

MPLS 网络流量工程的关键机制剖析^{*}

An Analysis of Key Techniques in MPLS Traffic Engineering

郑伟平 林 生

(华南师范大学计算机科学系 广州 510631)

Abstract In this paper, the technical evolution of network Traffic Engineering(TE) is reviewed at first. Then, the focus is placed on analyzing and discussing several key techniques in MPLS TE. It is finally pointed out that there are some unsolved problems in the implementation of MPLS TE yet.

Keywords MPLS, Traffic engineering, Constraint-based routing, Load balancing, QoS

1 引言

随着 Internet 的蓬勃发展,网络中新技术、新业务层出不穷,Internet 已从单一的数据传送网逐步演化成为综合业务网。网络业务的多元化导致了对网络需求的复杂化,要求网络提供高效、可靠的服务,能够保证服务质量,这对网络流量工程技术提出了极大的挑战,人们迫切需要一种有效且切实可行的解决方案。

MPLS^[1]在无连接的网中提供了面向连接的特性,使 MPLS 在实施流量工程方面有明显的优势。目前,大多数的流量工程解决方案都采用 MPLS 作为转发机制。许多研究人员对基于 MPLS 的流量工程技术寄予厚望。本文介绍了 Internet 流量工程技术的演化,重点对 MPLS 流量工程的关键机制进行分析讨论,并指出了 MPLS 在流量工程实施过程中存在的一些问题。

2 Internet 流量工程的演化过程

2.1 传统基于路由器网络的流量工程

90 年代初,IP 网络是基于路由器的核心网,它使用租用线路把路由器互连起来。该时期的网络规模较小,业务分布较为确定,很少产生临时热点。因此,可采用基于量度值(Metric)的流控技术,通过调节链路的量度值控制业务分布来解决流量工程问题。此方案存在很多局限之处。一是可伸缩性差,容易顾此失彼。大规模的网络中,量度值的调整解决了网络某部分的拥塞问题,但可能在另一处又导致了拥塞。二是路由只考虑简单的量度,忽略网络的负载情况,容易导致业务流过度集中,一些链路出现拥塞,而其它可用的链路则浪费闲置。

2.2 IP over ATM 模型

90 年代中期,为了提高核心网的交换速度,IP 网络普遍采用 IP over ATM 的模型。该模型中路由器被布置在 ATM 域周围,通过在 ATM 交换机上配置 PVC,可以把路由器两两互连起来,路由器看到的是虚拟的网络拓扑,对实际的物理拓扑一无所知。根据链路容量以及历史流量模式和预期流量模式,网络管理系统(NMS)对 PVC 的路径和属性进行整体控制和优化^[2]。NMS 对 PVC 网整体优化计算后,把配置参数下载到路由器和 ATM 交换机上,以构建全连接的逻辑

网。通过配置 PVC 网,可以精确地控制业务流的流动,从而使业务流在各链路上合理分布。ATM 交换机还提供每条 PVC 上的流量统计数据,使网络操作员能够监视流量模式,优化 PVC 的布局和管理。

但 IP over ATM 模型依然存在缺陷。第一,网络服务商必须维护管理两个不同的网络,一个是 ATM 交换网,另一个是覆盖在上面的逻辑 IP 网,这增大了网络管理的复杂度,提高了网络运作成本。第二,该模型需要提供全连接的 PVC 网,故存在 N^2 的问题,可伸缩性较差。第三,存在“信元税”的问题。第四,路由和流量工程运行在两个不同的系统中。即路由发生在路由器上,而流量工程却在 ATM 交换机上实施,使路由和流量工程难以结合起来。第五,ATM 路由器跟不上光通信网带宽发展的要求。ATM 曾引以为自豪的速率已成为其发展的制约因素,网络服务商不得不重新衡量 IP over ATM 模型的利与弊。

2.3 基于 MPLS 的流量工程

MPLS 承诺提供类似 IP over ATM 模型的流量工程能力,但以一种集成的方式提供,克服了 IP over ATM 模型的许多缺点。MPLS 把流量工程分解成三个基本问题^[3]。一是如何把数据包映射到 FEC 上;二是如何把 FEC 映射到流量主干(traffic trunk)上;三是如何通过 LSP 把流量主干映射到物理拓扑上。流量主干指在一条 LSP 上的同类运输流的聚合体,是有具体特征的流量的抽象表示。

MPLS 把流量工程实施为四个功能模块,每个模块都是独立的,模块间设有清晰的接口,改进、更换一个模块并不影响其它模块,具有很大的灵活性。这四个功能模块为:(1)包转发部件;(2)路由信息分发部件;(3)通路选择部件;(4)信令部件。

MPLS 流量工程采用 MPLS 作为包转发机制。MPLS 是一种综合 L2 层交换和 L3 层路由的技术,它把 IP 路由的灵活性、可扩展性和 L2 层的性能、服务质量以及流量管理能力结合起来。其最大的优点是提供了路由与转发的分离,用单一的转发算法就能为多种服务及流量类型使用。MPLS 的转发机制是这样的:当数据包进入 MPLS 域后,入口 LSR 通过包分类器把数据包归类到某一 FEC 中,加上一个 MPLS 头,其中包含与该 FEC 对应的标签,并送到 LSR 的对应接口上转发。当 MPLS 分组到达出口 LSR 时,出口 LSR 把它除去

^{*}本研究获得“综合业务网及关键技术”国家重点实验室开发研究基金资助(202895103)。郑伟平 硕士研究生,研究方向是 MPLS 技术与网络流量工程。林 生 教授。

MPLS 头,再做 L3 层路由转发。数据包在中间 LSR(即 LSP 上除入口 LSR 和出口 LSR 之外的其它任何 LSR)上的转发是一个标签交换的过程,本质上是一个从“入接口—入口标签对”到“出接口—出口标签对”的映射过程。

MPLS 流量工程在计算路由时不但考虑网络的拓扑状态而且考虑了网络的流量负载情况。这要求路由信息分发部件不仅能够分发拓扑信息而且能够分发链路的资源信息。路由信息分发部件可通过简单地扩展现有的链路状态路由协议得到。路由器把通过 IGP 扩展协议得到的链路资源信息存放在专门的流量工程数据库(TED)中,以备计算有资源约束的 LSP 时使用。即使在拓扑没有变化,传统 IGP 不需要发布信息的情况下,IGP 扩展协议也可能因为链路资源状态的变化而需要分发信息^[4],所以 IGP 扩展协议分发信息的频率比传统 IGP 大得多。这加大了网络的负担,造成不必要的开销。我们必须在“尽量准确地反映网络状态”和“避免过多散布通告信息”两者间作出选择。为了避免过度消耗网络资源,我们可以设置一个计时器来限制 IGP 散布信息的频率,或者,如果链路资源的变化量小于某个阈值则不产生 IGP 通告。另外,文[5]中提出了一种在不增加 IGP 通告的前提下提高拓扑数据库准确度的方法。

MPLS 流量工程为流量主干计算通路时使用基于约束的路由。基于约束的路由不同于 SPF,其计算得到的是在满足一系列约束的前提下最短的通路。约束可分成网络资源的性能约束和策略性的管理约束两种。基于约束的路由从服务质量路由^[6]发展而来,但服务质量路由只考虑通路是否满足服务

质量要求,而基于约束的路由不但要满足服务质量的性能约束,还要满足网络管理员的管理约束。基于约束的路由有利流量工程的自动实施,减少了手工对流量工程的配置和干预。然而,基于约束的路由只是根据网络当时的资源状态来计算流量主干的通路,并不考虑对其它流量主干的影响。为了弥补基于约束的路由的不足,可以把基于约束的路由和离线的流量工程工具结合起来使用^[2]。离线的流量工程工具必须知道链路的资源属性以及网络中所有流量主干的要求,它执行全局计算,比较不同的计算结果,力图找到一种对网络整体而言最优的方案。离线工具计算得到一个 LSP 的集合,网络操作员能以任意的次序建立这些 LSP。

IGP 扩展协议和基于约束的路由为 LSR 提供了源端路由计算的能力。计算出通路之后,MPLS 流量工程必须沿该通路建立转发状态。为了支持流量工程基于约束的路由,信令协议在建立 LSP 的同时必须修改链路的资源属性。目前有 RSVP-TE 和 CR-LDP 两个协议可供选择。RSVP-TE 和 CR-LDP 都是在现有协议的基础上扩展得到的。RSVP-TE 和 CR-LDP 各有自己的优缺点,它们之间孰优孰劣一直存在激烈的争论。RSVP-TE 与 CR-LDP 间最大的不同在于它们对传输协议的选择。RSVP-TE 和 CR-LDP 对传输协议的不同选择直接导致了两者在“安全”、“可伸缩性”、“故障检测”等上的差别,文[7]对它们进行了详细、全面的比较。

最后,我们给出了 MPLS 流量工程各功能部件间数据交换的情况,如图 1 所示。图中箭头的方向表示数据流动的方向。

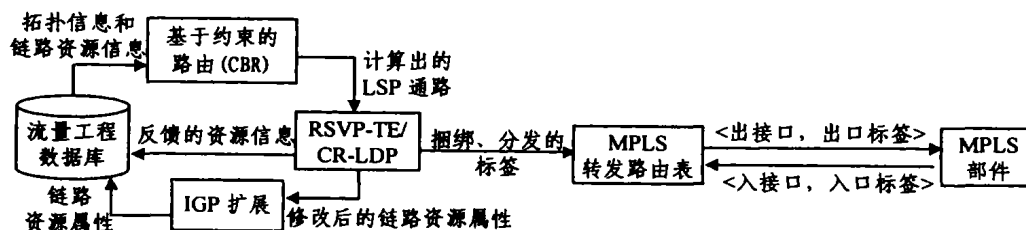


图 1 流量工程各功能部件间的数据交换

3 MPLS 流量工程中的关键机制

3.1 业务保护与恢复

目前,Internet 上开展的实时业务对网络的可靠性和可生存性(Survivability)提出了尖锐的挑战,提供可靠的网络操作已成为流量工程的重要目标之一^[8]。因此,网络经营者应当提供业务保护和恢复机制,使业务的服务质量在发生故障时不会受到太大的影响。传统 IP 网络依靠动态路由协议来恢复故障。当节点或链路发生故障时,IGP 协议把拓扑变化扩散给各路由器,网络收敛后路由器使用 SPF 算法更新路由表,就能将业务流绕过故障点传输到目的地,整个收敛过程需要秒级甚至上分钟的时间。对实时业务而言,目前比较一致认同的故障最大恢复时间为 60 微秒,秒级的恢复时间显然不能满足实时业务的要求。

作为下一代网络的理想技术,MPLS 能够在几十微秒级的时间内恢复故障,提供了类似于 ATM 交换机的故障恢复能力。下面,先给出几个关于业务保护的概念,然后介绍几个基于 MPLS 的保护方案,并对它们进行分析、比较。

3.1.1 保护切换与重路由 MPLS 通过为主通路建立可代替通路的方法来实现业务保护。业务开始时在主通路上

传输,故障发生后业务流被导到可代替通路上,从而保证了流传输的连续性,确保了业务的服务质量。根据可代替通路建立时机的不同,MPLS 业务保护分成两种基本的保护模型:保护切换和重路由^[9]。

保护切换是在故障发生之前就建立好可代替通路的保护模型。出现故障后,它只需通知相关的 LSR 把业务流切换到可代替通路上即可。这个在主通路和可代替通路间负责切换业务流的 LSR 被称为 PSL(Path Switch LSR)。重路由则在故障后才建立可代替通路,可代替通路的路由可以在故障后才计算,或通过通路预计算得到。通路预计算会预先计算出一条符合约束条件的备用通路的显式路由,但不真正建立这一通路。

保护切换只需在故障后通知 PSL 切换业务流,其恢复速度取决于 PSL 发现故障的快慢。相反,重路由时,PSL 收到故障通知后还得计算可代替通路的路由,并使用信令协议建立通路,故其花费的恢复时间比保护切换长得多。但保护切换要求事先建立通路,占用了一定的网络资源,而且网络状态的时刻变化可能使原来建立的备用通路变成非最优的通路。所以,从流量工程的角度考虑,重路由在资源利用和业务量均衡分布方面比保护切换更加有优势。

3.1.2 全局修复与本地修复 根据保护范围的不同,业务保护又可分为:全局修复和本地修复。全局修复的保护范围是整条被保护通路,能为通路上所有的节点和链路提供保护。全局修复在被保护通路的入口 LSR 和出口 LSR 间建立一条可代替通路,该可代替通路一般不与主通路交叉。全局修复的 PSL 一般是入口 LSR,检测到故障的 LSR 需要知会 PSL 进行切换。本地修复的保护范围一般只是某个节点或某条链路,通过建立一条绕过保护点的 LSP 来提供保护。本地修复中检测到故障的 LSR 就是 PSL,所以故障恢复速度比全局修复快,但成本也高得多。

3.1.3 几个方案实例及其比较 目前,已有人提出了多个基于 MPLS 的保护方案。下面介绍 Haskin 方案、Makam 方案和最小费用(Least-Cost)方案,并对其进行比较。

1999 年,D. Haskin 在文[10]中提出了一种保护切换的、全局修复的方案,我们称之为 Haskin 方案。该方案的实现思想是:检测到故障的 LSR 把业务流逆传回入口 LSR,由入口 LSR 将其切换到可代替通路上传输。Haskin 方案要先建立两个 LSP 保护段,它们衔接起来构成主通路的保护通路。第一

个保护段在主通路的倒数第二个节点和主通路的入口节点之间,方向与主通路相反。第二段是一条入口节点和出口节点都与主通路相同的通路,它的其它节点与主通路不交叉。

在文[11]中 S. Makam 提出了一个基于保护切换模型的方案。Makam 方案要事先建立一条可代替通路,入口 LSR 和出口 LSR 与主通路的相同,其它节点则与主通路不交叉。故障发生后,检测到故障的 LSR 通过反向通告树(RNT)给入口 LSR 发送故障通告消息,令 PSL 把业务流切换到保护通路上。

最小费用方案是 Gaeil Ahn 提出的一种基于重路由的、本地修复的方案^[12]。在此方案中,主通路上的每个节点都有一个对应的候选节点集,它包含主通路上该节点之后的所有下游节点(包括出口 LSR)。节点检测到故障后,它利用某些算法(如最宽-最短算法,最短-最宽算法,最短距离算法等)算出本身与候选节点集里的各节点间的候选通路以及各候选通路的费用。在算出的各候选通路中选出一条费用最小的作为保护通路,并通过信令协议建立它,之后把业务切换到此通路上传输。

表 1 几个方案实例间的比较

| 方案实例 | 反应速度 | 丢包情况 | 乱序情况 | 资源占用 | 建立时机 | 保护范围 | 复杂度 |
|--------|---------------|------------|---------------|--------------|--------|------|---------------------|
| Haskin | 快 ☆☆☆ | 很少 ☆☆☆ | 会乱序 ☆ | 开始就为保护通路分配资源 | 故障发生之前 | 全局 | 需配置:2个 LSP 保护段 |
| Makam | 由 FIS 的速度决定☆☆ | 丢失分组 ☆☆ | 几乎不会乱序 ☆☆☆ | 同上 | 同上 | 同上 | 需配置:(1)RNT (2)保护通路 |
| 最小费用 | 较慢 ☆ | 丢失分组 ☆☆ | 很少 ☆☆ | 必要时才分配 | 故障发生之后 | 本地 | 需计算并配置保护通路,路由计算复杂度高 |

从上面的比较可以看出,这几个方案各有优缺点。Haskin 和 Makam 方案对故障的反应速度快,但对资源的耗费也大。而最小费用方案在合理分配流量和资源利用方面更有优势,但反应速度相对就慢些。所以,对保护方案的选择要充分考虑业务特性、所需成本以及对原有协议的影响等各种因素。

3.2 提高拓扑数据库的准确度

由于网络状态的时刻变化,流量工程数据库中的链路资源信息总是“过时”的。其中,链路的带宽最易“过时”,因为任何 LSP 的创建或拆除都会影响链路的有效带宽。我们说拓扑数据库中的有效带宽是“乐观的”,则意味着数据库中的带宽超出了网络中的实际数量。相反,“悲观的”则表示数据库所显示的带宽值比实际有效带宽小。无论“乐观的”还是“悲观的”偏差对路由计算都会产生不良影响。“乐观的”偏差会计算出事实上不符合 QoS 要求的路由来,在 LSP 的建立阶段将会因为节点不能满足 LSP 的资源请求而宣告失败。“悲观的”偏差则会忽视某些链路,造成无法找到最佳通路,甚至完全找不到通路的情况。为了减少拓扑数据库中存在的偏差对路由计算的影响,文[5]中提出了利用信令协议来捎带链路实际有效带宽的机制。

信令协议通常需要给源 LSR 反馈一个确认消息,以告知其 LSP 建立请求的成功、失败或终止。该机制在这些消息中以 TLV 编码的形式捎带了其经过链路的实际有效带宽。当 LSR 收到一个反馈到源 LSR 的消息时,它选择适当的接口后,为与此接口连接的链路构造一个表示现有有效带宽的 TLV,然后插入到反馈消息中,并从此接口转发出去。在收到反馈信息的时候,无论源 LSR 还是中间 LSR 都可以利用反馈消息中携带的带宽信息来更新其拓扑数据库。下次 LSR 进

行源路由计算将使用这些更新过的链路信息,直到再一次 IGP 通告或新的反馈信息到来为止。

这种利用 LSP 反馈提高拓扑数据库准确度的机制有很多优点。首先,它不必增大 IGP“泛洪”(flooding)频度就能提高拓扑数据库的准确度,增大了流量工程 LSP 的建立成功率。例如,当源 LSR 因乐观估计链路的带宽而建立 LSP 失败后,反馈回的通知消息(以 CR-LDP 为例)将更正拓扑数据库,下次路由计算将不会选择那些原先使用的不符合条件的链路,避免犯同样的错误。在这种机制的协助下,源 LSR 不断地尝试和更正,直到要么找到并建立好该 LSP,要么发现网络中并不存在满足该约束集的通路。其次,这种机制对重路由尤其重要。如果重路由时建立通路失败,利用该机制可以及时更新拓扑数据库,从而能继续业务的重路由。否则,只有等下一次 IGP“泛洪”到来后才能继续重路由。另外,在支持抢占机制的网络中,利用该机制可以避免资源抢占的“多米诺效应”,有利于网络的稳定和性能的提高。

3.3 LSP 带宽的自动调整

我们知道,建立流量工程 LSP 时需要预留资源,其中主要是带宽资源。给一条 LSP 预留多大带宽是个难以确定的问题。因为通过资源预留协议给 LSP 分配带宽的方法是静态的,一旦确定就不能改变,太大会浪费带宽,无法充分利用资源,太小又不能满足业务需求,会致使服务质量下降。为此,有人提出了动态分配带宽的机制,即根据其实际使用的量来自动调整 LSP 的带宽。

实际上,这是一种利用前面得到的统计数据来预测后面业务流量的方法。例如,可以把业务的整个传输时间按时间间隔 M(例如 24 小时)分成多段,在 M 时间段内每隔 N 时间

(例如 5 分钟)就测量 LSP 实际使用的带宽平均量,记为 $AB_{n,i}$,表示在第 n 个 M 时间间隔里的第 i 个 N 时间间隔中测量到的带宽平均量。用 BW_{n+1} 表示第 $n+1$ 个 M 时间间隔里 LSP 所分配到的带宽量。 BW_{n+1} 可由下面的公式算出:

$$BW_{n+1} = \text{Max}(AB_{n,1}, AB_{n,2}, \dots, AB_{n,i}, AB_{n,i+1}, \dots)$$

第 $i(i \geq 1)$ 个 M 时间间隔结束后,需要对 LSP 重新分配带宽。它实际上为该业务请求另一条带宽值为 BW_{i+1} 的通路,假如建立成功就把业务从原来的 LSP 切换到该通路上,否则仍使用原来的 LSP 以避免业务中断^[13]。

自动调整 LSP 带宽的方法有利于简化 MPLS 流量工程的实施。一般地,在实施流量工程时,网络服务商必须先知道网络的流量负载模式,然后手工输入网络的拓扑结构以及流量的负载参数,再利用建模工具计算出流量主干的通路,最后配置建立通路。但利用上述机制的话,我们能先以任意的带宽值建立流量主干,再根据流量模式来动态、自动地调整流量主干的带宽,从而简化了流量主干的初始建立。

3.4 负载分布与负载均衡

3.4.1 负载分布 对负载分布的控制是流量工程网络的核心功能。负载分布的优化目标是减少链路或节点上的过载状况,使负载均匀地分布在网络中。负载分布可分成两个部件:一是对路由的标识,即找出业务流可走的通路;二是对流的分类,把业务分配到合适的路由上(在多通路路由的情况下)^[14]。负载分布的路由可以是动态或静态的,即可以是拓扑驱动的 IGP 路由,或者是指定的显式路由。路由部件的任务是在两端点间找出所有存在的可达通路。例如,文[15]给出的支持负载均衡的多通路路由算法。另外,负载分布的路由还应考虑通路建立次序对全局业务负载能力的影响,或者说,在通路选择时要选择一条对往后通路建立影响最小的通路。当两个端点间有多条并行通路时,必须解决如何把业务分配到通路上的问题,这就是所谓的负载均衡。负载均衡是负载分布的特定制子问题,其本质是流的分类问题。

3.4.2 负载均衡 可以是动态或静态的。动态负载均衡基于动态的管理控制,其业务的分配与网络的反馈信息有关,需要测量网络状态以得到某些控制参数。静态负载均衡则仅根据数据包的头部信息来对流分类。对应不同的背景,负载均衡有着不同的实现。

• ECMP(Equal Cost Multipath)。IGP 协议中已包含了负载均衡的实现。当两个端点间有多条费用相等的最小通路时,IGP 在这些通路间平均分配业务流量。IGP 中 ECMP 的流量分配可以是基于分组的或基于流的。基于分组的分配方式一般采用轮循的方法,容易导致数据分组的乱序。而基于流的分配方式对分组头部的某些域(如源地址和目的地址等)进行哈希运算(例如 CRC16),得到一个键值。所用的哈希函数把键值的可能取值空间分成 N 个区间,一个区间对应于 ECMP 中的一条通路。由得到的键值所在的区间就可为该分组确定应走的下一跳^[16]。ECMP 一般在这些等价的通路间平均分配流量,不考虑通路上的负载。而且,ECMP 不能在非等价的多条并行通路间负载均衡,故存在较大的局限性。

• OMP(Optimized multipath)。与 ECMP 不同,OMP^[17]在考虑各链路负载情况的基础上进行负载均衡,从而能得到全网范围的路由优化。为此,OMP 必须对 IGP 协议做相应的修改,使它能够在传播链路的负载信息。根据网络的反馈信息(主要是链路的负载信息),OMP 动态地调节业务流的转发策略。为了避免网络振荡,OMP 提供了一种逐渐调节转发的算

法。该算法逐步地调节转发以确保网络的稳定,又能及时跟踪负载分布状况的变化。

• 基于策略的负载均衡^[18]。在基于策略的网络结构中,策略是规则的集合,规则的形式为:If (条件) Then (动作),表示一定条件的满足将会引起一组操作的执行。策略执行点(Policy Enforcement Point, PEP)和策略决定点(Policy Decision Point, PDP)是基于策略的网络结构中的两个重要元素。PDP 负责作出策略决定,而 PEP 实际执行策略决定。PDP 和 PEP 间的通信由策略协议来支持。

基于策略的负载均衡是策略在网络 QoS 控制方面的应用。在基于策略的负载均衡中,PEP 处于并行通路的共同入口节点上并执行网络的负载均衡策略,负载均衡策略体现了 LSP 对数据流的接纳规则;PDP 则可存在于每个存放有均衡策略的 LSR 上,由它决定流量在通路上的分派。基于策略的负载均衡把负载均衡简化成对适当的策略的定义和选择。因此,基于策略的负载均衡在对分组的分类和对通路的选择上有更大的灵活性,可实现更精细的负载均衡算法,能满足新应用对网络的优化目标^[14]。

4 MPLS 流量工程存在的一些问题

MPLS 流量工程很好地弥补了传统 IP 网络的先天缺陷,优化了网络资源分配,提高了网络的性能。但是,MPLS 流量工程在实施过程中也存在着很多问题,需要进一步地研究和探索。例如:

(1) MPLS 标准尚未成熟,各厂商的设备互通性差,缺乏普遍的互操作性。

(2) 在 MPLS 体系结构中,还不能有效地跨越自治系统和 IGP 区域建立 LSP。

(3) 缺乏对流量策略的统一管理。

(4) 目前,MPLS 流量工程主要采用人为指定通路的方法。但人为配置容易出错,而且这种语义性的错误往往很难检查。而目前实现基于约束的路由的技术还没完全成熟,所以要实现流量工程自动化也困难重重。

(5) 流量工程与 TCP 流控在交互上可能存在冲突。流量工程技术和 TCP 技术都对网络的拥塞作出相应的反应,故很难正确判断一条通路是否已“过载”。

(6) 流量工程的重选通路会增加分组乱序的可能,从而引起 TCP 对分组重排序,严重降低了 TCP 的效率。

结束语 流量工程是一项非常复杂的技术,它涉及到测量、建模、特征描述和流量控制等各方面技术的应用。目前多数关于流量工程的研究还只停留在“约束路由”和“显式路由”上,流量工程的实施面临着重重困难。鉴于 ATM 技术已比较成熟,所以 MPLS 流量工程在发展过程中可以多借鉴 ATM 流量工程技术,另外对一些尚未成熟的机制应在网络测试床或模拟器环境下仿真,以进一步明确其性能。

参考文献

- 1 Rosen E, Viswanathan A, Callon R. Multiprotocol Label Switching Architecture. RFC3031, IETF, Jan. 2001
- 2 Traffic Engineering Solutions for Core Networks. A paper in the Backbone Network and Business Solution Series, Alcatel
- 3 Awduche D, et al. Requirements for Traffic Engineering Over MPLS. RFC 2702, Sep. 1999

(下转第 76 页)

从表中可以看出, IWMS 系统的响应时间与 FastCGI 大致相当, 稍好于 API 的时间, 是 CGI 方法响应时间的两倍, 这主要是因为 IWMS 系统对 CGI 采用了持续连接的处理方法, 并且对 CGI 采用嵌入运行的模式, 考虑到由于 FastCGI 不支持对 CGI 的持续连接处理, 在采用 TCP 慢启动^[9]的网络环境中并且应用程序产生大量的小文件的时候, IWMS 系统的表现应当会优于 FastCGI 的表现。

2) 吞吐量

为了测试吞吐量, 我们将客户端程序直接放在服务器上运行, 这样可以不用考虑网络延迟的影响, 客户端程序在运行后会产生 30 个进程以模拟用户的连接访问服务器, 测试结果如表 2 所示。

表 2 吞吐量

| | | |
|-----------------|-------------------|-------------------------|
| IWMS | 9.2ms/per request | 109 requests/per second |
| Apache/Fast CGI | 8.5ms/per request | 118 requests/per second |
| Apache/API | 15ms/per request | 67requests/per second |
| Apache/CGI | 35ms/per request | 29 reques/per second |

从表 2 可以看出, IWMS 的表现要好于 API 的方式, 并且是 CGI 方式的四倍, 但要稍差于 FastCGI 的吞吐量, 这主要是因为 IWMS 中加入了 session 的支持, 对每个用户的连接都要产生一个 session, 并且不是采用一个持续运行的 CGI 程序, 而是通过 CGI PLUGIN 在每个 session 内部运行, 而 FastCGI 不支持 session 的处理方法, 这方面的开销节省了。

5. 系统安全性的讨论

对于管理系统来说, 网络传输的数据安全性是非常重要的, 目前基于 Web 的管理系统都是通过 SSL 或 TLS 的方法实现 HTTP 数据的加密, IWMS 系统也采用该方法来实现数据安全, 该方法的优点是实现简单, 缺点是应用层协议需要提供额外的端口来支持 TLS, 对于加密的信息和普通信息采用

两种不同的服务并且不能实现面向用户的加密, Rohit khare^[6]提出了通过使用 http/1.1 的 Upgrade 机制来在一个已经存在的 TCP 连接上发起 TLS, 从而使加密和普通的 HTTP 通信共享一个端口的想法, 该想法的另一个好处就是使得提供面向用户、面向需求的加密服务变为可能, 用户不需要在连接的开始就协商加密方法, 而可以在连接的任何时间提出申请, 经过双方的协商进行数据加密, 服务端也可以通过发送强制加密的 Upgrade 标题来拒绝协商, 这种灵活的加密方法可以说是以后基于 Web 方式数据加密的一个方向。

结论 本文通过实现一个面向交互式系统管理的 HTTP/CGI Server 来解决目前基于 Web 的管理系统所面临的一些问题, 实践证明该系统在处理交互式的管理服务时具有较好的运行性能, 它的一些想法如嵌入式 CGI 的运行, session thread 的处理方式, CGI CACHE 的使用同样可以应用在其它的基于 Web 的交互式服务中。

参考文献

- 1 Jamie Cameron. <http://www.webmin.com/webmin/>
- 2 Open Market, Inc. FastCGI: A high-performance web server interface. Technical White Paper. April 1996
- 3 Brabrand C, et al. Aruntime system for interactive Web services. Computer Networks, 1999, 31: 1391~1401
- 4 Coar K. <http://cgi-spec.golux.com>
- 5 Fielding R, Frystyk H. rfc2068, Jan. 1997
- 6 Khare R, Lawrence S. Upgrading to TLS Within HTTP/1.1. Network Working Group Internet Draft January 5, 2000
- 7 Brabrand C. Synthesizing. Safety controllers for interactive web services. [M. Sc. thesis]. Dec. 1998
- 8 Sandholm A, Schwartzbach M L. Distributed safety controllers for web services. FASE98
- 9 Jacobsen V. Congestion Avoidance and Control. In: Proc. SIGCOMM'88 Symposium on Communications Architectures and Protocols, Stanford, CA, Aug. 1988. 314~329

(上接第 67 页)

- 4 Xiao X, Hannan A, Bailey B, Ni L. Traffic engineering with MPLS in the Internet. IEEE Network Magazine, March 2000
- 5 Ashwood-Smith P, et al. Improving Topology Database Accuracy With LSP Feedback. draft-ietf-mpls-te-feed-03.txt, work in progress, Nov. 2001
- 6 Crawley E, et al. A Framework for QoS-based Routing in the Internet. RFC2386, Aug. 1998
- 7 Brittain P, Farrel A. MPLS Traffic Engineering: A Choice of Signaling Protocols. Data Connection Limited, Jan. 2000
- 8 Awduche D O. MPLS and traffic engineering in IP networks. IEEE Communication Magazine, December 1999. 42~47
- 9 Sharma V, et al. Framework for MPLS-based Recovery. draft-ietf-mpls-recovery--frmwrk-03.txt, work in progress, July 2001
- 10 Haskin D, et al. A Method for Setting an Alternative Label Switched Paths to Handle Fast Reroute. draft-haskin-mpls-fast-reroute-00.txt, work in progress, June 1999
- 11 A Path Protection/Restoration Mechanism for MPLS Networks. draft-chang-mpls-path--protection-03.txt, work in progress
- 12 Ahn G, Chun W. MPLS Restoration Scheme using Least-Cost based Dynamic Backup Path
- 13 Cisco MPLS AutoBandwidth Allocator for MPLS Traffic Engineering: A Unique New Feature of Cisco IOS software. White Paper, Cisco Systems
- 14 Wright S, et al. Traffic Engineering of Load Distribution. draft-wright-load-distribution-00.txt, work in progress, July 2000
- 15 Lee S S, Gerla M. Fault tolerance and load balancing in QoS provisioning with multiple MPLS paths. Lecture Notes in Computer Science, vol. 2092
- 16 Hopps C. Analysis of an Equal-Cost Multi-Path Algorithm. RFC2992, Nov. 2000
- 17 Villamizar C. OSPF Optimized Multipath (OSPF-OMP). draft-ietf-ospf-omp-02.txt, work in progress, Feb. 1999
- 18 Wright S, et al. Policy-Based Load-Balancing in Traffic-Engineered MPLS Networks. draft-wright-mpls-te-policy-00.txt, work in progress, June 2000