

灵活以太网技术 白皮书 (2018版)



中国电信
CHINA TELECOM



1 概述

灵活以太网(FlexE)技术是基于高速Ethernet接口，通过Ethernet MAC层与PHY层解耦而实现的低成本、高可靠、可动态配置的电信级接口技术。该技术利用了业界最广泛、最强大的Ethernet生态系统，并且契合了视频、云计算以及5G等业务的发展需求，自2015年提出以来，受到业界广泛关注。



2 灵活以太网(FlexE)技术的产生

灵活以太网技术是在Ethernet技术基础上,为满足高速传送、带宽配置灵活等需求而发展的技术。

以太网概念由施乐公司于1972年首次提出,并基于载波侦听和冲突检测(CSMA/CD)技术逐步完善。自1980年代开始,Ethernet技术的发展完全遵循IEEE 802.3/1所制定的标准体系架构,并在产业技术与业务需求的共同驱动下快速发展,成为目前IT业界应用最为广泛、生态系统最为完善的L2互联技术。

Ethernet技术在接口层面遵循IEEE 802.3定义的MAC/PHY层标准,在2010年之前,基本按照“X10”倍速率发展,从10M-100M-1G-10G-40/100G发展。但是最近几年,随着业务需求与Serdes等技术的发展,Ethernet新增了25G-50G-200G-400G-800G的演进路径。而原有10M…100G路径也开始向100G-400G-800G方

向发展。以太网接口的发展路径如图1所示(参考文献[1])。

随着Ethernet接口技术的广泛应用,自2000年代开始,运营商城域网与广域网的Carrier Ethernet(电信以太网)技术得以发展与完善。Carrier Ethernet主要针对运营商网络的高可靠、可运行、可维护等需求,在MEF/IEEE/BBF等组织进行标准制定,从而使Ethernet技术具备了OAM、保护倒换、高性能时钟与QoS/QoE保障等电信级功能,广泛应用于城域网、广域网、移动承载网以及专线接入等场景。

近年来,随着云计算、视频以及移动通信等业务的兴起,人们对IP网络的诉求从以带宽为主逐渐转移到业务体验、服务质量和组网效率上。为满足上述需求,作为底层连接技术的Ethernet在保持既有低成本、高可靠、可运维等优势之外,还需要具备以下能力:

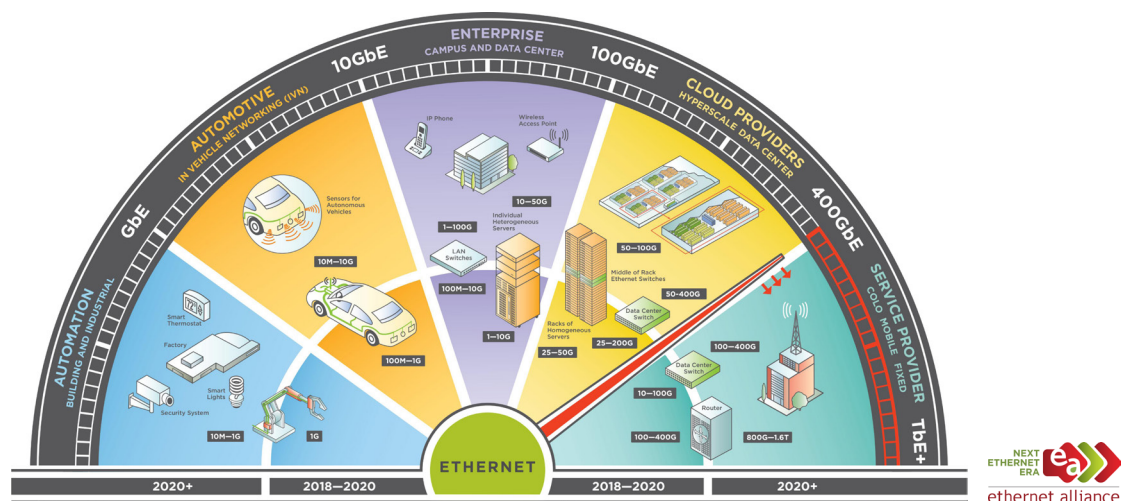


图1 以太网联盟发布以太网接口发展路径

- » 多粒度速率灵活可变：随着业务与应用场景的多样化，业界希望Ethernet接口可提供更加灵活的带宽颗粒度，而不必受制于IEEE 802.3标准所确定的10-25-40-50-100-200-400GE的阶梯型速率体系。业界甚至出现了800G、1.6T等超高速Ethernet接口需求，而这些接口标准尚未形成，需要寻求其它接口类解决方案。
- » 与光传输能力解耦：Ethernet接口能力与光传输设备能力发展并不同步。IP设备通过高速Ethernet接口组网时，经常受制于光传输网络能力。如果Ethernet接口速率与光传输网络速率解耦(即不需要光传输网络的DWDM链路速率与UNI接口的以太网速率保持严格的匹配)，就可以最大限度地利用现有光传输网络实现对新型超大带宽Ethernet接口的传输和承载。
- » IP与光融合组网：在Ethernet与传输能力解耦基础上，通过Ethernet与光传输网络之间的简单映射承载，简化网络，提高灵活性(这种场景可应用于大型IDC之间跨地域组网，也是FlexE技术最初提出的应用场景)，并进而实现流量灵活疏导与调度优化(参考文献[2])。
- » 面向多业务承载的增强QoS能力：多业务承载条件下用户体验增强，是高速Ethernet技术发展的重要关注点。Ethernet如果能在物理层接口上提供通道化的硬件隔离功能，就可以在物理层保证业务基于不同分片的隔离，进一步与上层网络/应用配合，结合高性能可编程转发以及层次化QoS调度等功能，即可在多业务承载条件下实现增强QoS能力。

灵活以太网(FlexE)技术也由此应运而生。



3 灵活以太网(FlexE)技术关键实现

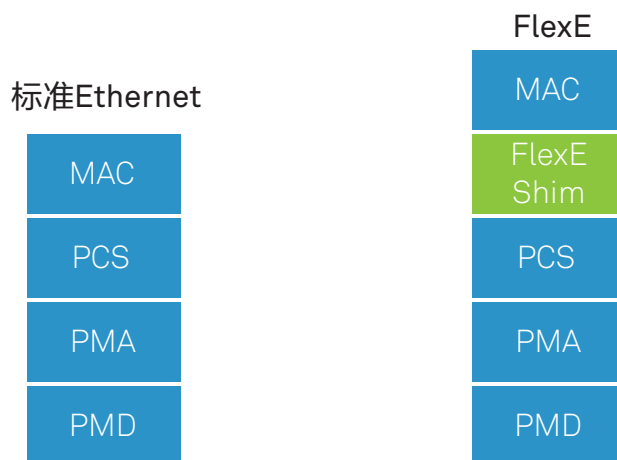


图2 标准Ethernet与FlexE结构

FlexE技术通过在IEEE802.3基础上引入FlexE Shim层实现了MAC与PHY层解耦(其实现如图2所示),从而实现了灵活的速率匹配。

灵活以太网基于Client/Group架构(如图3)定义,可以支持任意多个不同子接口(FlexE Client)在任意一组PHY(FlexE Group)上的映射和传输,从而实现上述捆绑、通道化及子速率等功能(参考文献[3])。其中:

- » **FlexE Client:** 对应于网络的各种用户接口,与现有IP/Ethernet网络中的传统业务接口一致。FlexE Client可根据带宽需求灵活配置,支持各种速率的以太网MAC数据流(如10G、40G、n*25G数据流,甚至非标准速率数据流),并通过64B/66B的编码的方式将数据流传递至FlexE Shim层。

- » **FlexE Shim:** 作为插入传统以太网架构的MAC与PHY(PCS子层)中间的一个额外逻辑层,通过基于Calendar的Slot分发机制实现FlexE技术的核心架构。

- » **FlexE Group:** 本质上就是IEEE 802.3标准定义的各种以太网PHY层。由于重用了现有IEEE 802.3定义的以太网技术,使得FlexE架构得以在现有以太网MAC/PHY基础上进一步增强。

以FlexE点对点连接场景为例,多路以太网PHY组合在一起成为FlexE Group,并承载通过FlexE Shim分发、映射来的一路/多路FlexE Client数据流。

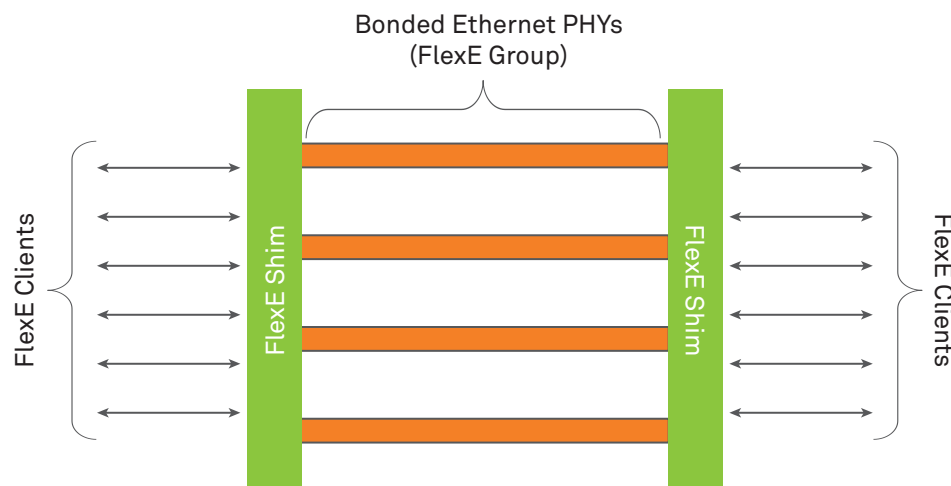


图3 FlexE通用架构

FlexE的核心功能通过FlexE Shim层实现，它可以把FlexE Group中的每个100GE PHY划分为20个Slot(时隙)的数据承载通道，每个PHY所对应的这一组Slot被称为一个Sub-calendar，其中每个Slot所对应的带宽为5Gbps。FlexE Client原始数据流中的以太网帧以Block原子数据块(为64/66B编码的数据块)为单位进行切分，这些原子数据块可以通过FlexE Shim实现在FlexE Group中的多个PHY与时隙之间的分发。

按照OIF FlexE 标准，每个FlexE Client的数据流带宽可以设置为10、40或者 $m \times 25$ Gbps。由于FlexE Group的100GE PHY中每个Slot数据承载通道的带宽为5Gbps粒度，FlexE Client理论上也可以按照5Gbps速率颗粒度进行任意数量的组合设置，支持更加灵活的多速率承载。

FlexE Shim通过Calendar机制实现多个不

同速率FlexE Client数据流在FlexE Group中的映射、承载与带宽分配。FlexE按照每个Client数据流所需带宽以及Shim中对应每个PHY的5G粒度Slot的分布情况，计算、分配Group中可用的Slot，形成Client到一个或多个Slot的映射，再结合Calendar机制实现一个或多个Client数据流在Group中的承载。具体到比特流层面(如图4所示)，每个64/66B原子数据块承载在一个Slot时隙中(此处Slot作为承载64/66B数据块的基本逻辑单元，可与图中的Block概念等同)。FlexE在Calendar机制中，将“20blocks”(对应slot0到slot19)作为一个逻辑单元(如图4中绿色数据块所示)，并进一步将1023个“20blocks”作为Calendar组件。Calendar组件循环往复最终形成了5G为颗粒度的Slot数据承载通道。

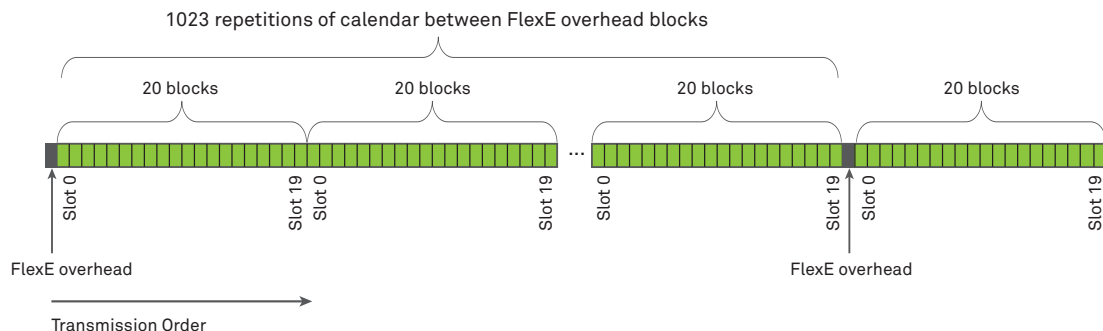


图4 FlexE帧结构示意图

FlexE Shim层通过定义Overhead Frame/MultiFrame的方式体现Client与Group中的Slot映射关系以及Calendar工作机制。FlexE Shim层通过Overhead提供带内管理通道，支持在对接的两个FlexE接口之间传递配置、管理信息，实现链路的自动协商建立。具体而言，一个开销复帧(Overhead MultiFrame)由32个开销帧(Overhead Frame)组成，一个开销帧则由8个开销时隙(Overhead Slot)组成。Overhead Slot如图4中黑色数据块所示，实际上是一个64/66B的原子数据块。Overhead Slot每隔1023个“20 Blocks”出现一次，但每个Overhead Slot中所包含字段是不同的。开销帧中，第一个Overhead Slot中包含“0x4B”的控制字符与“0x5”的

“O Code”字符等信息。在信息传送过程，对接的两个FlexE接口之间通过控制字符与“O Code”字符的匹配确定第一个开销帧，从而在二者之间建立了一个独立于图4绿色Slot的数据通道之外的管理信息通道，实现对接的两个接口之间配置信息的预先协商、握手等。例如，某个FlexE Client数据流在发送端的FlexE Shim/Group中的数据通道Slot映射信息、位置等内容传送到接收端后，接收端可以从数据通道中根据发送端的Slot映射等信息恢复该FlexE Client的数据流。FlexE的带内管理还可以交互两个接口之间的链路状态信息，传递RPF(Remote PHY Fault)等OAM信息。





FlexE通过为每一个Client提供Slot/Calendar配置可更改机制，实现所需带宽的动态调整。FlexE中，对接的两个接口之间通过开销管理通道实时传递体现Client在Group中映射关系的两种不同Calendar配置信息：A和B(分别由“0”或“1” bit表示)。两组Calendar A/B可以动态切

换，从而实现对应Client的带宽可调整。任意一个Client的带宽在两组Calendar A/B之间可能是不同的，通过切换，并进一步结合系统应用控制可以实现无损带宽调整。Calendar A/B的切换通过开销管理通道内嵌的Request/Acknowledge机制实现。

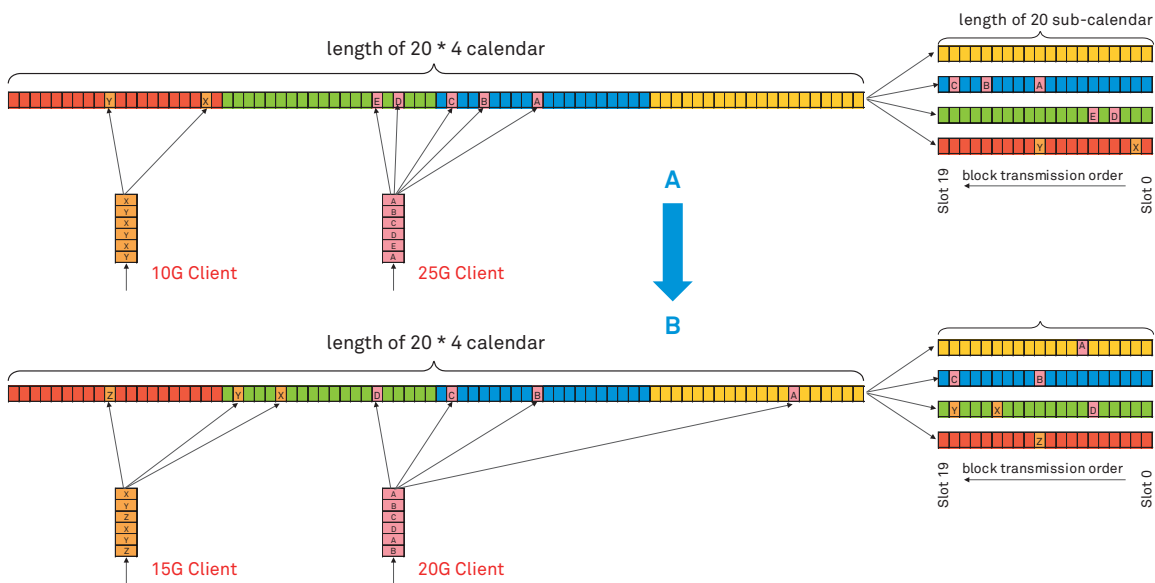


图5 FlexE Calendar切换实现带宽调整示意图

4 灵活以太网(FlexE)的应用

根据FlexE的技术特点，Client可向上层应用提供各种灵活的带宽而不拘泥于物理PHY带宽。根据Client与Group的映射关系，FlexE可提供三种主要功能(如图6所示)：

- » **捆绑(Bonding)**：多路PHY一起工作，支持更高速率。如8路100GE PHY实现800G MAC速率。
- » **通道化(Channelization)**：多路低速率MAC数据流共享一路或者多路PHY。如在100G PHY上承载25G、35G、20G与20G的四路MAC数据流，或者在三路100G PHY上复用承载125G、150G与25G的MAC数据流。
- » **子速率(Sub-Rate)**：单一低速率MAC数据流共享一路或者多路PHY，并通过特殊定义的Error Control Block实现降速工作。如在100G PHY上仅仅承载50G MAC数据流。

子速率功能从某种意义上讲是通道化功能的一个子集。该功能在FlexE接口通过光传输网络连接时，可实现与DWDM链路速率的一致性匹配，并简化相应的映射处理过程。具体而言，就是当MAC数据流速率低于PHY的速率时，FlexE开销帧将未使用的时隙标记为unavailable slots，并在calendar中相应的时隙填充Error Control Block。在FlexE Aware模式下，这些被标记为unavailable slots的时隙将被丢弃。

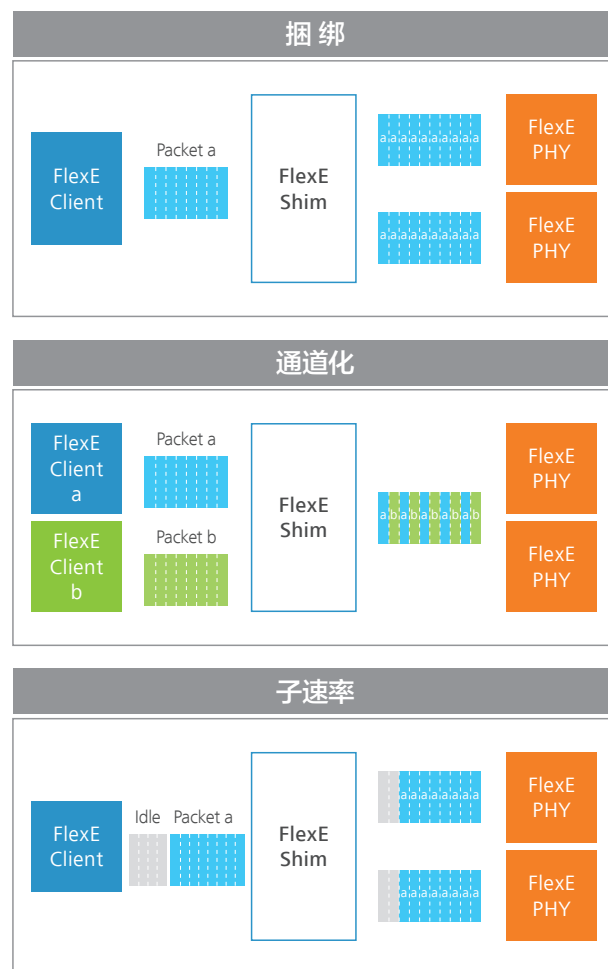


图6 FlexE功能示意图

4.1 通过FlexE捆绑实现超大带宽接口

IP网络的融合承载已成为大势所趋。随着新应用、新业务的不断涌现，以统计复用为特征的IP网络仍遵循着通信行业标准化的发展节奏，在按需快速组网、资源灵活配置以及指定资源保障等方面，面临新的挑战。

以FlexE三大功能为基础，该技术可在IP网络中通过大带宽接口、网络分片、通道化子接口物理隔离等特性，可以实现带宽按需分配、硬管道隔离以及低时延保障等方案，同时结合SDN技术，支持基于业务体验的未来网络架构，以支撑未来的高带宽视频、VR/AR、5G等业务发展。

近期，FlexE可通过超大带宽接口、IP+Optical融合、大客户专线以及网络分片等场景作为切入点逐步推广部署。

IEEE 802.3的以太网标准工作，基于业务需求与技术发展等因素，具有一定的周期性。另一方面，IEEE 802.3所制定的以太网标准为固定速率(如10GE、100GE等)，无法满足基于灵活带宽组网的需求。可以基于FlexE捆绑技术，通过接口速率组合，构造更大带宽的链路(如5*100GE，10*100GE等)。

FlexE Bonding本质上是一种“L1 LAG(Link Aggregation)”技术。由于其基于精细的64/66B Block进行捆绑工作，不存在传统LAG¹ (Link Aggregation Group)中以逐流、或者逐包方式在多条物理链路上分发导致的流量不均衡问题，可以达到100%的带宽分配均衡，并且不存在传统LAG的带宽浪费(一般业界认为LAG会浪费10%-30%带宽)，因而相对LAG技术更具优势。

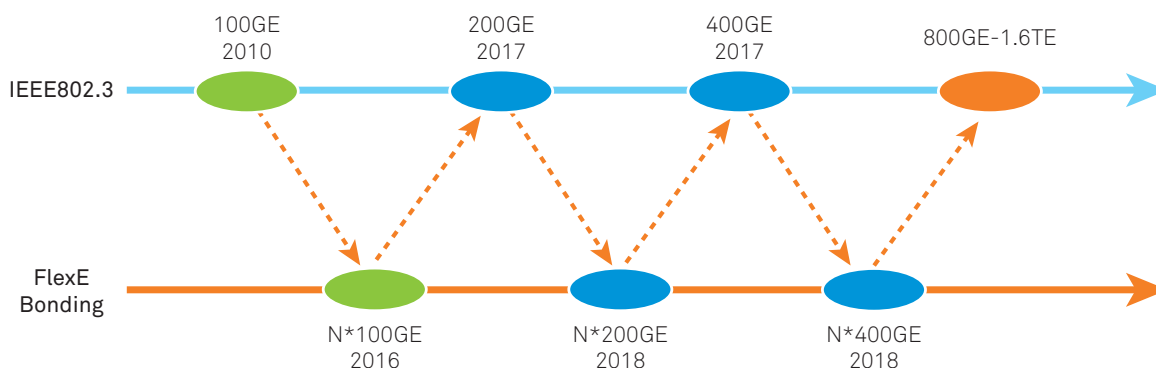


图7 IEEE802.3标准周期性与FlexE Bonding解决方案

¹ LAG技术是指将一组物理以太网接口捆绑在一起作为一个逻辑接口(链路聚合组)来增加带宽并提供链路保护的一种方法。目前广泛使用的链路聚合标准是IEEE 802.1AX协议。针对LAG的逻辑接口，承载的数据流在组内按照负载分担算法(一般为Hash算法)被分配给不同的链路成员(物理以太网接口)。由于数据流的带宽及行为变化性和不确定性，链路聚合的不同物理以太网接口之间很多情况下存在负载均衡问题，并不能获得100%带宽利用率。



4.2 通过FlexE实现IP+Optical灵活组网

FlexE作为路由器与光传输网络设备之间的UNI接口，可以通过速率匹配实现UNI接口实际承载的数据流带宽与光传输网络NNI接口WDM链路承载带宽的一一对应，从而极大简化路由器的FlexE接口在光传输网络传输设备的映射，降低设备复杂度以及投资成本(CAPEX)和维护成本(OPEX)。

OIF Flex Ethernet标准对于灵活以太网在光传输网络中的映射定义了三种模式(参考文

献[3]): Unaware、Termination和Aware。其中Unaware模式与传统以太网接口在光传输网络中通过PCS Codeword Transparent Mapping一致。这种情况类似于光传输网络透明承载灵活以太网接口。这种模式可以充分利用旧现有光传输网络设备，在无需硬件升级的情况下实现对FlexE的承载，并可基于FlexE Bonding功能实现跨光传输网络的端到端超大带宽通道。

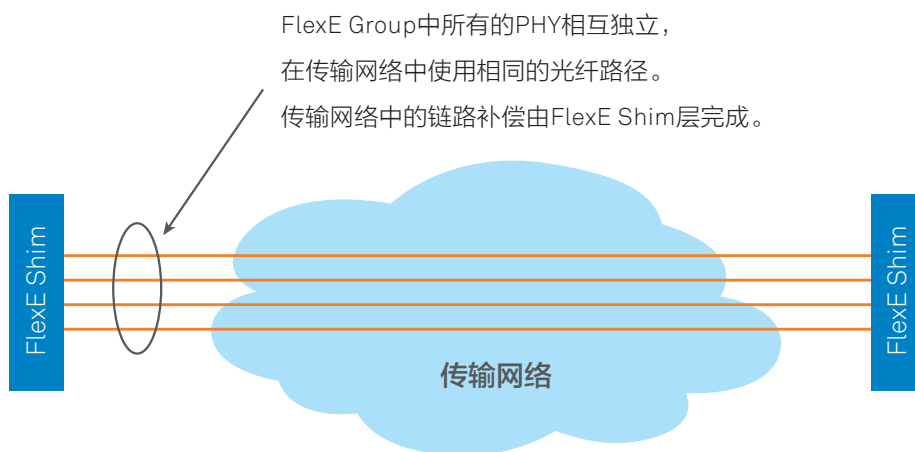


图8 FlexE Unaware模式在传输网络映射示意图

FlexE Termination模式下，光传输网络感知FlexE UNI接口并恢复出FlexE Client数据流，再进一步映射到光传输网络中进行传输承载。这种模式与传统以太网接口在光传输网络上的承载一致，可以在光传输网络中实现对不同FlexE Client流量的疏导等功能。

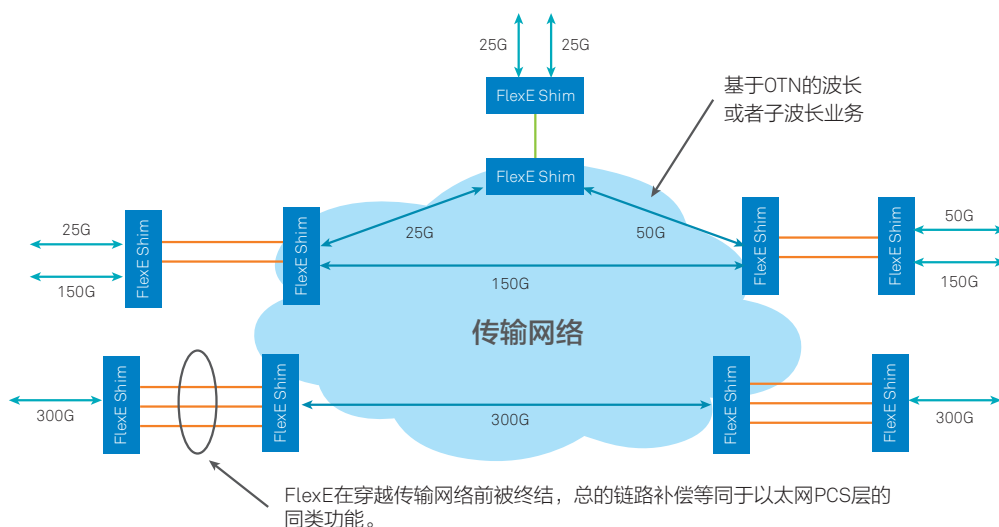


图9 FlexE Terminate模式在传输网络映射示意图

FlexE Aware Transport模式主要利用了FlexE的子速率特性。这种模式下FlexE将unavailable slots通过填充特殊的Error Control Block数据块标识。当作为UNI侧的灵活以太网接口通过Aware模式在光传输网络中映射时，光传输网络直接丢弃unavailable slots，按照原始数据流带宽提取需要承载的数据，进而映射到速率匹配的光传输网络DWDM传输管道。光传输网络设备需要与作为UNI侧的FlexE接口配置保持一致，从而感知FlexE UNI接口并进行承载传输。图10为这种模式的应用原理示意：

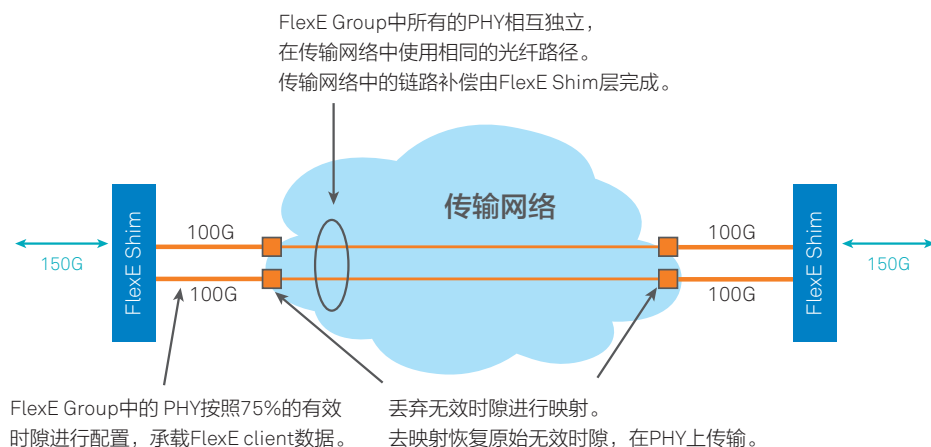


图10 FlexE Aware模式在传输网络映射示意图

4.3 基于FlexE实现硬管道大客户专线

在IP网络中通过硬管道技术，对于重要专线、低时延敏感业务等的承载可以基于FlexE通道化功能构建端到端刚性管道。在统计复用的IP网络中，这种端到端FlexE硬管道专线可在充分利用现有网络基础设施基础上，提供特定高价值客户业务的服务质量保证。

以太网虚拟专线(EVPL, Ethernet Virtual Private Line) 服务已经在企业网和城域网中得到广泛应用，尤其是用于连接地理位置分散的区域(比如企业总部和不同分支之间)。随着基于网络的业务种类的不断增加，对专线服务的质量要求也在不断提高。例如，某些服务要求确保独享带宽和极低延迟，而一些服务却重视隐私保护和高安全性。基于FlexE的专线服务，可以很好地满足这些新的需求。图11展示了一

种针对地理位置相对分散的企业网中FlexE的应用方式，各地区办公室之间是通过FlexE建立连接，而且每条连接都可以根据数据流量来保证所需带宽得到满足。

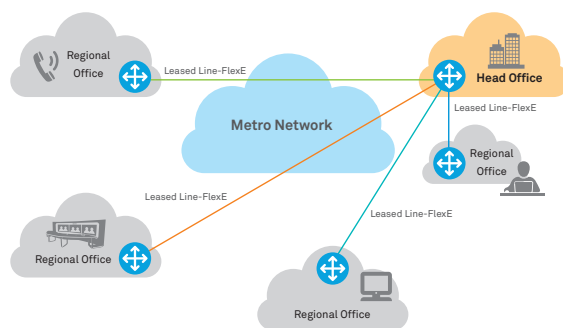


图11 基于FlexE的大客户专线示意图

4.4 基于FlexE实现5G网络分片

网络分片(Network Slicing)通过网络资源的分割来满足不同业务的承载需求，并保证服务的SLA(如带宽、时延等)。按照NGMN发布的5G白皮书(参考文献[4])，分片可以实现不同业务(如eMBB增强宽带、自动驾驶、uRLLC及海量

IoT互连等)在同一个IP网络中承载。FlexE的通道化技术提供了接口级不同FlexE Client之间的物理切分及相互隔离，进一步与路由器架构结合，构建端到端网络分片。

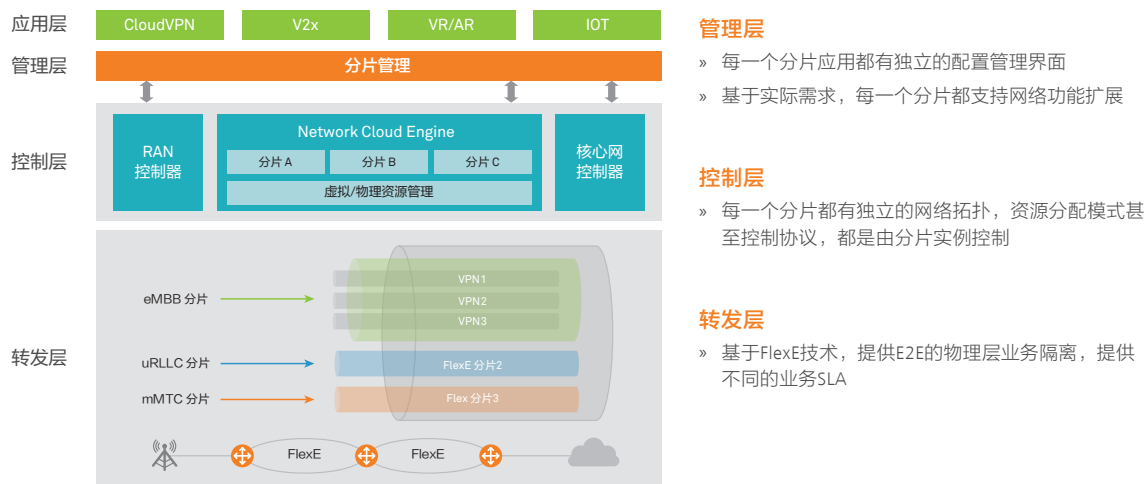


图12 基于FlexE的5G网络分片

5 FlexE标准化及技术发展

FlexE标准于2015年初在Optical Interworking Forum(OIF)启动,并于2016年发布了FlexE IA1.0标准。该标准定义了对100GE PHY的支持,是业界关于FlexE的首个标准,一经发布即引起各方的广泛关注。目前,OIF关于FlexE IA2.0的标准已经发布(参考文献[5]),该标准相对IA1.0版本,增加了200GE/400GE PHY的支持,保持了与FlexE IA1.0相兼容的复用帧格式以及100/200/400GE PHY速率适配的Pad机制,并且基于移动回传应用场景增加了对IEEE 1588V2时间同步的支持。

在OIF标准工作之外,包括ITU-T、IETF、BBF等标准组织也开始启动了FlexE相关标准化工作:

- » ITU-T Q11/15和Q13/15工作组正在定义FlexE的Unaware、Terminate和Aware模式在OTN上的映射,预计会通过G.709标准的补充版本发布。其中,基于FlexE Unaware模式下的映射承载参照了100GBASE-R在OTN中的PCS Codeword透明传输模式;Terminate模式下的传输本质上就是现有传输设备承载以太网数据的方案,可以支持通过Idle/Padding机制进行速率调整;Aware模式在OTN上的映射通过最新定义的Idle Mapping机制实现,支持UNI侧 Client数据流与DWDM链路速率的调整与适配。FlexE的时间和频率同步功能在OTN映射中的机制也在讨论中(参考文献[6])。
- » BBF(BroadBand Forum)于2017年5月启动了“Network Services in IP/MPLS Network using Flex Ethernet”的标准项目(参考文献[7]),定义如何在IP/MPLS网络中通过FlexE接口实现增强型QoS功能架构,以及如何基于现有网络兼容支持FlexE接口的隧道技术以更好地承载带宽和延时保证型业务。在BBF2017年Q3会议期间,多个基于FlexE的技术方案被接纳。这些提案包括FlexE在

IP/MPLS网络中部署的技术方案和架构、基于FlexE的网络分片以及结合Segment Routing实现更灵活路径提供/管理等。

- » IETF(Internet Engineering Task Force)启动了FlexE控制平面标准的制定工作,目标是结合IETF的IP/MPLS技术将FlexE从接口技术扩展成为一种可提供端口级硬隔离效果的端到端技术,以实现网络分片、大客户专线等技术方案。目前IETF的工作聚焦在FlexE框架上,主要讨论端到端FlexE技术的架构、场景以及为实现端到端FlexE路径需要完善/扩展的信令及路由协议。信令扩展聚焦在RSVP-TE与Segment Routing领域;路由协议扩展主要包括ISIS、OSPF与BGP-LS的协议的扩展。(参考文献[8])
- » 随着5G uRLLC承载、时间敏感应用等新兴业务的出现,在IP/Ethernet中保障最恶劣情况下时延目标(guaranteed latency, worst case latency)的确定性网络技术开始出现。其中,二层技术IEEE 802.1TSN和三层技术IETF DetNet定义了IP/Ethernet网络中的拥塞管理机制、基于时延信息的调度算法、显式路径建立以及提供高可靠性的冗余链路技术等,进一步结合FlexE技术,以提供有保障的时延(Bounded Latency)、零丢包的确定性业务承载也成为了研究热点(参考文献[9]、[10])。

随着OIF FlexE 2.0标准的正式发布,以及FlexE技术在数据通讯领域相关标准组织的系统应用、架构级扩展, FlexE技术引起了业界广泛关注,芯片/设备制造商积极投入芯片/设备研发、产品测试及演示等环节,网络运营商/大型OTT业务提供商也在积极通过标准推动、技术合作、方案验证等方式参与其中,相关产业链正在形成(参考文献[11])。

6 总结

作为基于以太网和产业链扩展的技术架构，FlexE技术完全重用了现有IEEE 802.3以太网物理层标准，在MAC/PCS逻辑层通过轻量级增强，实现灵活的多速率接口，并与IP技术实现无缝对接，在IP/Ethernet技术体系下较好地满足了大带宽、灵活速率以及通道隔离等需求，符合技术与产业发展趋势。视频、5G等业务的兴起，以及FlexE技术的完善与功能增强，正在加速FlexE产业链的形成。我们有足够的理由相信，FlexE作为未来IP/Ethernet体系的基础性技术，必将会得到长足发展与广泛应用。

参考文献：

- [1] <http://www.ethernetalliance.org/roadmap/>
- [2] “How can Flexibility on the Line Side Best be Exploited on the Client Side?”, OFC 2016 © OSA 2016, Tad Hofmeister, Vijay Vusirikala, Bikash Koley Google, Inc.,
- [3] OIF Flex Ethernet Implementation Agreement: IA OIF-FLEXE-0.10
- [4] NGMN 5G White Paper, 2015 https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf
- [5] OIF Flex Ethernet Implementation Agreement: IA OIF-FLEXE-0.20
- [6] ITU-T G.709/Y.1331 Amendment 1 (11/2016): Interfaces for the optical transport network Amendment 1
- [7] <https://wiki.broadband-forum.org/display/BBF/BBF+Quarterly+Newsletter>
- [8] <https://tools.ietf.org/html/draft-izh-ccamp-flexe-fwk-03>
- [9] <http://www.ieee802.org/1/pages/tsn.html>
- [10] <https://datatracker.ietf.org/wg/detnet/about/>
- [11] EE Times: Ethernet Flexes Network Muscles:
http://www.eetimes.com/document.asp?doc_id=1330553&page_number=2



缩略语

缩略语	全称	说明
BBF	Broadband Forum	宽带论坛
BGP-LS	Border Gateway Protocol - Link State	BGP链路状态协议
DWDM	Dense Wave-length Division Multiplexing	密集波分复用
eMBB	Enhanced Mobile Broadband	增强移动宽带
FlexE	Flexible Ethernet	灵活以太网
IEEE	Institute of Electrical and Electronics Engineers	电气及电子工程师学会
IETF	Internet Engineering Task Force	Internet工程任务组
IoT	Internet of Things	物联网
ITU	International Telecommunication Union	国际电信联盟
ITU-T	ITU-Telecommunication standardization sector	国际电信联盟电信标准化组
MAC	Medium Access Control	介质访问控制
MEF	Metropolitan Ethernet Forum	城域以太网论坛
MPLS	Multiprotocol Label Switching	多协议标记交换
NGMN	Next Generation Mobile Network	下一代移动网络
OAM	Operation, Administration and Maintenance	运行管理和维护
OIF	Optical Internetworking Forum	光互联网论坛
OTN	Optical Transport Network	光传送网络
PCS	Physical Coding Sublayer	物理编码子层
PHY	Physical Layer	物理层
PMA	Physical Medium Attachment	物理介质接入子层
PMD	Physical Media Dependent	物理介质关联层
QoS	Quality of Service	服务质量
QoE	Quality of Experience	感知体验质量
RSVP-TE	Resource Reservation Protocol-Traffic Engineering	针对流量工程扩展的资源预留协议
SerDes	Serializer/Deserializer	串行/解串器
UNI	user-to-network interface	用户侧接口
NNI	Network to Network Interface	网络侧接口
uRLLC	Ultra-Reliable and Low-Latency Communication	超高可靠极低时延通讯