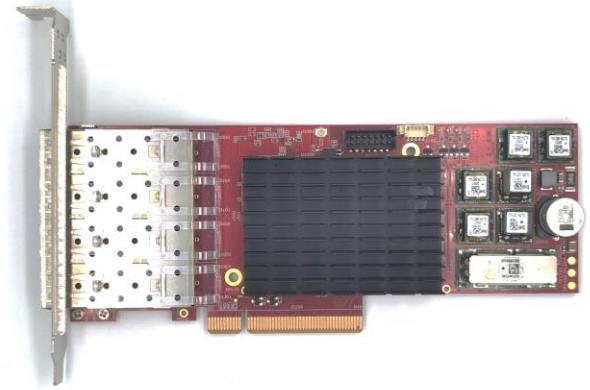
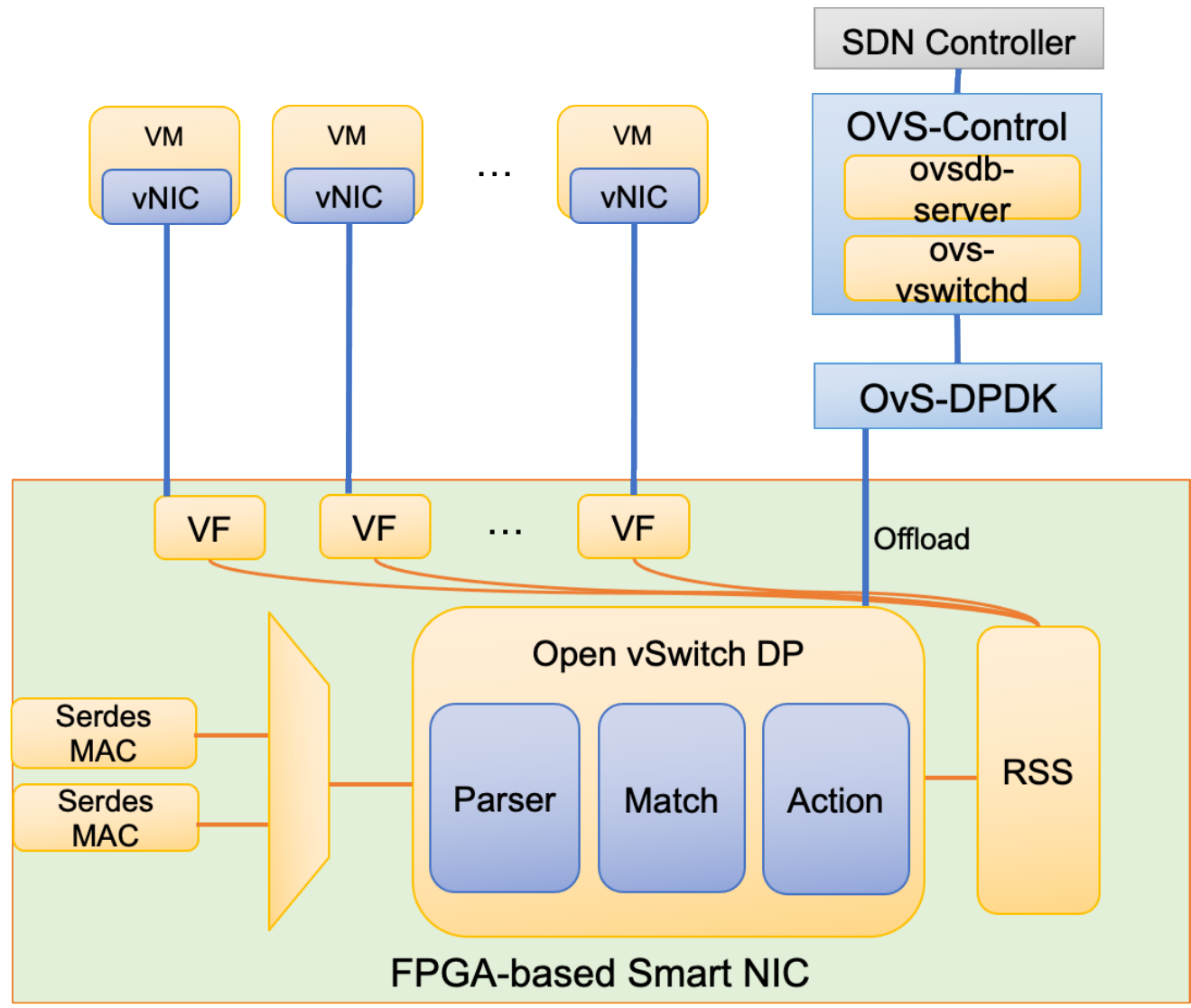
在多虚拟机环境下，Open vSwitch提供了虚拟机到虚拟机的软件层交换，然而当虚拟机数量增加，带宽需求增大，软件虚拟交换机OvS的性能会有很明显的下降，通过智能网卡的卸载应用，把耗费CPU大量资源的数据层查找卸载到智能网卡中，可提升超过30%吞吐量。

➤ VirtIO卸载：裸金属场景

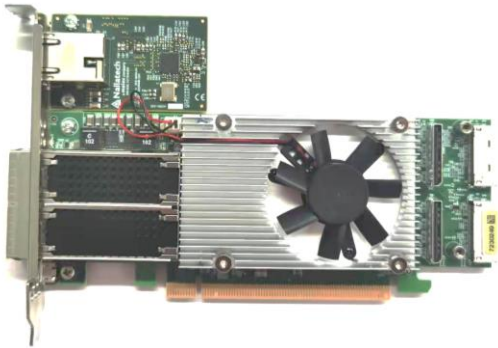
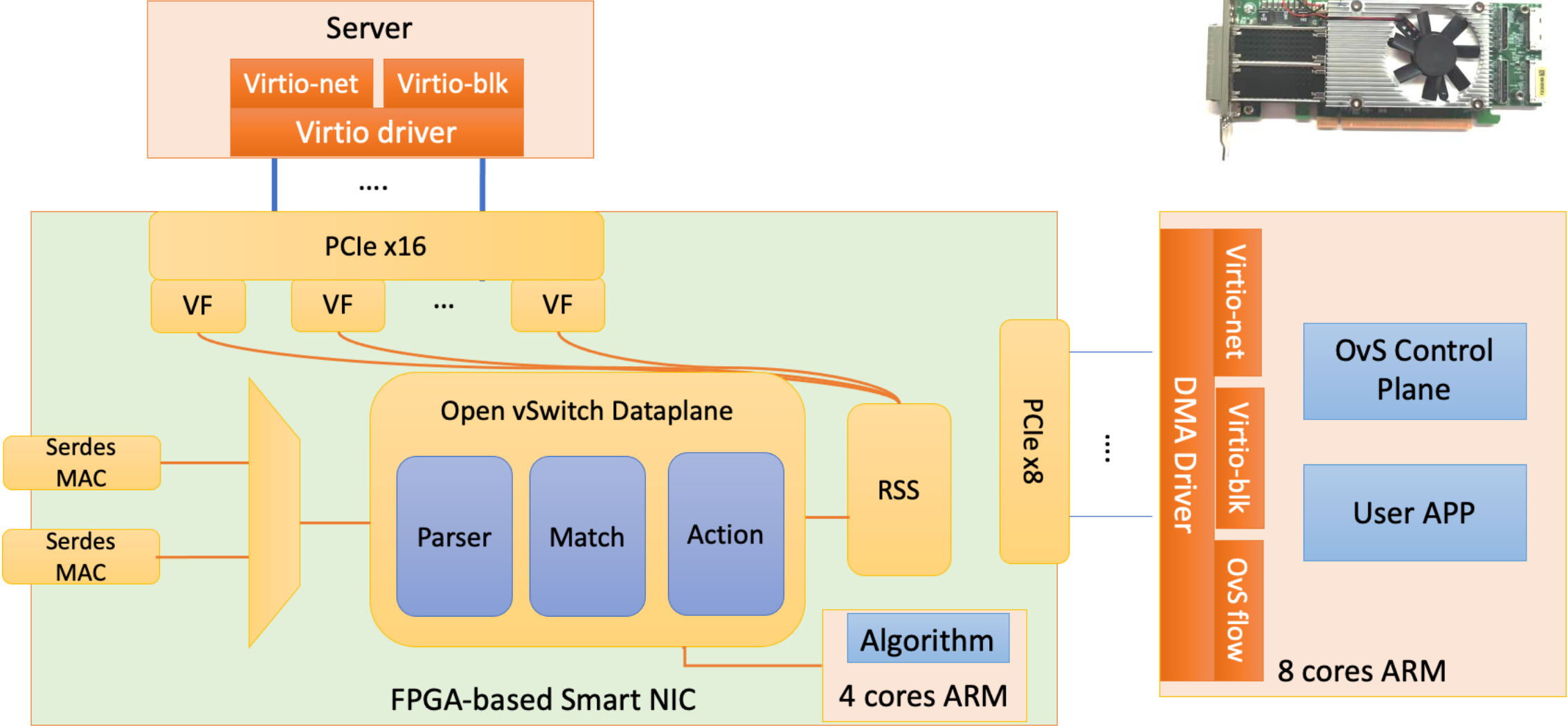
P4和OvS数据平面的逻辑结合



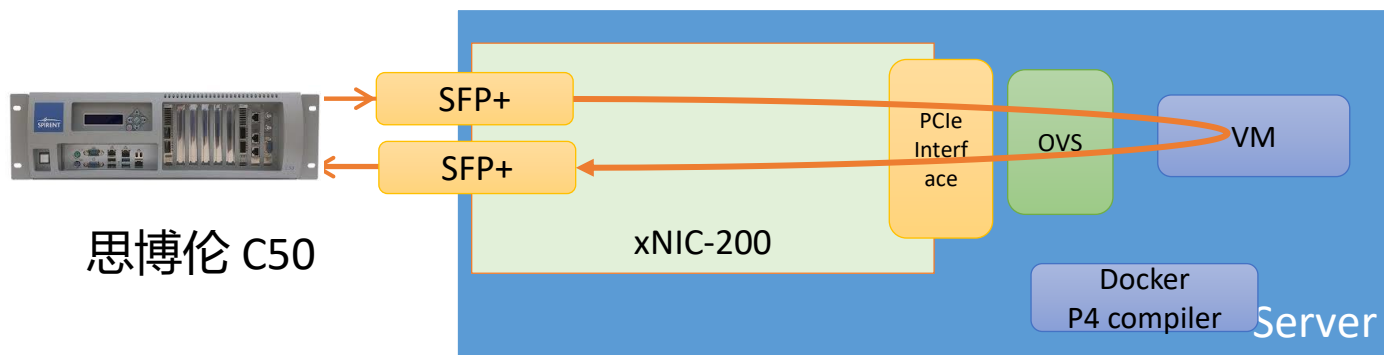
OvS数据平面卸载构架



OvS全卸载构架

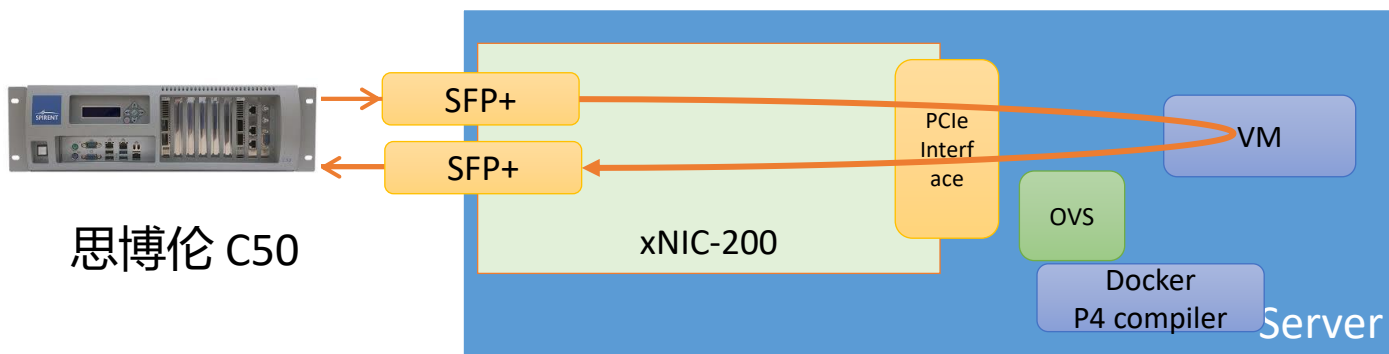


OvS卸载 vs Virt IO无卸载 --在国产飞腾CPU上性能比较



(a)

- a) Offload disabled
- b) Offload enabled

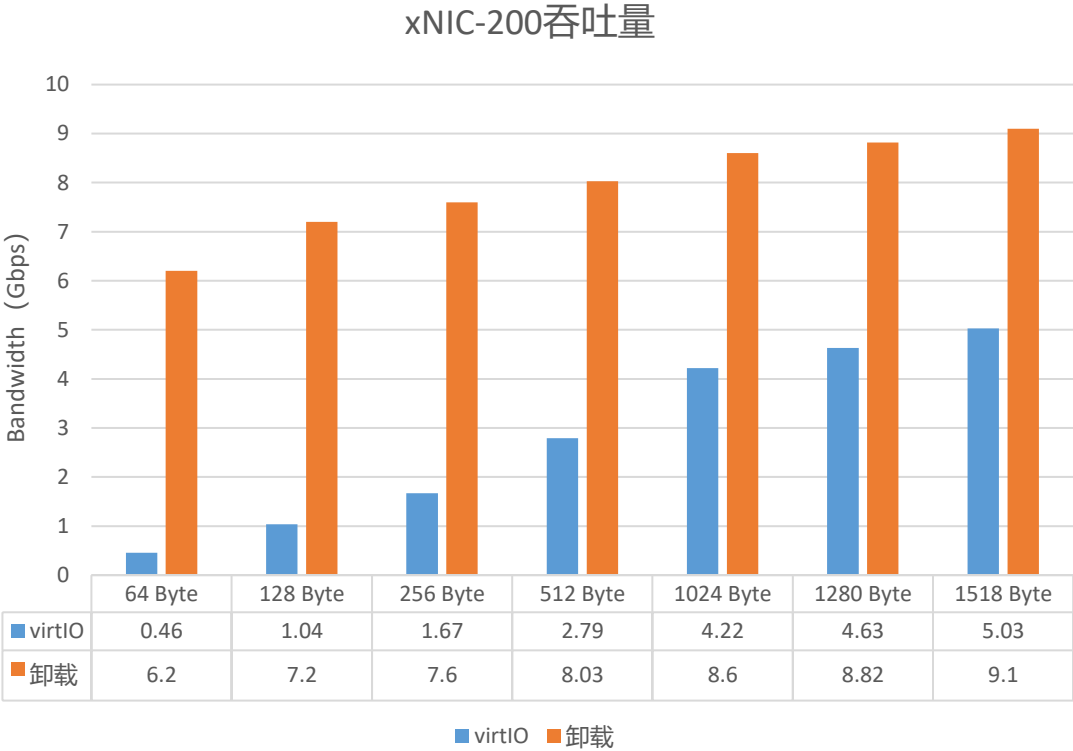


(b)

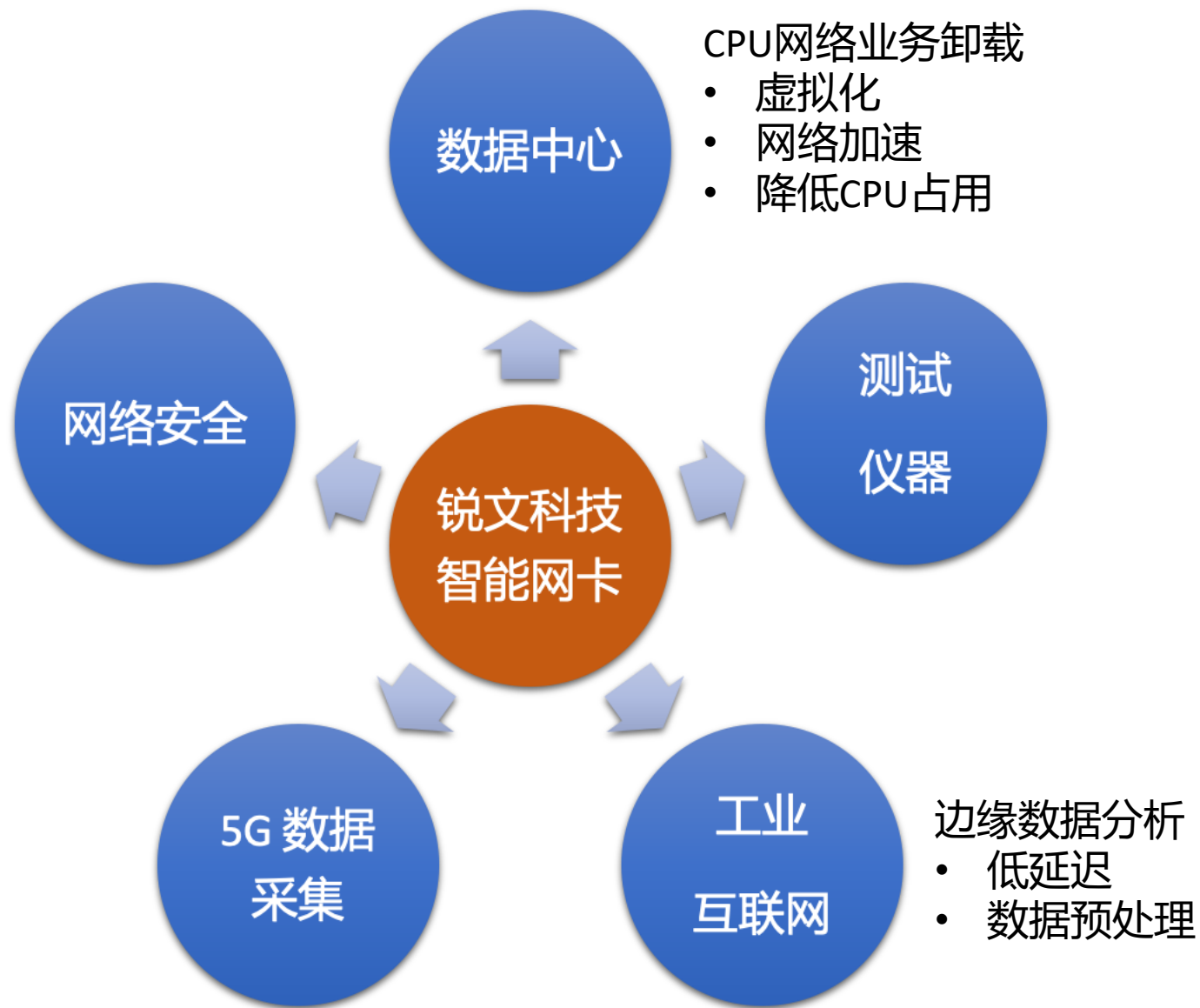
- DPDK testpmd

OvS卸载 vs Virt IO无卸载 --在国产飞腾CPU上性能比较

测试	测试2-锐文科技xNIC-200 10G 两口单核无丢包性能
服务器	飞腾FT2500 CPU服务器
CPU	飞腾FT2500处理器
系统内存	256G DDR4内存, @2666MHz
网卡	一个xNIC-200 10G 两口SFP+, PCIe3.0x8
操作系统	银河麒麟4.19.90-19.ky10
硬盘	250G SSD
DPDK版本	18.11.2
测试配置	1网卡, 2口 每口配置1个队列, 共2个队列, 每个队列分配一个逻辑核, 共2个逻辑核 testpmd



应用及市场



- 国产CPU加速
- 国产FPGA应用

RAYMAX 锐文科技

says

THANKS

