

基于可编程交换芯片构建随需而变的云网络

传统交换芯片限制上层业务软件

计算机网络从软件转发时代发展到 ASIC（Application Specific Integrated Circuits，专用集成电路）芯片时代是一次质的飞跃。在 ASIC 芯片时代，网络的报文处理和转发逻辑全部被蚀刻在硬件芯片中，由完全线速的芯片逻辑完成，软件不再参与这些过程，从而让网络的性能得到突飞猛进的提升。

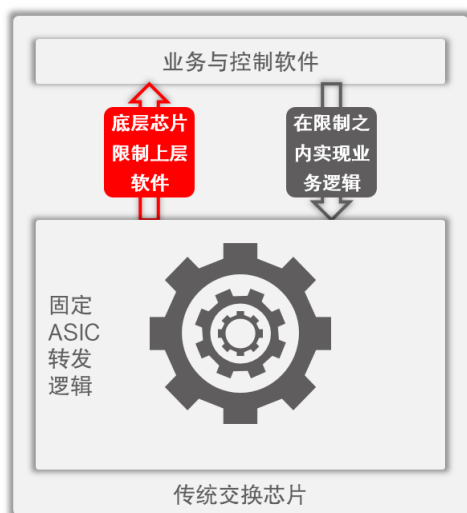


图 1：传统交换芯片限制上层业务软件

随着 IT 业务的不断发展，上层业务与控制软件对底层网络的特殊需求也逐渐涌现出来。例如，如果能够将一些类似于 NAT（Network Address Translation，网络地址转换）或 SLB（Server Load Balancer，服务器负载均衡）等常用的功能下移到底层交换芯片中实现，就能够将整个业务在网络上的处理性能大幅提升，并且大幅降低业务自身的部署复杂度，从而使得系统的整体效率得到提升。

遗憾的是，这样的需求在传统交换芯片时代是不可能被满足的，因为传统芯片的报文处理和转发逻辑在生产时刻全部被固化在硬件中，使用这样芯片的开发者是不可能芯片的工作过程中对其流程和逻辑做成任何更改的，而且，因为此类需求的复杂性与多样性，传统的交换芯片也不可能将这些功能集成到芯片中。如此一来，上层业务系统就只能在芯片能力所限定的范围内设计、开发，无法按需自行调整。

可编程交换芯片就是在这样的背景下应运而生的。

可编程交换芯片为上层业务软件解锁

顾名思义，可编程交换芯片与传统交换芯片最大的区别就在于其可编程性。与传统交换芯片将报文处理和转发逻辑固化在芯片硬件中不同，可编程交换芯片的报文处理和转发逻辑是能够通过软件来按需调整的。交换系统的设计开发者通过可编程交换芯片提供的高级编程语言，按照上层业务系统的需求开发芯片的报文处理和转发逻辑，然后将这样的逻辑下载到可编程交换芯片的处理流水线中，可编程交换芯片就可以按照这些定制的流程来处理网络流量和报文。

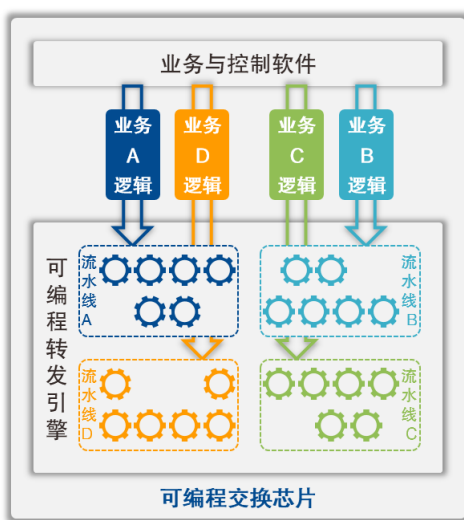


图 2：可编程交换芯片为上层业务软件解锁

可编程交换芯片带来的好处也是显而易见的。在以可编程交换芯片为核心的交换系统中，业务与控制软件不再受限于底层芯片的能力，可以根据自身的需求进行开发与定制；通过为不同的需求、不同的场景定制不同的报文处理和转发逻辑，芯片的各条流水线能够协同工作，在不损失系统整体性能前提下，将这些场景的需求在芯片的报文转发层面实现，从而使得系统整体的效率大幅提升。

当然，可编程交换芯片的硬件载体依然是 ASIC 与 FPGA (Field Programmable Gate Array, 现场可编程门阵列) 的结合体，集成在芯片中的硬件资源 (片载内存、流水线等) 毕竟是有限的，因此也就不可能无限地将所有的网络功能全部在芯片中编程实现。但是，在当前的网络应用环境中，将一些常见的网络功能转移到芯片中以可编程的形式实现已经是绰绰有余了。

从其实现逻辑和所达成的效果来看，可编程交换芯片不仅是一次交换芯片硬件技术的发展，更是 SDN (Software Defined Network, 软件定义网络) 这一理念在支撑网络的硬件芯片层面的一次伟大革新。可编程交换芯片让网络在保持高性能的前提下，前所未有地接近了软件定义这一未来发展趋势。

基于可编程交换芯片构建随需而变的云网络

在云网络中，构成底层网络的交换机其实在不同的位置上扮演者不同的角色，云网络其实也对这些不同的角色

提出了不同的需求。

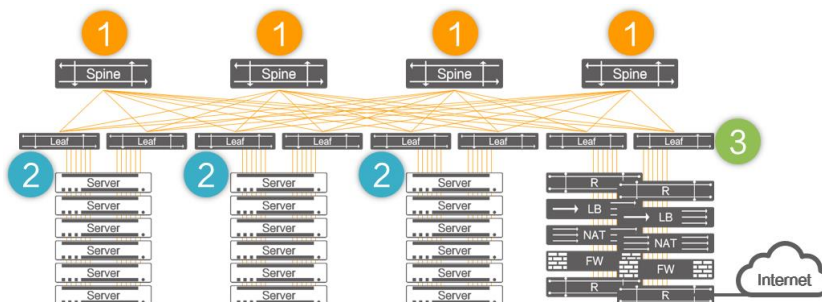


图 3：云网络中的交换机扮演的不同角色

图 3 是一个典型的 Spine-Leaf 架构的云网络。在这个云网络中，所有的交换机其实被分为三种不同的角色（在实际环境中，因为 Leaf 交换机往往被部署在每个机架的顶部，因此也被称为 ToR（Top of Rack）交换机）：

1. Spine 交换机，构成云网络的骨干交换机。
2. Server ToR，位于物理服务器机柜的顶部、将所有服务器接入云网络的交换机。
3. Gateway ToR，位于网关设备机柜的顶部、连接云与外部世界（Internet 或云的其他部分）的交换机。

显而易见，上述的三种角色在云网络的不同位置承担了不同任务，它们需要处理的流量的特征、所需要花费硬件资源的配比等都是不同的。在传统交换芯片构成的网络中，我们只能用一套固定的处理逻辑来应对这些不同的需求，但是在可编程交换芯片的网络中，情况就不同了。

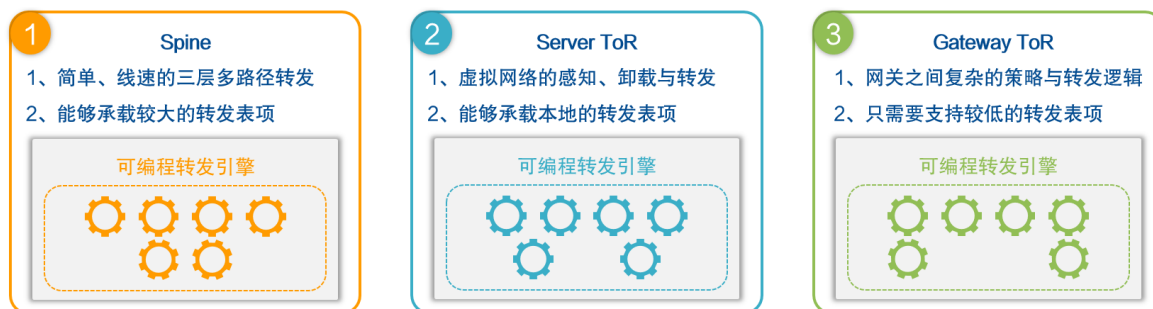


图 4：根据不同的场景动态加载不同的转发逻辑

可编程芯片能够为不同角色的交换机编制不同的转发逻辑、硬件资源分配配比：

1. 对于 Spine 交换机来说：最关键的任务在于提供简单、线速的三层多路径转发通道，同时，在 Asterfusion 的协议无限云网架构中，要求其能够承载较大的二层转发表项即可。
2. 对于 Server ToR 来说：最关键的任务在于提供虚拟计算节点世界的感知、虚拟网络的卸载和转发，同时，在 Asterfusion 的协议无限云网架构中，只要求其能够承载本地的二层转发表项即可。
3. 对于 Gateway ToR 来说：最关键的任务在于承载、处理网关之间复杂的控制、转发策略和大容量的路由

能力，同时，在 Asterfusion 的协议无限云网架构中，对于其二层转发表项的能力是非常低的。

当然，在上述架构中，我们完全也可以采用集中式或分布式的模型，将 NAT、SLB 等一些功能分布到 Gateway ToR 和/或 Server ToR 上。

(注：协议无限云网架构 (Protocol Infinity Cloud Fabric Architecture, PICFA) 是 Asterfusion 为下一代云网络提出的创新网络架构。PICFA 采用独创的分布式路由算法和与之相配合的转发逻辑，完全重构了云网络的控制平面与数据平面，彻底抛弃了传统网络中低效的集中式存储结构与转发逻辑，将云网络对云中虚拟计算节点的容量支持一举提升 100 倍至千万量级，同时大幅提升转发性能，使网络不再成为云计算容量的限制因素，从而为云网络从计算空间向底层物理网络的迁移打下坚实的基础。请参考其他相关文档以了解 PICFA 的细节)

Asterfusion 采用可编程交换芯片架构云网络交换机产品

Asterfusion 的全线云网络交换机产品，均采用可编程芯片架构。基于 Asterfusion 的云网络交换机，云计算的运营者能够轻松搭建软件可编程、随需而变的云网络，为云中业务构建一个灵活敏捷的网络平台。



=====

关于星融元数据

星融元数据技术有限公司（Asterfusion Data Technologies Co., Ltd.）为云计算提供领先的、真正意义上的软件定义网络解决方案。凭借所拥有的专利技术，基于高度软件定义的整体架构、完全开放透明的操作系统、突破传统限制的硬件平台，星融元数据帮助用户重新定义云计算的网络基础设施，并且为云计算的使用者赋予真正弹性与超高性能的虚拟网络、为云计算的管理者提供纯粹的开放接口和自动部署调度、为云计算的开发者构建可编程和可视化的业务环境。

星融元数据的核心团队来自于中美两国多家知名的通信和 IT 企业，共执对客户需求的透彻理解、对产品技术的笃定热爱、对下一代云计算网络的美好梦想，源于中国、面向全球，致力于成为中国最优秀的云计算公司。星融元数据坚信云计算将为用户、产业与我们带来多方共赢的美好未来。

联系方式

苏州（总部）

苏州市工业园区星湖街 328 号
2 栋 B401
0512-62982976

北京

北京市海淀区宝盛南路 1 号
奥北科技园 20 号楼 207
010-62672668

西安

西安市曲江新区旺座曲江
C 座 2605
029-89834058

希望获取更多有关星融元数据公司、产品及解决方案的信息，请登录 www.asterfusion.com、或发送邮件至 sales@asterfusion.com、或关注我们的官方微信、微博。

“星融元数据”、“Asterfusion”、“ASTERFUSION”，及其徽标均为星融元数据技术有限公司在中国的商标或注册商标。其他所有商标为其各自所有者之财产。本文件所包含的信息可能会发生修改，恕不另行通知。未经书面许可，本文件所含内容不作为合同或许可证的一部分。

©2018，星融元数据保留一切权利。