

面向光传输系统的交换结构

作者：Gary Lee，Fulcrum Microsystems

交换芯片实现的细节根据支持特定业务需求而所选择的芯片的交换容量的不同而千差万别。

运营级光传输系统（OTS）的背板交换结构是承载 TDM 业务的，该业务已经存在了很多年，即使在目前面向 IP 数据业务的系统中仍以某种形式存在。当我们完全向基于包的世界转型的过程中，传统的基于 TDM 的 OTS 通过使用宽带宽的以太网交换芯片，演变成为更高效的背板结构，其中 TDM 数据通过包交换架构传输。

设计者需要知道 OTS 的要求如何映射到以太网交换芯片的功能中。这包括接口协议、端口数、业务分类、阻塞管理，以及如何在已有业务隧道方法中选择出包含特殊业务流的专有帧头。

这些交换芯片必须支持很多功能，也要尽可能的低成本。从客户的角度来讲，需要支持一系列的接口和协议，包括以太网和存储，例如光纤信道和基于以太网的光纤信道。客户业务必须有效地汇聚以通过光承载网传输，还要有业务层的性能保证。从运营商网络层面来讲，交换芯片必须为光传输网（OTN）或 SONET/SDH 提供宽带宽连接。

以太网交换芯片的要求

传统 OTS 背板采用基于信元的架构，在输入端将有限的包业务分割成信元，在输出端再重新组装成包。随着时间的推移，包业务的数量和链路带宽持续增长，TDM 背板架构进行分割和重组变得十分昂贵，导致需要在系统背板采用 TDM 和包的双重架构。

大多数系统设计者认为，双重架构仅仅是一个临时解决方案，逐渐出现的以太网架构可以同时支持包业务和 TDM 业务，从而减少了系统开支和复杂度。这种以太网架构才刚刚开始支持这些需求。

如果 OTS 包括了以太网汇聚功能，它必须将多个低速率端口组合到宽带宽链路中，支持例如 OTN 等的光传输协议。例如，64 个 G 比特以太网（GbE）端口可以通过 OTU3 映射，汇聚到动态有保护的 40G OTN 链路中。在这种情况下，应用单个多端口以太网交换机比采用多芯片更便宜。由于可能会有潜在的过量认购 40G 链路，这

种交换机需要支持高级的分类能力和提供排序功能，以实现用最小的带宽来保证业务监控和整形。

为了支持 TDM 流，以太网架构需要支持固定时延的无损操作。这是由于在传统 LAN 交换机中，帧不能被丢弃，输出去抖 FIFO 需要按照最大时延边界来确定容量。

令人惊奇的是，针对这些需求的解决方案可以从数据中心市场上借用，如图 1 所示。通过采用 IEEE 优先级流控（PFC）帧结构，TDM 业务可以在交换和流控制时，分别通过特殊存储分区区分出来，保证了 TDM 业务的无损。IEEE 增强传输选择（ETS）细化了输出调度，可以给不同业务等级分配最少的带宽。通过给 TDM 业务提供最小带宽，最大延迟也可以被保证。另外，业务整形可以对非 TDM 业务使用，以减小包交换拥堵。

交换机的存储结构也会影响 TDM 的性能。真实的输出队列共享存储提供了低时延抖动和可预测的队列延时。

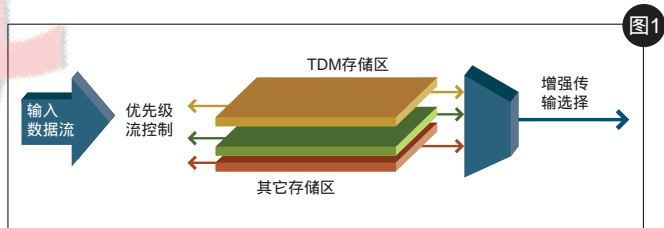


图1 TDM业务保护。

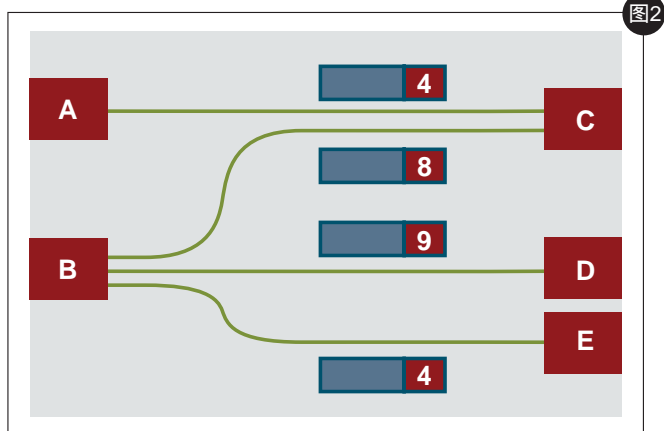


图2 TDM业务转送。

由于内部存储性能的限制，一些以太网交换机采用混合的输入输出队列(CIOQ)架构。这种架构不得不面对长时延、内部阻塞等问题，导致高时延抖动，需要更大的输出时间恢复 FIFO，从而成本更高。

业务隧道方法

直到近来，SPI 4.2 一直是 10G TDM 业务的标准信道化接口协议。然而这个接口对于多端口设备来说过宽，不能直接通过背板连接。Interlaken 逐渐成为 SPI 4.2 的替代品，然而大的以太网背板架构并不包含 Interlaken 接口。因此，TDM 映射器和成帧器供应商开发了增强的 XAUI 标准，例如 SPAUI 或者扩展 XAUI，包含带内信道信息和信道化流控制。这些方法对于常常包含 XAUI 物理接口的以太网交换架构非常理想。

输入业务流在头中包括了信道编号。一般情况下，以太网头被用来减少带宽预算。以太网交换机必须解析这个信道编号以及输入端口号，并根据它将帧转发到相应的输出端口。

如图 2 所示，每个交换机输入端口都能使用一套独立的信道编号。如图所示，输入端口 A 和输入端口 B 都采用了信道编号 4，而交换机用 4 加上输入端口号，将它们映射到不同的输出端口。

下一代 OTS 交换机必须支持信道化的流控制。图 3 表示了帧不仅仅是按照信道编号被转发到输出端口，而且在帧管道内信道编号还指代不同的业务等级，映射到输出队列中。通过采用 ETS，每个信道都被指派了不同的最小带宽。

输出映射器可将信道化流控制信息发送回交换机。通过解析和 PFC 类似的性能，帧可以在输出业务等级 (CoS) 队列中，根据信道编号被暂停。这种帧管道可以基于信道编号和端口信息，将这些流控制帧转发到相应的输入映射器。如果多个信道都需要流控制，也会通过广播的方式发送到合适的交换机输入端口。

在光汇聚系统中，往往需要给光链路上的流打上标签用来区分。这可以采用传统的方法实现，例如 MPLS 或者虚拟专有线路业务 (VPWS)，也可以采用专有交换机间链路 (ISL) 标签。如果类似的传输系统在链路两侧都使用，这些标签的全局地址可以用来围绕两个分立的物理位置，创造一个大的虚拟以太网交换机。

融合的TDM——一个包架构例子

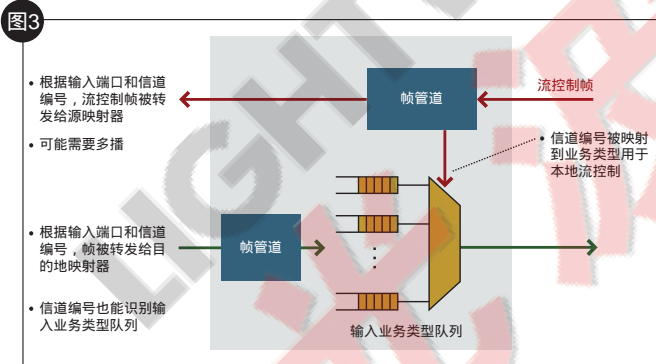
图 4 为一个光传输系统 100G 线路卡的例子。映射器目前已经可以实现单重、双重、四重的 10G 光模块接口。在另外一侧，它们可能包括 SPAUI 或者扩展 XAUI 接口到线路卡的交换芯片。这个交换芯片包含宽带上行到背板的多重交换卡，提供了 10T 比特的系统可扩展性。一些帧可能被转移到附加的 NPS 来缓存以实现更精细的业务管理，这个卡还可以将 TDM 和包业务转发到背板。

系统也可能包含传统的 SONET/SDH 线路卡，它必须包含连接到协议处理器的映射芯片，以实现在包架构中通过隧道传送 TDM 流。

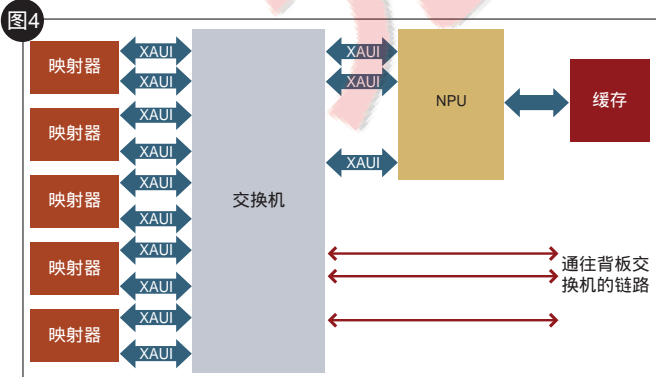
系统还可能包含以太网接入卡，它包含与背板高带宽连接的以太网交换机。这些线卡将标准以太网格式的业务转发到 OTS 线卡，在交换机中应用不同的存储分区，有效地扮演了分立虚拟交换机的角色。

如图 4 所示，当线卡收到包，映射器会在帧外加上 OTU 封套，用于光传输。在接收端，映射器去掉 OTU 封套，将帧再插入以太网虚拟交换机中，以标准以太网格式转发到接入线卡。

NPU 可以被用来做缓存、高精细度业务管理以及例外处理。在大多数情况下，软件可以合适地将 TDM 业务配置到 OTS 交换架构中。当业务暂时超出了映射器带宽或交换机带宽，业务又会被转移到 NPU 缓存中。



信道化转发和流控制。



100G OTS线卡例子。

了限幅放大器和 VCSEL 驱动芯片，适用于 10G 以太网、4G/8G 光纤通道，完全符合 SFF-8472 接口标准，支持 SPI 处理器接口规格。该芯片还可以支持对内部电压、工作温度、偏置和调制电流、发送功率等许多关键指标的监视功能。该芯片也可以配合 ESI-1010 芯片支持 FP/DFB 激光器工作。支持 VCSEL 时，ESI-1002 最高可以支持 28mA 的调制电流。整个芯片的功耗不超过 800mW。

ESI-1010 是单通道的激光器驱动芯片，可以支持各种不同 RC 时间常数的激光器工作。最高可以支持 100mA 偏置、50mA 调制电流。通过其分流结构，可以支持低阻抗的激光器。该芯片采用低功耗 CMOS 技术，功耗不到 50mW。

第二代四通道 TIA 和 VCSEL 驱动芯片

IPtronics 宣布其 IPTA12G4C TIA 芯片和四路 IPVD12G4C VCSEL 驱动芯片开始接受样品申请。这两款芯片将是 IPtronics 现有的 4 通道光模块芯片的补充。IPVD12G4C 和 IPTA12G4C TIA 功耗低到每对芯片 0.45W，支持 12.5Gbps × 4 速率，还具有多速率可调的功能。IPxx12G4C 芯片同时可以支持芯片到芯片通信、有源光缆、可插拔 QSFP+ 模块等不同应用。该芯片组完全符合 InfiniBand 以及以太网标准，内置 PROM 信息设置、温度控制、速率选择等功能，支持 I²C 和 RSSI 等接口。

10G EPON 芯片

Broadcom 推出了其第一款 10Gbps EPON 芯片 BCM55030。该产品主要针对多用户环境的 ONU 应用，也可以针对 LTE 网络的支撑网应用。目前该芯片已经接受样品申请，预计从今年 2 季度开始量产。BCM55030 主要来自他们收购的 Teknovus 公司的技术。Broadcom 表示他们的技术可以实现对同步数据信号以及对时间敏感的移动支撑网信号传输更好的支持。BCM55030 将高集成度、超低功耗、小尺寸以及经济性等结合到一起，对于电信运营商和 OEM 厂商的 FTTB 设备开发都能起到良好的作用。

50GHz 可调复用解复用器

CoAdna Photonics 推出了基于液晶技术的 50GHz 1 × N 可调复用解复用器 (TDMX)。CoAdna 号称这是业界第一个产品。该产品最多可以支持 20 个端口。该产品基于 CoAdna 专利的 LightFlow 平台，支持 hitless 波长可调、远程控制，具有极低的功耗。该产品可以用于波长无关、

方向无关、内容无关的上下路（分插）应用。

25GHz 波长选择开关

CoAdna Photonics 推出了基于液晶技术的 25GHz 间隔波长选择开关 (WSS)。该器件采用了该公司专利的 LightFlow 平台，具有优异的性能指标，支持可编程的带宽设置，支持 hitless 波长选择，并可以实现远程遥控。该器件还具有低功耗、低插损的特点。

深度存储的示波器

安捷伦科技公司提升了旗下 Infiniium 系列示波器的存储深度：全部 30 款型号目前均提供业界最深的标配存储深度和最深的存储深度选件。配有更深采集存储器的示波器具有以下两个优势，能够提供更深入的洞察力，从而帮助开发和验证团队快速将产品上市：一、相比存储器深度较浅的示波器，可以在固定采样速率上提供更长的采集时间；二、相比存储器深度较浅的示波器，可以在固定采集时间内提供更高的采集速率。

安捷伦提供 Infiniium9000、90000 和 90000X 系列实时示波器（带宽 600MHz 至 32GHz）的广泛选择。混合信号示波器和数字存储示波器目前配备 20Mpts 标配存储深度，是行业标准的两倍。数字信号分析仪目前配备业界领先的 50Mpts 标配存储深度。Infiniium9000 和 90000 系列示波器提供同类产品中最高的 1Gpts 的存储深度选件，Infiniium90000X 系列示波器提供高达 2Gpts 的存储深度选件。安捷伦的最深存储深度选件是竞争产品存储深度的 4 至 8 倍。

上接第13页

有时候也可能需要映射器和交换机配置更精细的业务管理。在这种情况下，交换机可以区分这些业务的类型，然后将它们转移到可以支持数以千计的 CoS 队列的 NPU 中。

最后，特殊的包也可以被映射器和 / 或交换机识别，送到 NPU 中处理。这包括以太网操作、管理、维护连接性检查和性能管理包，以及在进入和离开接入网时需要特殊安全处理的帧。

由于业务类型的转变和成本因素，光传输网正在向基于包交换的架构演进。先进的以太网交换架构通过提供同时支持包和 TDM 业务的功能，适应了这个角色。为了支持这个功能，可以开发支持映射器、NPU 和交换机之间直接连接的高效 OTS 线卡。[LWC]