

新型公共网络中的流量工程

| | |
|---|----|
| 介绍..... | 2 |
| 流量工程..... | 2 |
| 流量工程的应用..... | 3 |
| 展望..... | 3 |
| 过去：传统的路由器核心网络..... | 3 |
| 流量工程的性能..... | 4 |
| 传统路由器核心网络流量工程的局限..... | 4 |
| 现在：IP 覆盖型网络..... | 5 |
| IP 覆盖型网络的运行..... | 5 |
| IP - over - ATM 模型在服务提供商网络中的优势..... | 7 |
| IP - over - ATM 在 OC - 48 光互联网中的局限..... | 7 |
| 未来：Juniper 网络公司的流量工程体系..... | 9 |
| 流量工程解决方案中的组成部分..... | 9 |
| 包转发部分..... | 10 |
| 基于标记交换的 LSR 包转发..... | 10 |
| 数据包在 MPLS 骨干网中传输的例子..... | 11 |
| MPLS 的优势..... | 12 |
| 信息发布部分..... | 12 |
| 路径选择部分..... | 13 |
| 信令部分..... | 15 |

| | |
|-----------------------------|----|
| 灵活的 LSP 计算及配置..... | 16 |
| 成功流量工程解决方案在操作上的要求..... | 17 |
| Juniper 网络公司的流量工程体系的优势..... | 18 |
| 结论..... | 19 |
| 参考..... | 20 |

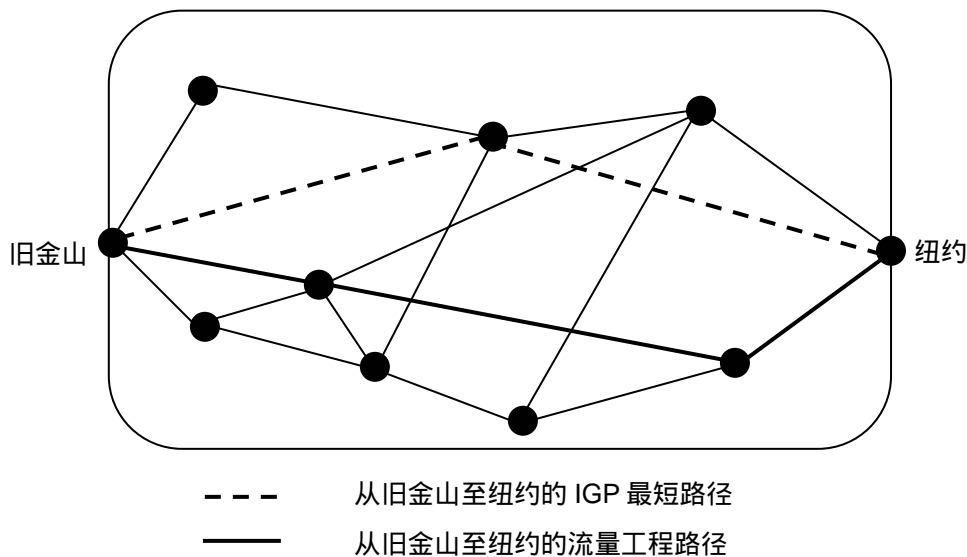
介绍

Internet 服务提供商面临的挑战主要来自于如何使他们的客户满意并保持高速增长。首先，也是最基本的要求，ISP 需要在一定地域内提供许多具有不同带宽的线路。换句话说，ISP 必须配置一个能够使他们的客户连接到他们网络上的物理拓扑结构。

在网络部署完毕后，ISP 必须将客户的业务流映射到网络的物理拓扑上。90 年代初，将业务流映射到网络物理拓扑上并不是以一种科学的方法来实现。这种映射的实现只是基于产品的路由配置 - - 业务流只是简单地被分配到由 ISP 所使用的内部网关协议 (IGP) 计算出的最短路径上去。这种不规则映射的局限是通过当某条链路发生阻塞时，提供过量带宽来解决的。现在，ISP 网络越来越大，线路上支持的 IP 越来越快，同时，客户的需求也变得越来越高。因此，将业务流映射到物理拓扑上的任务需要以一种完全不同的方式来实现，只有这样，网络上传输的负载才能通过一种受控和有效的方式得到支持。

流量工程

将业务流映射到现有物理拓扑上的任务被称作流量工程。目前，流量工程作为一个课题在 Internet 工作组和一些大型 ISP 内部被热烈地讨论。如果一个流量工程的“应用”能够实现一组正确的功能，它将使 ISP 在其路由域内对业务流的分布实现精确的控制。特别地，流量工程还可以在 ISP 网络内实现将业务流从通过 IGP 选择的最短路径，转移至另一条潜在的、具



有更少阻塞的物理路径上去 (图 1)。

图 1：服务提供商网络中的流量工程路径与 IGP 最短路径比较

流量工程是 ISP 的一个强有力的工具，ISP 通过它可以在网络中不同的链路、路由器和交换机之间平衡业务负荷，使所有这些成分即不会过度使用，也不会未充分使用。这样，ISP 可以有效利用整个网络所提供的带宽资源。流量工程应当被看成是路由结构中的一个辅助，它能够在沿网络中备选路径转发业务时提供辅助信息。

流量工程的应用

由于客户对网络资源需求的空前增长、IP 应用中的重要任务的性质，以及 Internet 市场中不断增加的竞争性，使流量工程在 ISP 内成为一个重要的问题。现有的 IGP 在建立转发表时，并未将带宽的可用性和业务特点考虑进去，因此会使网络出现阻塞。ISP 清楚流量工程可有效地增强网络的运行和性能。他们希望流量工程具有以下功能：

- 对主路径进行路由时，绕过网络中已知的瓶颈和阻塞点。

- 当主路径发生一个或多个故障时，为业务如何进行重新路由提供明确控制。
- 通过确保网络的附属设备不会被过度使用，同时，潜在的备选路径上的网络附属设备不会未被充分使用，从而对可用的集成带宽和长距离光纤进行有效利用。
- 通过使运行有效性最大化而另运行费用降至最低，使 ISP 在市场中更具竞争实力。
- 通过使包丢失最小化，将阻塞的保持时间最小化和使吞吐量最大化的方法增强网络中以业务为导向的性能特性。
- 增强网络中将来用于支持多业务 Internet 的统计约束性能特性（如，丢失率，迟延变化，传输时延等）。
- 为客户提供更多的选择、更低的费用和更好的服务

展望

就在 ISP 们努力使自己跟上不断增长的 Internet 业务量的同时，流量控制已成为 ISP 们一个非常重要的工具。为增强读者对流量工程的理解及其在支持将来的 Internet 中的重要角色，本白皮书将从如何在传统的基于路由器的骨干网中实现流量工程开始。然后讨论在今天的 ATM 和帧中继“覆盖”型网络中如何实现流量工程及相关的技术、优势和局限性。

在讨论完现在普遍使用的配置解决方案后，白皮书将介绍一种专门为运行在光 Internet - 一个由密波分复用 (DWDM)、OC - 48 和 OC - 192 速率的接口、IP - over - SONET、IP - over - glass 和 Internet 骨干网路由器组成基础结构的骨干网环境 - 而设计的流量工程实现方法。最后一部分将介绍 Juniper 网络公司和 IETF 的基于多协议标记交换协议 (MPLS) 和资源预定协议 (RSVP) 技术的流量工程解决方案。

过去：传统的路由器核心网

90 年代初期，ISP 的网络通过使用租用线 - T1 (1.5Mbps) 和 T3 (45Mbps) 链接 - 将路由器互连而组成网络。当 Internet 开始它的爆发性增长时，对带宽需求的增长要比单条网络链接速率快得多。ISP 们对这个挑战的反应是提供更多的链接以提供额外带宽。从这一点上

看，流量工程对 ISP 变得越来越重要了，因此，当存在多条并行或备选路径时，ISP 们可有效地使用集成网络带宽。

流量工程的功能

在早期基于路由器的核心网中，流量工程是通过简单地使用路由量度值来实现的。因为那时无论从路由器数量、链接数及业务流量来讲，Internet 骨干网都是非常小的，所以，基于度量的控制在当时是足以胜任的。同时，在万维网普遍流行之前，Internet 拓扑层次也强制业务通过网络中较为确定的路径及事件，不会产生临时的热点。

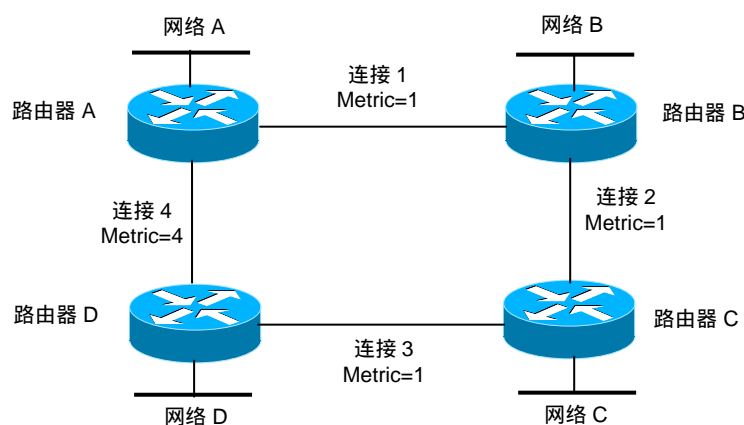


图 2：基于量度的流量控制

图 2 描述了基于量度的流量工程是如何运行的。假设网络 A 发送了大量业务给网络 C 和网络 D。参考图 2 所示的量度值，链接 1 和链接 2 可能发生阻塞，因为网络 A - 网络 C 和网络 A - 网络 D 的业务都将流过这些链路。如果链接 4 的量度值变为 2，网络 A - 网络 D 的流量将转移至链接 4，但网络 A - 网络 C 的业务将继续留在链接 1 和 2。结果，在不中断网络中任何处理的情况下修正了“热点”。

一直到 1994 或 1995 年，基于量度的流量控制都提供了充足的流量工程解决方案。但从那以后，ISP 的网络规模越来越大，他们无法再通过使用基于量度的流量控制和基于传统路由器的核心网来继续发展自己的网络。在一个巨型网络中，对网络某个局部的量度值进行调整是否会引起网络其它部分的新问题将变得非常难以判断。同时，对于基于传统路由器的核心网，当 ISP 计划增大他们的核心网时，路由器无法提供 ISP 们所需的高速接口和确定的性能。

传统路由核心网流量工程的局限性

传统的路由核心网在为流量工程提供可扩展性的支持上存在着许多局限：

- 由于传统路由器的汇集带宽和包处理能力有一定的局限性，因此，传统的、基于软件的路由器在高负荷的情况下可能成为潜在的瓶颈。
- 基于量度处理的流量工程不具有可扩展性。当 ISP 网络变得具有更多的链接时（即，更大、更密集的结网和更多的冗余），这种情况下很难保证对网络某个部分量度的调整而不致在网络的其它部分引起问题。基于量度处理的流量工程对于增加的复杂问题提供的是一个跟踪 - 纠错的解决方式，而不是一个科学的解决方案。
- IGP 计算是通过拓扑驱动的，它只基于一个简单附加量度，如跳数或某个管理值。IGP 并不发布类似于带宽可用性和业务特征等信息。这就意味着，当 IGP 计算其转发表时并不考虑网络上的业务负载。结果，业务不能在网络连接中平均分配，导致昂贵的资源未能被有效使用。一些链路可能发生阻塞的同时，另一些链路未被充分利用。这种情况在稀疏连接的的网络中也许能满足客户的需求，但对于一些复杂连接网络，对业务所使用的链路进行控制以确保链路的负荷均衡变得非常必要。

现在：IP 覆盖型网络

在 1994 或 1995 年左右，Internet 业务量的不断增长使 ISP 需要令他们的网络主干能够支持高于 T3 (45Mbps) 的速率。幸运的是，这时在交换机和路由器上 OC - 3 (155Mbps) 速率的 ATM 接口出现了。为了获得所需的速率，ISP 被迫重新设计他们的网络，使他们能够使用由交换 (ATM 和帧中继) 核心网提供的更高速率。一些 ISP 将他们的网络从 DS - 3 点到点连接转移至在网络边缘使用带有 OC - 3 速率 ATM 接口的路由器，而在网络核心部分使用 OC - 3 速率的 ATM 交换机的网络结构。大约在 9 个月之后，ATM 交换机之间的连接速率升级到 OC - 12 (622Mbps)。另一些 ISP 开始在他们的 DS - 3 帧中继网络中增加节点。当他们开始

从帧中继转移至 ATM 时，他们在网络边缘使用 OC - 3 链接，但不久便在核心部分配置了 OC - 12 速率的交换机间的连接。（见图 3）

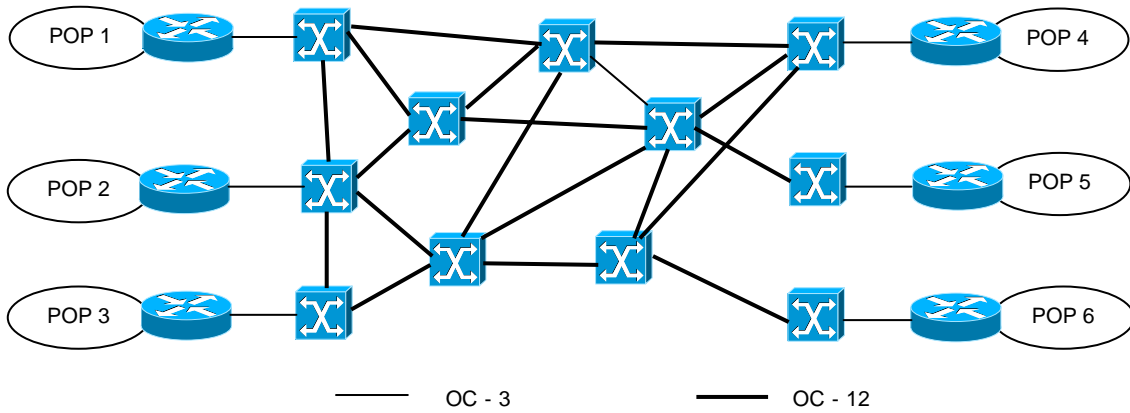


图 3：1997 年和 1998 年典型的大型 ISP 核心网物理拓扑结构

IP 覆盖型网络的运行

当 IP 运行在 ATM 网络上时，路由器在 ATM 网络的边缘环绕。每个路由器通过一系列经由 ATM 物理拓扑配置的永久虚电路 (PVC) 与其它路由器通信。PVC 就象逻辑电路一样工作，为边缘路由器提供连接。路由器并不能直接访问 ATM 结构中 PVC 的具体物理拓扑信息。路由器仅了解特定的 PVC 就象出现在两个路由器之间的简单的点到点电路。图 4 说明了 ATM 核心网物理拓扑与逻辑 IP 覆盖拓扑的区别。

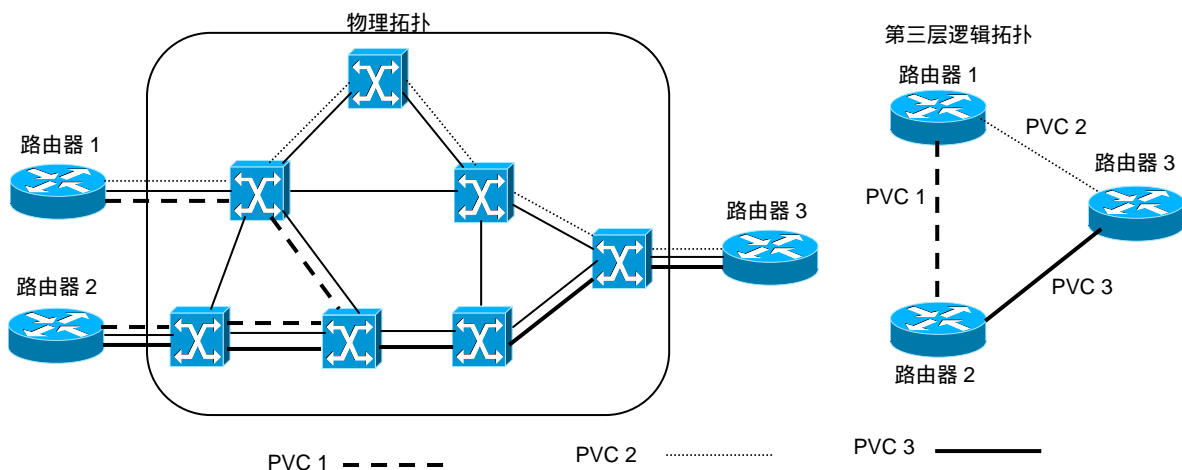
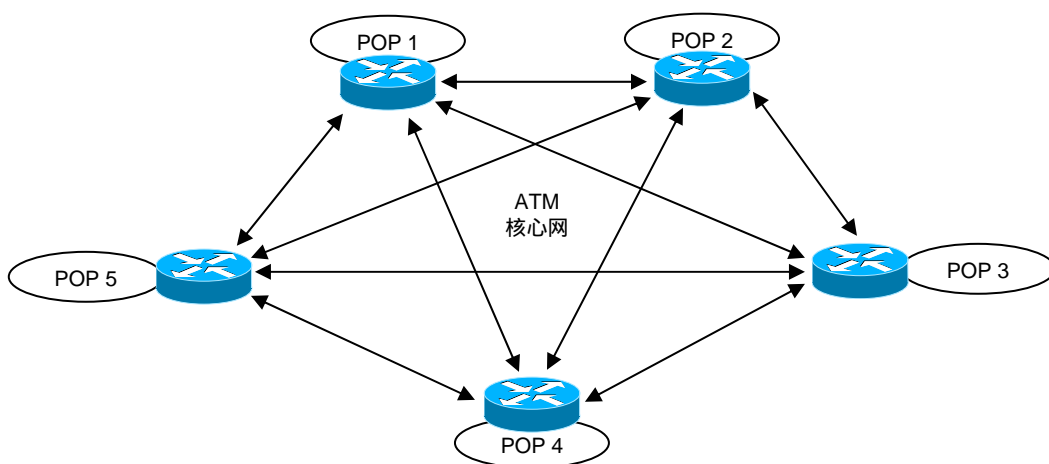


图 4：ATM 物理拓扑与第三层覆盖拓扑比较

对于大型 ISP，ATM 核心网由 ISP 完全拥有，并专用于支持 Internet 骨干网业务。这种核心网的基础结构与运营商的其它专用数据业务完全分离。因为网络由 ISP 完全拥有并专门

用于 IP 业务，所有的业务通过 ATM 核心时使用不确定比特率 (UBR) ATM 服务等级 - 没有策略，没有业务整形，没有峰值信元率，没有维持信元率。ISP 只是简单地将 ATM 交换结构作为一个高速传输系统，而并不依赖 ATM 的业务和阻塞控制机制。对于 ISP 们，没有什么理由需要他们使用这些先进的“特性”，因为每个 ISP 都拥有自己的骨干网，他们不需要对自己加以限制。

PVC 覆盖的物理路径通常通过离线配置计算获得的，它使用基于需求的方式 - - 当阻



塞发生，新增一条干线或配置一个新的 POP。一些 ATM 交换机厂商在考虑流量工程时，会使用一些专有的技术对 PVC 进行路由。但是，这些方案并不成熟，ISP 经常需要进行完整的离线状态路径计算以达到预期的结果。PVC 路径和特性通过使用基于链接容量和历史业务参数配置而对其进行整体优化。离线状态配置的使用也可以帮助设定一组备用 PVC，以在故障条件下准备作出反应。最后，在完成 PVC 结网的整体优化计算后，配置将被下载到路由器和 ATM 交换机以提供单个或两个全闭合结网的逻辑拓扑。

图 5：ATM 核心网上的逻辑 IP 拓扑

当 ATM 的 PVC 被映射到路由器的子端口时，分离的 ATM 网络和 IP 网络相遇。路由器子端口与 ATM 网的 PVC 互相协调，然后路由协议在子端口上与 IP 报头 (路由) 协调工作。实际上，离线配置的使用生成了路由器和交换机的配置，确定 PVC 号码的一致性，并且进行了适当的映射。

最后，通过在 PVC 中运行 IGP 以建立对等关系并交换路由信息，使 ATM 的 PVC 集成到 IP 网中去。在任何一对路由器间，主 PVC 的 IGP 量度配置要比备用 PVC 配置更为首选。这便保证了只有在主 PVC 无效时才会使用备用 PVC。同时，如果主 PVC 在故障后重新恢复，业务将从备用 PVC 回到主 PVC 上。

IP - over - ATM 模型在服务提供商网络中的优势

90 年代中期，ATM 交换机可提供解决 ISP 需要更多带宽以应付不断增加的业务负载需求的方案。那些决定转移到基于 ATM 核心网以继续经历成长的 ISP 们发现，ATM 的 PVC 提供了当业务通过网络时对其进行明确控制的工具。ISP 开始依赖于 ATM 交换机所提供的高速接口、确定的性能和 PVC 功能性，以成功地对其网络的运行进行管理。

当与传统的基于软件的路由器进行比较时，ATM 交换机提供了更高速率的接口和明显的更高的汇集带宽，因此避免了在网络核心部分因路由器引起瓶颈的潜在可能性。同时，速率和带宽为 ISP 提供了确定的性能。在那时，路由器是不可能提供类似功能的。

一个基于 ATM 的核心网完全支持流量工程，因为它可以对 PVC 进行明确的路由。PVC 的路由是通过在网络底层的物理拓扑上提供随机的虚拟拓扑实现的，而在网络底层的物理拓扑上，通过对 PVC 进行路由以使业务分配到所有链路上去，以致链路被平均使用。这种实现避免了业务全部汇集到低花费路由上去，从而避免了链路的过分使用或未充分使用。由 ATM PVC 提供的流量工程性能使 ISP 在他们的市场范围内更具竞争性，允许他们为其客户提供低费用和更好的服务。

由 ATM 交换机提供的每条 PVC 的统计信息，简化了监测用于优化 PVC 布局及管理的业务参数的过程。网络设计者最初为支持特定的流量工程目的而提供每条 PVC，然后，他们将连续监测每条 PVC 上的业务负载。当一条特定的 PVC 发生阻塞时，ISP 具有所需要的信息，使其能够通过修改虚拟或物理拓扑结构以适应偏移的业务负荷，对发生的事件进行补救。

在 OC - 48 光互联网中，IP - over - ATM 模型的局限性

在过去几年中，ATM 交换机增强了 ISP，允许他们扩大市场份额并增加利润。90 年代中期，ATM 交换机因其独一无二的高速接口、确定的性能，和通过对 PVC 进行明确的路由以

实现流量工程的性能而被选择。但是，今天，曾一度只为 ATM 交换机所拥有的一些特性，同样也能被 Internet 骨干网路由器所支持。路由技术的最新发展，使 ISP 重新评价其是否要继续容忍覆盖模型的局限性：管理费用，设备费用，运行可靠性和扩展性。

基于 ATM 的核心网的一个最根本的局限性是它需要对两个不同的网络进行管理：ATM 基础结构和逻辑的 IP 覆盖。通过在 ATM 网上运行 IP 网络，ISP 不仅增加了网络的复杂性，而且加倍了开销，这时因为 ISP 必须管理和协调两个分离网络的运行。同时，路由和流量工程分别在不同的系统上来完成 - - 路由在路由器上执行，流量工程则在 ATM 交换机上完成 - - 因此，将流量工程完全与路由集成在一起将是非常困难的。最近的新技术发展使 Internet 骨干网路由器能够提供以往只能在 ATM 交换机上才能发现的高速链接和确定的性能。当考虑将来需要升级至 OC - 48 速率时，ISP 们必须要决定是继续使用昂贵和复杂的设计，还是用一套设备即可完成相同功能的基于路由器的完全集成的核心网。

ATM 路由器接口未能跟上光学带宽的最新发展。已商品化的最快的 ATMSAR 路由器接口是 OC - 12。今天，OC - 48 的 POS 路由器接口已经实现，但是，OC - 48 速率的 ATM 路由器接口在短期内并不会被实现。很快，OC - 192 (~10Gbps) 的 POS 路由器接口会被推出，但是，OC - 192 的 ATM 路由器接口可能永远不会商品化，因为在如此高的速率上实现 SAR 功能是非常昂贵和复杂的。SAR 在扩展性上的这些局限，意味着当 ISP 们试图使用 IP - over - ATM 模型提高网络的速率时，将必须订购大型 ATM 交换机和带有大量较低速率的 ATM 接口的路由器。当 ISP 们今后考虑向 OC - 192 转移时，他们将付出巨额开支扩展其网络，而且会增加网络的复杂性。

以包为导向的协议，如 IP，在 ATM 结构上运行时，将引入信元税的概念。假设有 20 % 的 ATM 消耗用于组帧，分配包尺寸，则对于一个 2.488Gbps，OC - 48 的链路，1.99Gbps 将用于用户数据，而 498Mbps，约一个 OC - 12 将被用于 ATM 开销。当 10Gbps，OC - 192 接口出现时，1.99Gbps - 将近一个完整的 OC - 48 - 的容量将为 ATM 开销占用。当 ISP 业务需要将其网络升级至 OC - 48 和 OC - 192 速率时，他们必须决定是继续支付 ATM 信元税而使他们处于一个很不利的竞争位置，还是采用一个能将浪费的开销用于客户业务的，基于路由器的核心网络。

一个配置了全闭合 ATMPVC 的网络将产生传统的“ N^2 ”问题。对于一个小型或中型的网络，“ N^2 ”问题并不是一个主要的问题。但是，对于一个具有数百个路由器的核心 ISP，这种挑战将变得十分突出。例如，当将一个小型网络从 5 个路由器扩展到 6 个时，ISP 只需将 PVC 的数量从 20 条增加到 30 条。但是，将路由器的数量从 200 个增加到 201 个时，则需要配置 400 条新的 PVC - 从 39800 条增加到 40200 条 PVC。应该强调的是，这个数量并不包括备份 PVC 或网络在运行多种业务时所需的附加的 PVC，因为在提供一些业务时将在两个路由器间需要多于一的 PVC。由于“ N^2 ”问题的存在，将导致一系列操作上的问题：

- 必须要将新的 PVC 映射到物理拓扑上
- 必须要协调新增的 PVC，以使得他们对已有 PVC 产生最小的影响
- 巨大数量的 PVC 可能超出 ATM 交换机的配置及实现能力
- 必须修改核心部分的每一个交换机和路由器的设置

90 年代中期，因为路由器较低的接口速率和确定性能上的缺陷，避免内部路由器跳转的要求需要一个全闭合的 PVC 结网。现在，Internet 骨干网路由器已经克服了这些历史的局限，将“ N^2 ”问题留给了传统的 IP - over - ATM 结构。

配置一个全闭合 PVC 同样也给 IGP 带来压力。这种压力在于需要维护大量的对等关系，故障时需处理“ N^3 ”的链接状态更新，和为一个包含大量逻辑链接的拓扑进行 Dijkstra 计算的复杂性。任何时候将拓扑配置成为一个全闭合的结构，对 IGP 产生的压力都会使得拓扑结构难以维护。同时，当 ATM 核心网扩大时，“ N^2 ”将对 IGP 的复合产生压力。在基于路由器的核心网里避免了“ N^2 ”的压力。同信元税一样，IGP 压力是 IP - over - ATM 模型的产物。

基于覆盖模型的流量工程要求支持交换和 PVC 的第二层技术的存在。在一个混合媒体网络中，流量工程对于特殊的第二层技术（ATM）的依赖，使其难以提供可行的方案。如果一个 ISP 希望在 POS 或光网络中实施流量工程，则第二层传输不能提供流量工程，因为它被跳过了。混合媒体网络的增长和减少 IP 与光纤间的层数的目的，要求流量工程在第 3 层实现，以提供一种集成的途径。随着 ISP 不断地使用光互联网络模型建立他们的网络，IP - over - ATM 结构的局限性变得越来越突出。

总之，最初支持基于 ATM 核心网配置的基本假设将不再有效。当其它一些可选的模型不断出现时，继续坚持 IP - over - ATM 模型的诸多不利随之出现。高速接口，确定的性能，和使用 PVC 的流量工程不再使 ATM 交换机明显区别于 Internet 骨干网路由器。而且，基于路由器的核心网的配置解决了 ATM 模型的一些固有问题 - - 协调两个分立系统的复杂性和较高的费用，ATMSAR 接口的带宽局限性，PVC 的“N²”问题，IGP 压力，不能在混合媒体结构中运行的局限性，和不能实现第 2 层和第 3 层之间的无缝连接的不利因素等。关于高速接口和确定的路由器性能的题目已在 Juniper 网络的白皮书“Internet 骨干网路由器及相关的 Internet 设计”中讨论过。本白皮书余下的部分将着重讨论如何在基于路由器的核心网中最好地实现流量工程。

未来：Juniper 网络的流量工程结构

当 ISP 开始计划转移到更高速的网络时，他们应当仔细检查选择的方案，使他们过去的带宽和流量工程决策不会约束将来网络的增长和运行。对于运行在 OC - 48 速率上的高性能骨干网，问题变化得非常快，以至于保持相同策略只采用辅助的（或主要的）增强方式去修改或调整网络方案已不大可能。最后，所有的技术都将到达他们生命周期的某一点，那时，他们将不再能够扩展，网络设计人员将意识到是该停下来、重新开始、考虑新的解决方案的时候了。

我们可以非常清楚地看到，任何基于路由器的流量工程的实现方案，必须要提供与现存的 IP - over - ATM 等同的功能性。ISP 已经知道如何依赖于 ATM 的高速接口，确定的性能以及 PVC 配置能力的流量工程，如果没有其它过人之处，他们将保持现状不变。基于路由器的实现必须提供一套集成的方案，使之能够避免对两个分离网络进行协调和管理的复杂性和费用。最后，被提议的方案必须能够提供自动流量工程处理选项，使 ISP 能够提供增强的客户服务和可靠性，同时减少运行费用。

为在一定时间内满足 ISP 的需求，任何将来的实现途径都可以通过 IETF 工作组提供的已存在的工作结果来支持。如果方案中的一些组成部分已经存在，则被鼓励采用他们，而不是

回避，努力地“重新启动车轮”。需要进行开发的新技术应相应地简单和明确，使得他们在执行和配置过程中具有最小的风险。最后，新的流量工程方案的开发应该组合相关的简易性和方便的配置技术，为光互联网增长的可扩展性提供稳定的方案。

流量工程方案的组成部分

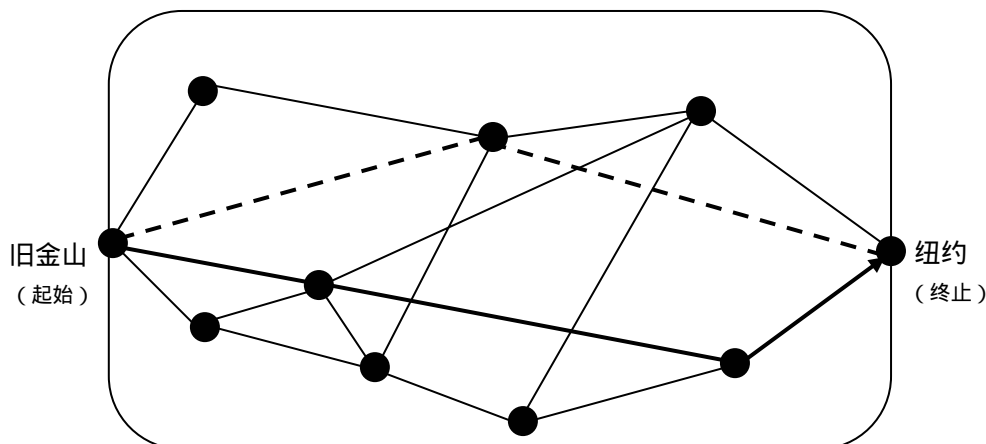
为实现基于路由器的流量工程实施方案，Juniper 网络公司一直积极地参与多协议标记交换 (MPLS) 标准的制定及相关的 IETF 工作组。我们认为使用 MPLS 的流量工程的策略包括 4 个基本组成部分：

- 包转发
- 信息发布
- 路径选择
- 信令

每个功能单元都是一个独立的模块。Juniper 网络公司的流量工程体系在四个功能模块之间提供了开放的接口。这种模块化的组合与开放的接口，为单个模块在有更好的方案出现时，按需要进行更改提供了灵活性。

包转发单元

Juniper 网络流量工程结构中的包转发单元是多协议标记交换 (MPLS)。MPLS 负责引导 IP 包流按一条预先确定的路径通过网络。这条路径被称作标记交换路径 (LSP)。LSP 本质上与 ATM PVC 相似，即业务从起始路由器按一定方向流向终止路由器。双工业务需要两条



LSP，每条 LSP 用于承载一个方向上的业务。LSP 的建立是通过串联一个或多个标记交换跳转点来完成，允许数据包从一个标记交换路由器（LSR）转发到另一个 LSR，从而穿过 MPLS 域（图 6）。LSR 是一个支持基于 MPLS 转发的路由器。

图 6：穿过 MPLS 域的 LSP

当起始 LSR 收到一个 IP 包后，它为此包加上一个 MPLS 报头，然后将其转发到 LSP 上的下一个 LSR。被标记的包被每个 LSR 沿 LSP 转发，直至到达 LSP 的终止处，在那一点上，MPLS 报头被去除，包基于第 3 层的信息进行转发，如基于 IP 目的地址。这个过程的关键之处在于，LSP 的物理路径并不为通过 IGP 选择的到达目的 IP 地址的最短路径所制约。

基于标记交换的 LSR 包转发

每个 LSR 进行的包转发处理都是基于标记交换的概念。这个概念与发生在 ATM 交换机中的 PVC 相似。每个 MPLS 包都带有一个 4 个字节的封装头，其中包含了 20 个字节的固定长度标记区域。当一个包含标记的包到达 LSR 时，LSR 检验标记并且将它在 MPLS 转发表中作为索引。每个转发表中的条目都包含了一个接口 - 内部标记对，其映射了所有在特定接口上到达并具有相同内部标记包的一系列转发信息。

MPLS 转发表

| 输入接口 | 输入标记 | 输出接口 | 输出标记 |
|------|------|------|------|
| → | → | → | → |
| 3 | 21 | 4 | 18 |
| 3 | 56 | 6 | 135 |
| → | → | → | → |

图 7：MPLS 转发表的例子

图 8 描述了 LSR 使用的标记交换算法的运行情况。LSR 在接口 3 上收到一个包，其包



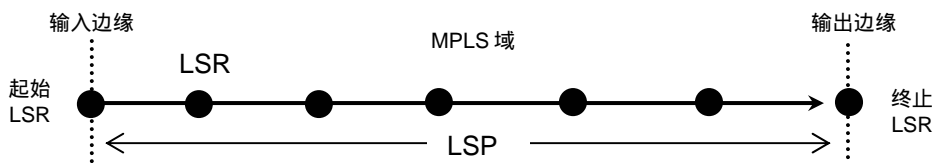
含的标记为 21。LSR 使用图 7 中转发表所包含的信息，将标记值更换为 18 并将包从接口 4 转发到下一跳的 LSR。

图 8 : LSR 转发一个包

包如何在 MPLS 骨干网中传输的例子

这一部分将阐述一个 IP 包在 MPLS 骨干网中传输时是如何被处理的。我们将在网络中的 3 个分立点上检测包运行的情况：1.当包到达 MPLS 骨干网的输入边缘时，2.当包沿着 LSP 被每个 LSR 转发时，3.当包在到达 MPLS 骨干网的输出边缘时。

在 MPLS 骨干网的输入边缘，起始 LSR 检验 IP 报头。基于这个分析，包被分类，并分配给它一个标记，以 MPLS 报头进行封装，然后转发给 LSP 的下一跳。MPLS 在为 IP 包分配 LSP 的方法上提供了极大的灵活性。例如，在 Juniper 网络公司的流量工程实现中，所有到



达起始 LSR 且在同一输出 LSR 离开 MPLS 域的包都会沿同一条 LSP 转发（见图 9）。

图 9 : MPLS 骨干网络

一旦包开始在 LSP 中传输，每个 LSR 使用标记作出转发决定。要记住，MPLS 的转发决定并未考虑最初的 IP 报头。而且，输入的接口和标记在 MPLS 转发表中被用作查找关键字。旧的标记被新标记替换，然后，包沿 LSP 转发至下一跳。LSP 中的每个 LSR 都重复同样的处理，直到包到达终止 LSR。

当包到达输出 LSR 时，标记被去除，包离开 MPLS 域。包接着按原来 IP 报头中包含的目的 IP 地址和由传统的 IP 路由协议计算出的最短路径进行转发。

在这部分里，我们并没有讨论标记是如何在沿 LSP 的 LSR 中进行分配和发布的。我们将在 Juniper 网络公司流量工程结构中的信令部分中讨论这个重要的任务。

MPLS 的优势

人们一般都相信 MPLS 可明显地增强 LSR 的转发性能。更确切地说，精确查找，例如由 MPLS 和 ATM 交换机所提供的，要比由 IP 路由器提供的最长匹配查找快。但是，最近芯片技术的进步使基于 ASIC 的路由查询引擎与 MPLS 或 ATM 的 VPI/VC1 查找引擎运行速度相同。

MPLS 技术的真正优势在于它提供了路由（即，控制）和转发（即，转移数据）间的完全分离。这种分离允许只使用单一的转发算法 - MPLS - 便可对多种服务和业务类型进行配置。将来，当 ISP 们需要开发一种新的增值服务时，MPLS 转发结构可以被保留，新的业务可通过更换包被分配至 LSP 中的方法而简单地建立。例如，当包被分配至一条 LSP 时，可基于目的子网和应用类型的组合，源和目的子网的组合，特殊的 QoS 需求，IP 多点传送组，或虚拟专网（VPN）识别号。基于这种方式，新的业务可以简单地被加入到通常应用的 MPLS 转发结构中。

信息发布部分

因为流量工程需要有关网络拓扑和网络负荷的动态信息的细节，新的流量工程模型的主要需求是用于信息发布的框架。这部分可简单地通过定义相关的 IGP 扩展来实现，这样，链接特性可包含在每个路由器的链接状态广播中。IS - IS 扩展可通过定义新的类型长度值（TLV）来实现，而 OSPF 扩展可通过不透明 LSA 来实现。链接状态 IGP 所使用的标准扩散算法保证链接特性被发布至 ISP 路由域中的所有路由器。

每个 LSR 通过一个特殊的流量工程数据库（TED）对网络链接特性和拓扑信息进行管理。TED 专门用于计算 LSP 通过物理拓扑时的外在路径。一个分离的数据库被维护以使并发的流量工程计算与 IGP 和 IGP 链接状态数据库独立。同时，IGP 继续无改变地运行，通过路由器链接状态数据库所包含的信息进行传统的最短路径计算（见图 10）。

LSR 组成框图

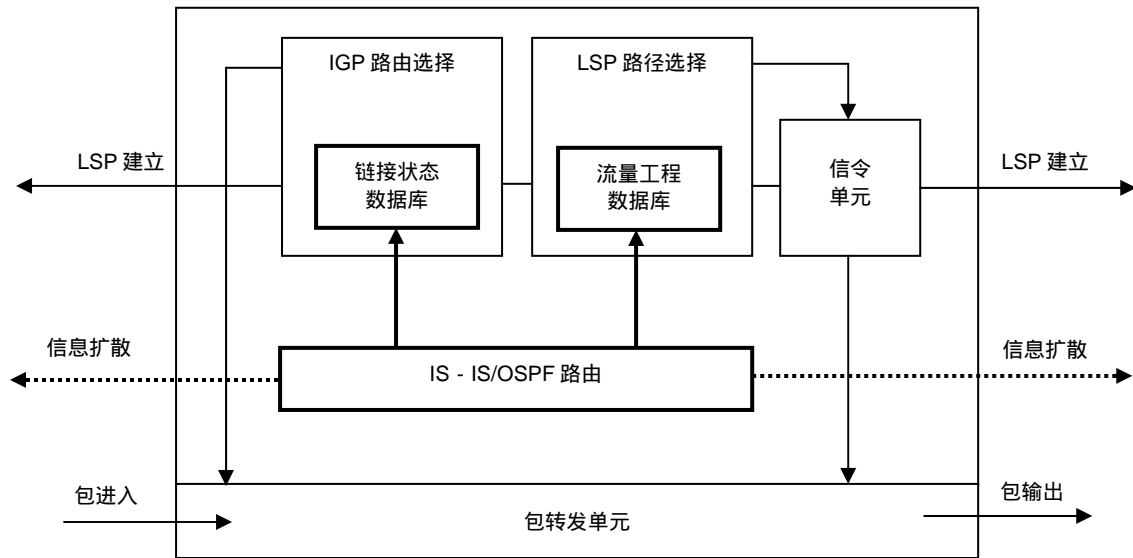


图 10 : 信息发布单元

一些需要加到 IGP 链接状态广播中的流量工程扩展包括：

- 最大链接带宽
- 最短预留带宽
- 当前带宽预定
- 当前带宽使用
- 链接着色

路径选择部分

在网络链接特性和拓扑信息通过 IGP 进行扩散并存储到 TED 中去之后，每个起始 LSR 使用 TED 计算出属于它的穿过路由域的一组 LSP 路径。每个 LSP 的路径可表示成精确的或疏松的外在路由。一个外在路由是通过作为 LSP 物理路径一部分的一系列 LSR 预先设置而成的。如果输入 LSR 确定了 LSP 中所有的 LSR，则 LSP 被认为是通过精确外在路由确定的。如果起始 LSR 只规定了 LSP 中的几个 LSR，则 LSP 是通过疏松的外在路由描述的。同时支持精确和疏松外在路由允许路由选择处理既能在可能的情况下给予最大的自由度，又可以在需要的情况下给予约束。

起始 LSR 通过对 TED 中的信息使用约束最短路径优先 (CSPF) 算法来决定每条 LSP 的物理路径。CSPF 是一种改进的最短路径优先算法，它是一种在计算通过网络的最短路径时，将特定的约束也考虑进去的算法。CSPF 算法的输入包括：

- 从 IGP 获得并在 TED 中维护的拓扑链接状态信息。
- 由 IGP 扩展承载并储存在 TED 中的与网络资源状态有关的特性 (如总链接带宽，预定链接带宽，可用链接带宽，和链接颜色)。
- 从用户设置得到的，用来支持当业务通过建议的 LSP 时所需要的管理特性 (如，带宽需求，最大跳转数，和管理策略需求等)。

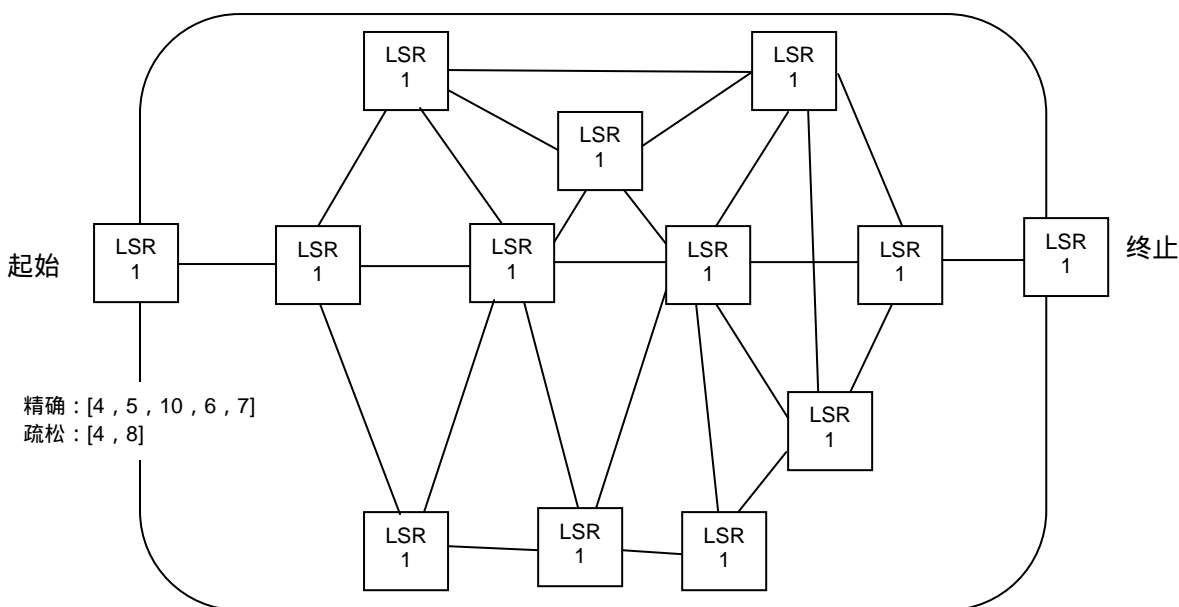


图 11：起始 LSR 计算明确路由

当 CSPF 考虑一条新的 LSP 的每个备选节点和链接时，它可基于资源的可用性或所选部分是否违反用户策略约束而对特定的路径组成部分接受或拒绝。CSPF 计算的输出是一个外在路由，该外在路由包含了一组通过网络的最短路径并满足约束的 LSR 地址。这个外在路由随即传递给信令部分，信令部分在 LSP 中的 LSR 建立转发状态。CSPF 算法在每条 LSP 内被要求发生的起始 LSR 中重复。

尽管在线路径计算减少了管理工作，但为了优化全局流量工程，还是需要离线的计划和分析工具。在线计算将资源约束考虑进去，每次计算一条 LSP。这种实现的挑战是其确定性。LSP 计算的次序在决定 LSP 穿过网络的物理路径时将作为一个重要的角色。早些计算出的

LSP 比早些计算出的 LSP 具有更多的有效资源，因为早先计算的 LSP 消耗了网络资源。如果 LSP 计算的次序改变，则 LSP 的物理路径结构也会随之改变。

离线的计划和分析工具同时检验每条链路对于资源约束以及每条输入 - 输出 LSP 的需求。离线实施可能需要花费几个小时来完成，它提供全局计算，比较每个计算的结果，然后为网络选出一个全局性的最佳方案。离线计算的输出是一系列优化了的网络资源使用的 LSP。在离线计算完成后，LSP 可以以任何次序建立，因为 LSP 的所有安装都是遵循着全局优化方案的规则进行的。

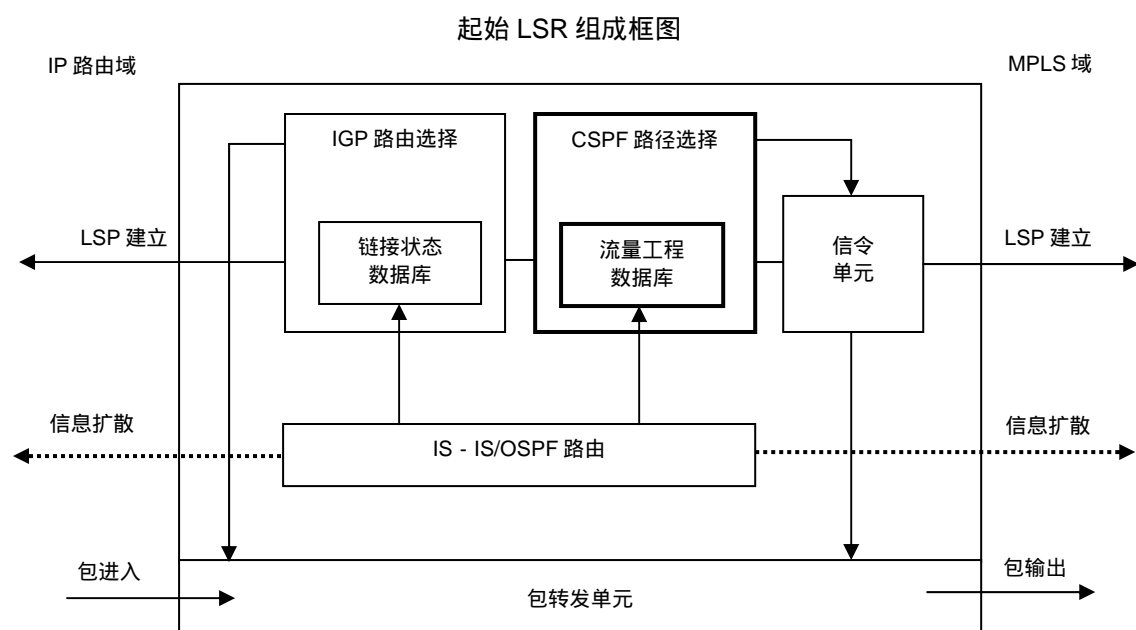


图 12：路径选择单元

信令部分

因为驻留在起始 LSR 的 TED 内关于网络状态的信息在任何时候都是过期的，CSPF 计算出的路径只是被认为是可以接受的。只有在 LSP 被信令部分真正建立之后，才能知道这条路径是否真正可以工作。负责建立 LSP 状态和标计分配的信令部分依赖于资源预定协议

(RSVP) 的一些扩展：

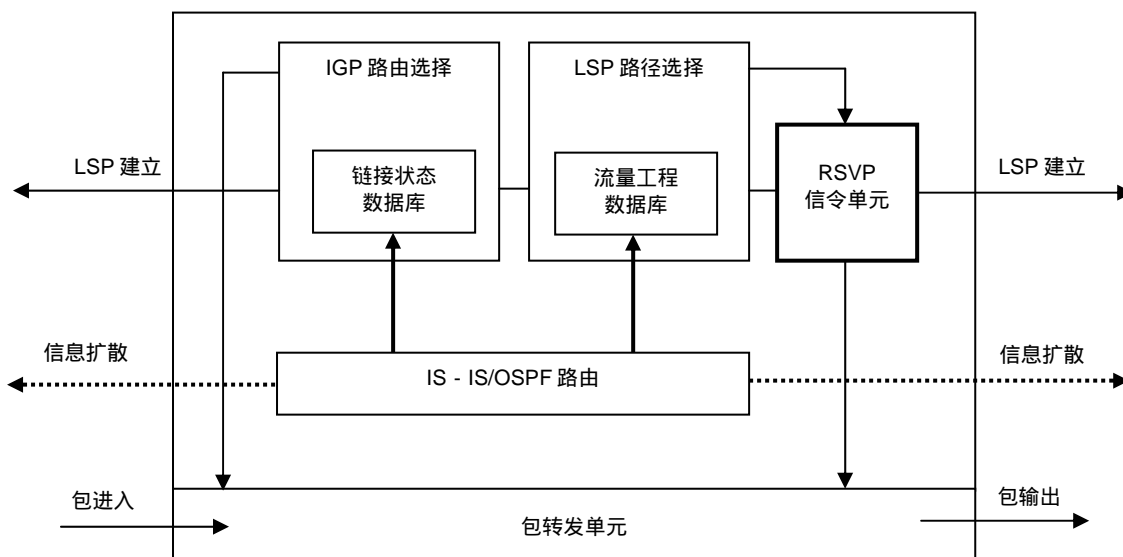
- 外路由对象允许 RSVP 路径 (PATH) 信息在与传统的最短路径 IP 路由独立的外在 LSR 序列中传输。

- 标计请求对象允许 RSVP 路径 (PATH) 信息向中间 LSR 要求提供用于 LSP 建立的标计捆绑。
- 标计对象允许 RSVP 在不改变现存机制的情况下支持标计的分配。因为 RSVP 的 RESV 信息跟随 RSVP 路径信息的预定路径，标计对象支持从下行节点到上行节点的标计分配。

RSVP 作为建立 LSP 用的信令协议是很理想的：

- RSVP 是一个标准的 Internet 资源预定协议，它通过一些附加的新对象类型，特别为支持增强功能而设计。
- RSVP 的软件状态能够在 MPLS 环境中可靠地建立和维护 LSP。
- RSVP 允许将网络资源明确地预定和分配给一条给定的 LSP。
- RSVP 允许建立明确路由的 LSP，它能够提供与原来由 ATM 和帧中继所提供的流量工程和负载均衡能力相等同的功能。
- 穿过 MPLS 域的边缘 - 边缘 RSVP 信令是可扩展的，因为 LSP 与域内边缘 LSR

中间 LSR 组成框图



数量的关系，要比路由表中的项目数和终端系统业务流量的关系密切得多。

图 13：信令单元

灵活的 LSP 计算和配置

流量工程的本质是将业务流映射到物理拓扑上去。这意味着通过 MPLS 提供流量工程的核心是为每条 LSP 决定物理路径。这条路径可通过离线设置来决定或通过在线的基于约束的路由来决定。与物理路径的计算方法独立，转发状态可通过 RSVP 的信令功能在网络中安装。

Juniper 网络基于 MPLS 的流量工程策略支持对 LSP 的不同的路由和设置方式：

- ISP 可以离线地对 LSP 进行全路径计算，并对 LSP 中的每个 LSR 单独地进行带有所需的静态转发状态的设置。这与现在的一些 ISP 们对 IP - over - ATM 的设置相类似。
- ISP 能够离线地对 LSP 进行全路径计算并对起始 LSR 进行静态的全路径配置。起始 LSR 则将 RSVP 作为动态信令协议，为 LSP 中的每一个 LSR 安装转发状态。
- ISP 可依赖基于约束的路由为 LSP 提供在线的动态计算。在基于约束的路由中，网络管理人员为每条 LSP 配置约束，然后，网络自己决定能够最好满足这些约束的路径。正如早先讨论过的，Juniper 网络公司的流量工程策略允许起始 LSR 可以基于确定的约束计算全部的 LSP，随后在网络中对信令进行初始化。
- ISP 可以离线地计算出 LSP 路径的一部分，使用路径中的 LSR 的一个子集对起始 LSR 进行静态的配置，然后，然后允许在线计算决定完全的路径。例如，假设 ISP 有一个包含两条东西向横穿美国的路径的拓扑：一条在在北部通过芝加哥，另一条在南部通过达拉斯。现在，假设 ISP 希望在分别位于纽约和旧金山的两个路由器间建立起一条 LSP。ISP 可以为 LSP 配置部分路径，其包括在达拉斯的一个具有单一疏松路由条转的 LSR，其结果是 LSP 被按照南部的路径路由。起始 LSR 使用 CSPF 计算完全路径，并使用 RSVP 沿着上面的 LSP 安装转发状态。
- ISP 可以毫无约束地对起始 LSR 进行配置。这种情况下，通常的 IGP 最短路径路由被用来决定 LSP 的路径。这种配置不提供任何流量工程的价值。但是，配置非常简单，它可能在虚拟专网 (VPN) 等业务中变得有用。

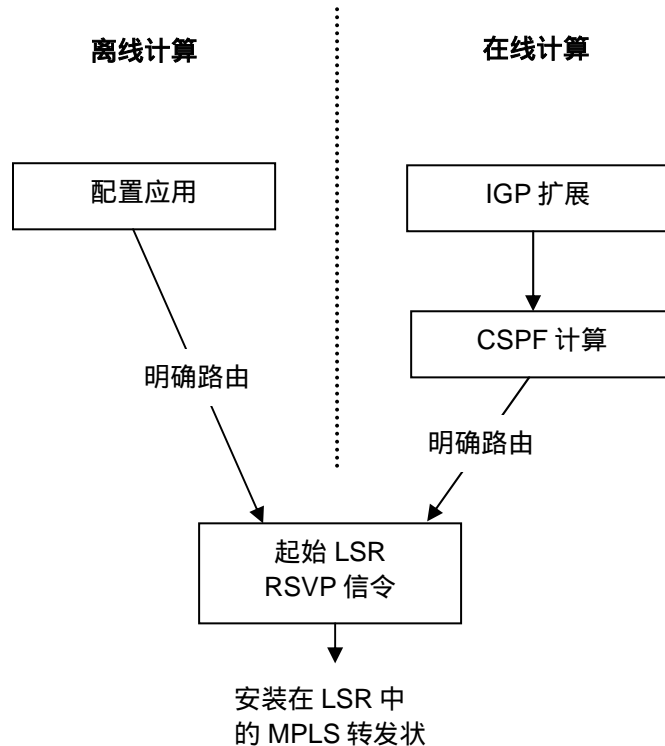


图 14：离线及在线 LSP 计算与配置

所有这些情况中，任何数量的 LSP 都可以定义为主 LSP 的备份。为了在故障情况下建立备份 LSP，两个或更多的实现可被组合在一起。例如，主路径可通过离线进行明确的计算，第二条路径可以是基于约束的，第三条路径可以是无约束的。如果主 LSP 的一条线路出现故障，起始 LSR 从下行 LSR 收到的错误标志信息或 RSVP 软件状态超时注意到故障的发生。起始 LSR 能够动态地将业务转移到一条热备份 LSP 或请求 RSVP 为新的备份 LSP 建立转发状态。

成功的流量工程方案的运行要求

为提供一个强大而且用户友好的工具，Juniper 网络公司的流量工程结构是为能够支持大范围的客户需求而设计的。这种实现需要允许网络操作者：

- 为 LSP 提供许多对于流量工程处理非常重要的特殊操作：
 - 建立一条 LSP。
 - 激活一条 LSP，使其开始转发业务。

- 终止一条 LSP，使其停止转发业务。
- 修改 LSP 的属性（例如带宽，跳转限制，和 CoS）以管理它的性能特性。
- 重新路由 LSP，使其改变通过网络的物理路径。
- 拆除一条 LSP，使网络收回所有分配给 LSP 的资源。
- 为 LSP 配置疏松或精确的明确路由。对于这种选择的支持，允许路径选择处理在可能的情况下获得很大的自由度，或在需要是进行约束。
- 对于给定的 LSP，给出支持它的备选物理路径的次序。例如，可建立一个路径列表，列表的第一条路径被认为是主路径，如果主路径未能建立，则次序列列表中的第二条路径将被尝试建立。
- 在不工作时允许或不允许对 LSP 的重新优化。
- 为一条 LSP 的物理路径定义一组必须明确的被包括或不被包括的资源。一个资源组可被看成是分配给一个链接的一种“颜色”，带有同一种颜色的一组链路属于相同的类。例如，网络策略可以规定一条给定的 LSP 不能够通过金色的链路。
- 按优先级次序建立 LSP，这样可使 LSR 首先建立优先级高的 LSP，然后再建立优先级低的 LSP。
- 决定一条 LSP 是否能够依据 LSP 的属性和优先级从一个给定的物理路径上抢占另一条 LSP。抢占允许网络撤销现存的 LSP 去支持一个新建的 LSP。
- 通过使用基于约束的路由参数，自动获得一个 LSP 布局问题的解决方案。
- 在每条 LSP 的级别上访问计费和业务统计。这些统计值可用于表征业务，最佳化性能，和计划容量。

Juniper 网络流量工程结构的优势

Juniper 网络流量工程结构相对于现在 IP - over - ATM 模型可提供一些优势：

- 支持高速光接口。
- 明确的路径允许网络管理员定义 LSP 通过服务提供商网络的确切物理路径。
- 支持动态故障恢复到一个预先计算的，热备用的备份 LSP。

- 因为 LSP 和基于连接的虚电路非常类似，LSP 可直接用于已有的离线网络计划和分析工具。这些工具的输出可转化为建立 LSP 物理路径的设置。
- 每条 LSP 的统计值将作为将来网络扩容计划和分析的工具，用来分析网络瓶颈和中继线的利用率。
- 基于约束的路由提供了许多增强的功能，它允许 LSP 在其建立前便能满足特定性能的需求。

除了支持并扩展了覆盖模型的优势，Juniper 网络公司的流量工程结构避免了现有的覆盖模型的可扩展性问题上的局限性，允许 ISP 将他们的网络扩展至 OC - 48 及以上的速率：

- 这种结构提供了一个集成的方案，将覆盖模型中的第 2 层和第 3 层网络合并成一个单一的网络。这种集成避免了协调两个分离网络的管理负担，允许路由和流量工程发生在同一平台上，减少了网络的运行费用。另外，LSP 来自 IP 状态，而不是第 2 层状态，所以网络可以更好地反映 IP 业务的需求。由于 ISP 继续增长，Juniper 网络公司的流量工程策略由于在一个单独的集成网络上提供了相同的功能性，因此不必去订购，配置，管理及调试两套不同的设备。
- 这种结构不会因开发 OC - 48 速率的 ATMSAR 路由器接口技术上的挑战而被限制在 OC - 12 的连接上。这意味着缺乏高速 ATMSAR 路由器接口并不能阻止 ISP 将他们的网络速度提高至 OC - 48 或更高。
- 因为 ATM 不再作为第 2 层技术而被需要，信元税被完全避免。这意味着过去被 ATM 信头所占用的 15~25% 的带宽，现在可被用于承载其它的客户业务。
- 基于 MPLS 组成的路由核心网不会有类似于 ATM 的“ N^2 ” PVC 全闭合结网的问题，因此也不会对 IGP 产生压力，进而导致复杂的设置问题。现代的 Internet 骨干网路由器不再出现为保证网络性能而使 ISP 将配置全闭合结网放在第一位的性能问题。
- 这种结构不需要支持交换及虚电路的特殊的第二层技术（ATM 或帧中继）。因此流量工程可在第三层提供，支持混合媒体网络并减少了 IP 和“光纤”之间的层数。

- Juniper 网络公司基于 IGP 扩展，CSPF 路径选择，RSVP 信令和 MPLS 转发组合结构的开发，在不引入新技术的情况下，促进了 IETF 进行的工作。这种方案的发明来自于相关的简单而方便的配置技术的组合，它可以提供与需要更多人为参与的流量工程相同的控制等级，但是只需要很少的人为参与，因为网络本身也参与了 LSP 的计算。
- 最后，Juniper 网络公司的结构为 ISP 如何在其网络中选择流量工程的实现方法提供了极大的灵活性。LSP 可以通过离线或在线计算得到，而且他们可以通过手动或由 RSVP 信令动态地安装到 LSR。对于全局的优化方案仍然需要离线计算。

结论

几年来，Internet 核心网经历着指数级速度的飞速增长。今天，快速增长的业务量迫使一些 ISP 每 3 个月就得使其网络的容量加倍。那些在不断的变化环境中不断增加其市场份额的成功的 ISP 们，正是那些具有洞察力和灵活性的，能够将其骨干网转移至满足不断增长的客户需求的新技术上的人们。

在 90 年代早期，ISP 们依赖于使用量度来对通过路由器核心网的业务流分布进行管理。基于量度的流量控制提供一个满意的流量工程方案，直到 90 年代中期，核心拓扑的备份容量开始限制了方案的可扩展性。同时，在 1994 年或 1995 年左右，ISP 需要增长其网络，配置更宽的通道，并从中间系统中得到确定的性能。这时，ATM 核心交换机和路由器的 OC - 3 及 OC - 12 接口出现了，它们可以提供所需的带宽。这在 90 年代中期成为 ISP 市场发展的一个重要转折点。那些意识到现存结构的局限性，并通过重新设计其网络而转移到覆盖模型的 ISP 们能够平滑地扩展他们的市场份额并增加了利润率。

在 90 年代后期，ISP 在计划将网络升级至 OC - 48 或更高速率时，再次面临选择。继续按照 IP - over - ATM 模型，其结果是可观的费用和增加的复杂性。Juniper 网络公司的流量工程结构提供了 ATM 核心网的业务管理性能，同时又避免了 ATM 的性能和可扩展性的局限。那些考虑到现有 IP - over - ATM 方案的局限性，并且考虑了 MPLS/RSVP 备选方案的优势的 ISP 们能够理解：他们过去的流量工程决定会影响他们网络将来的成长和利润率。