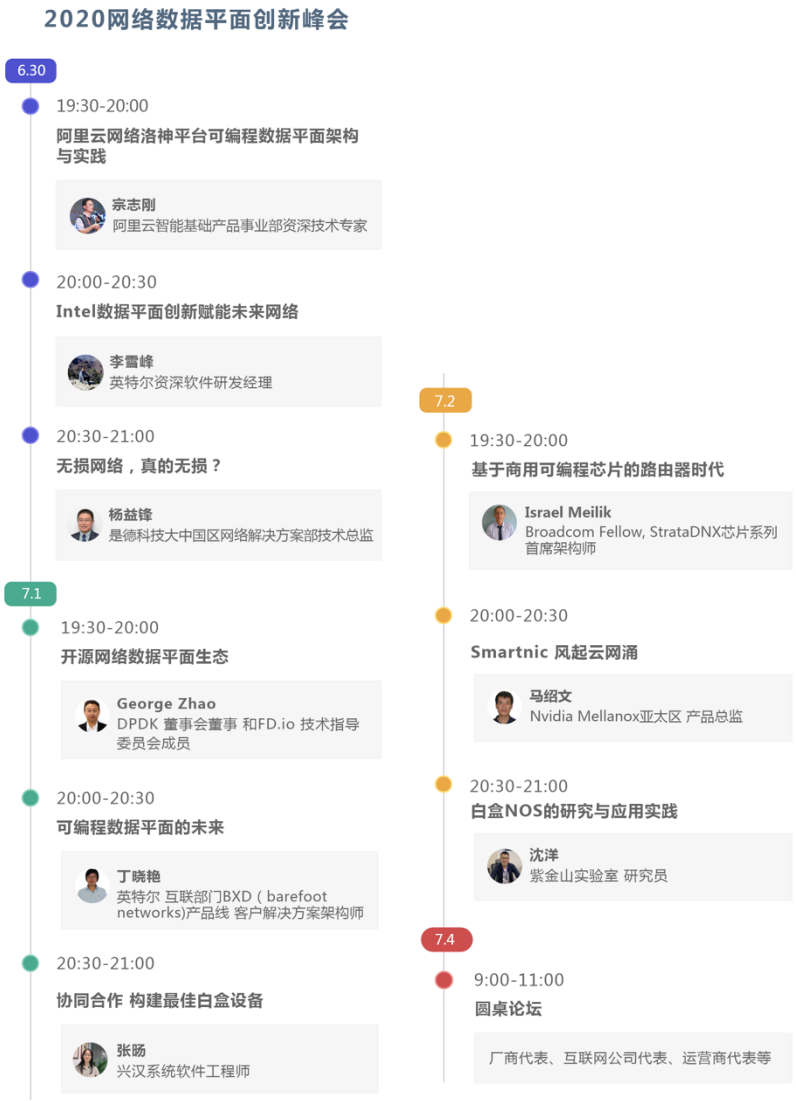
# 2020 网络数据平面创新峰会

**2020 年 6 月 30 日～2020 年 7 月 4 日**

## 1 会议主要内容

软件定义网络（SDN）旨在将网络设备的控制面与数据面分离，从而实现网络流量的灵活控制，使网络作为管道变得更加智能，为核心网络及应用的创新提供良好的平台，此为 SDN 的第一阶段。SDN 的第二阶段聚焦于数据平面，SDN 之父美国斯坦福 Nick Mckeown 教授提出的 P4 语言就是充分解放数据平面的编程能力。围绕网络数据平面，学术界和产业界开始了对交换机、芯片、网卡等硬件及网络可编程语言 P4 的研究与探索。

会议所有专题以及报告者信息如图 1




## 2 部分专题记录

### 2.1 《无损网络，真的无损？》——杨益峰

问题：什么是无损网络？为什么需要无损网络？无损网络的特点是什么？如何测量无损网络真的无损？

AI 时代的数据高效处理诉求中，传统数据中心网存在的问题：如图 2，IP 网络、存储网络、计算网络是分离的。IP 网络规模大，可用性好，带宽高，成本低；计算网络用的主要技术是 Infiniband，具有低时延、高可靠、高服务质量的特点，但是是专有的产品，私有的协议，兼容性不是很好，运维成本高；存储网络主要采用的技术是 SAN 和 FC，高可靠，但是带宽不高，升级不容易，成本昂贵。

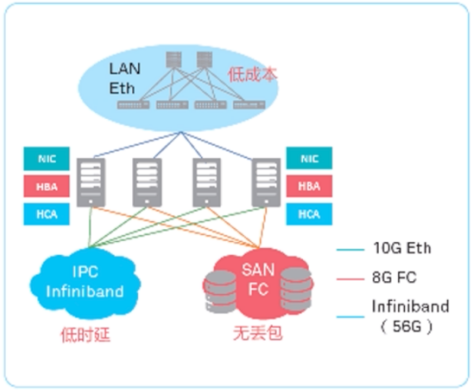



图 2 传统数据中心网

在 AI 智能无损网络中，如图 3，将计算、存储和网络融合在一起，可满足 AI 需求的无损网络，特点是低成本、无丢包、低时延。采用的主要技术是 RoCE(RDMA over Converged Ethernet)。

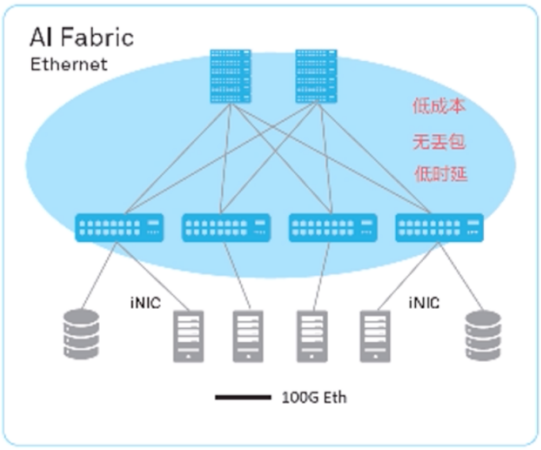



图 3 AI 无损网络

其中关键来看 **RDMA**(Remote Direct Memory Access)的，如图 4，传统的 TCP/IP 网络存在来自协议栈处理的高时延问题，且需要主机 CPU 参与多次协议栈内存拷贝，网络规模越大会导致 CPU 持续高负载。RDMA 的内核旁路机制允许应用网卡之间的直接数据读写，可以将服务器内的数据传输时延降低至接近 1 微秒；并且 RDMA 的内存 0 拷贝机制允许接收端直接从发送端的内存拷贝数据，极大减少了 CPU 的负载。所以 RDMA 具有低时延、降低 CPU 负载和高带宽的优势。

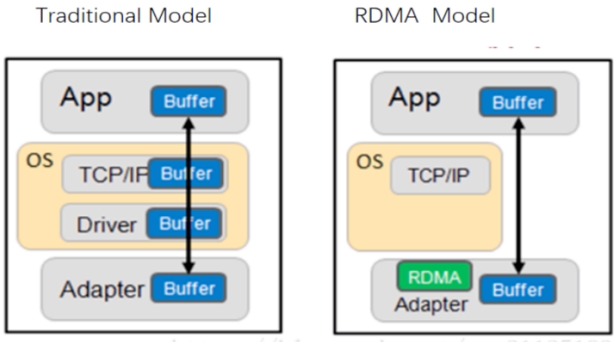



图 4 RDMA

无损网络的发展，从 InfiniBand，是基于 IP 网络的，成本高，兼容性差；继而出现了 RoCE，以及现在使用的 RoCE v2，如图 5.

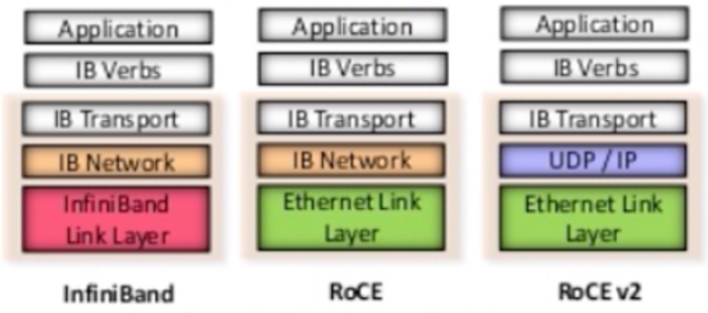



图 5 无损网络的发展

**无损网络的特点是无丢包、低时延、高吞吐**，那么怎么测试无损网络满足这些特性呢？通常用于高带宽、上千台服务器的大规模网络，如何模拟这样的真实场景来进行测试，财力人力以及网络的变动性都是挑战。

RoCE 有一系列协议来保证无丢包、低时延、高吞吐的特性。

1）DCB PFC，用于实现两种流量在以太网中共存时，存储流量无丢包，且对其他流量无影响

2）IP ECN，用于拥塞控制

3）IB Congestion Notification

4）NIC DCQCN

无损网络的具体测试方法如图 6.

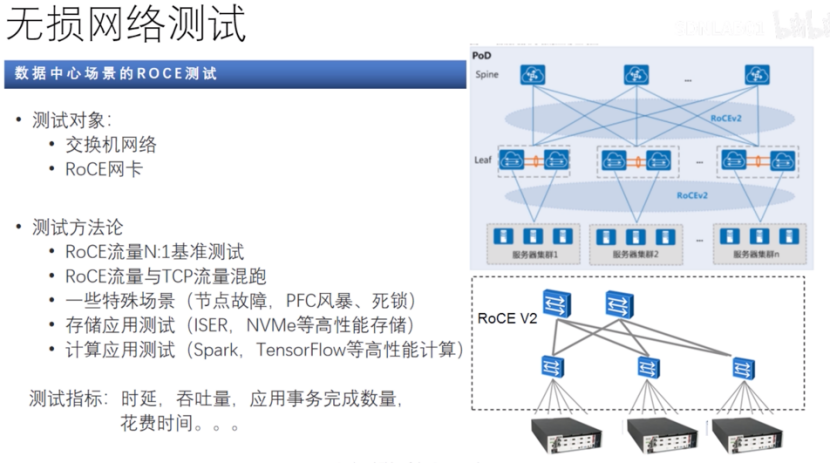



图 6 无损网络测试

（产品 Keysight，可模拟环境和协议，进行图 6 中的各种测试方法）

2.2 《开源数据平面生态》——George Zhao

George Zhao 联系方式：george.zhao@futurewei.com

微信：

这篇报告主要分享了，在 SDN 时代发展下的一些项目。

OpenDaylight：网络开源项目和生态的开始，将网络的开源推上了一个台阶，从个人开源进入到公司企业开源时代。

LFN：linux 基金会网络开源伞项目，包括 ONAP、OPNFV、OpenDaylight、Open Switch、Panda、SNAS，tungsten fabric 如图 7.（下面一半是非 LFN 项目，但也属于 Linux 基金会的项目）


图 7 开源项目

Panda：集成了多个开源项目的大数据分析平台，使用到了 kafka、Spark 等开源技术项目，将这些项目进行整合测试，提出了这样一个大数据平台框架，在应用的时候，就不需要自己去一个个测试这些平台和开源项目，可以直接用 Panda，他也提供了很多设计好的接口。

SNAS.io：Streaming Network Analytics System 流网络分析系统，收集、跟踪并存取多条实时路由对象的流网络分析系统，架构如图 8.

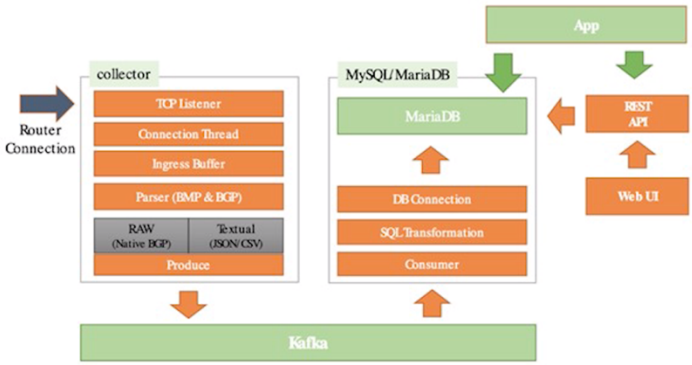



图 8 SNAS.io

ONF 基金会下的开源项目：STARATUM、TRELLIS、NG-SDN、P4、ODTA、OAOS。

OpenStack 是曾经最火热的开源社区和生态，而现在是 CNCF KubeCon.

数据面开源项目，如图 9。

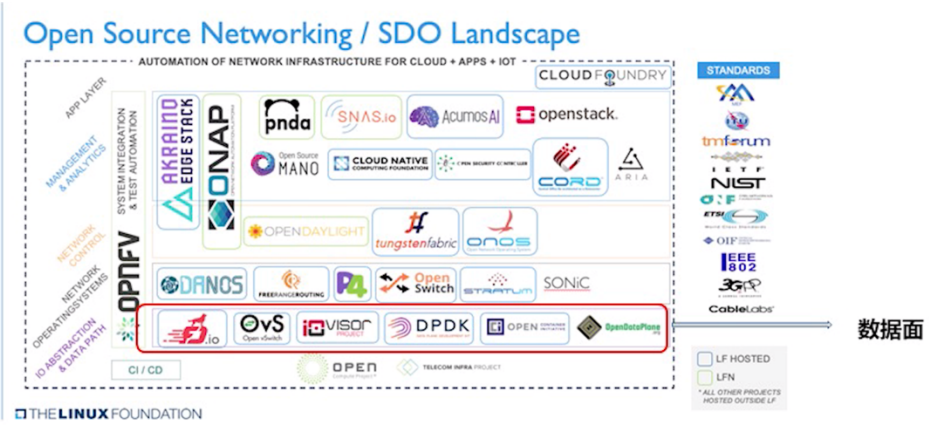



图 9

DPDK：节省了用户间和内核间上下文切换的损耗，使数据包能够更加快速的处理。

FD.io：如图 10，在 FD.io 最初成立的时候，DPDK 是 FA.io 的一个子项目。FD.io 也可以理解为在比 DPDK 高半层的位置。其中 VPP 是一个性能接近商业产品的软转发者，数据软转发是网络通信的基础。

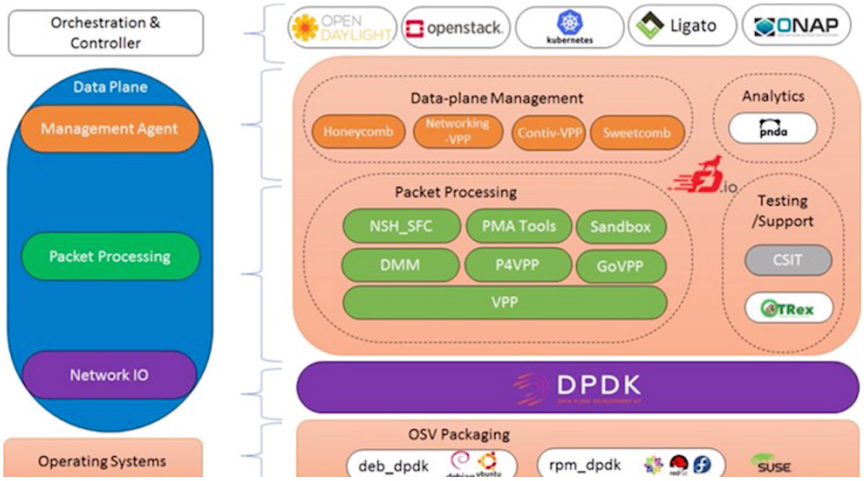


图 10 FD.io

IOVisor：和 FD.io 在同一个层面，做的事情也很类似，最大的区别是 IOVisor 使用的是 Linux 的内核态，FD.io 主要用的是用户态，IOVisor 主要用 XTP（内核数据包处理框架）和 EPDF 的技术。

NSM：Network Service Mesh 网络服务网格，如图 11，以云原生的方式呈现部署和管理网络数据平面，使用云原生的一个设计概念作为连接微服务的通信技术架构，服务网格在安全面，和控制面数据面分离。NSW 网络服务网格就是效仿了服务网格的做法，用云原生的设计方法，比如微服务、容器化等，来作为调控器，控制应用的生命周期，把数据面的网络服务，用云原生包装了一下。这是一个比较创新的方法，处于初期阶段。其实就是用更好的方式让 CT 和 IT 技术沟通

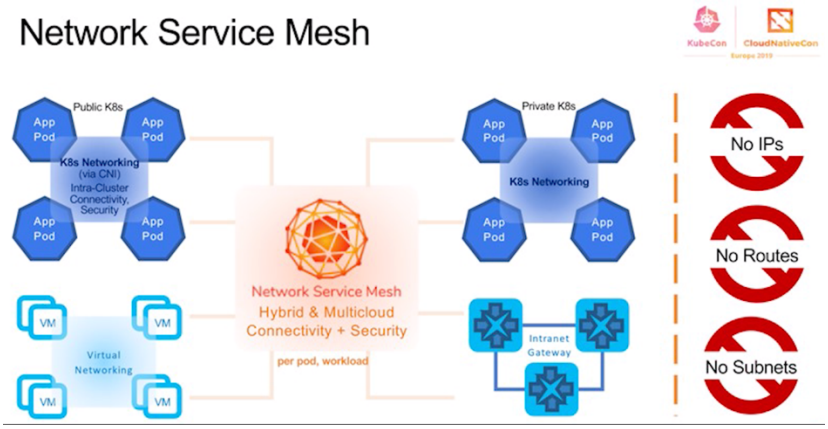



图 11

## 3 会议总结

通过听的两个报告，主要了解了目前 SDN 的一些发展近况，具体到有哪一些项目，每个项目在研究什么，处理什么问题，有什么优点，数据平面可以解决的问题是什么。就是可以对目前 SDN 数据平面的现状有一个快照式认识。

## 4 会议直播视频

https://space.bilibili.com/522278407/channel/detail?cid=138154

（bilibili 上 SDNLAB 的账户，相关会议的视频都会有）