

广域网络冗余机制实现方式的分析与比较

时长江 (山东检验检疫局)

方 军 (国家质检总局)

摘要:分析单条窄带链路组成的广域网络应用面临的问题,比较广域网络冗余机制的四种模式的实现方式和条件、优点和不足。为广域网络冗余机制下的多媒体应用如何实现负载均衡、出现故障时自动切换等不同需求提供了相关参考模式。

关键词:负载均衡 广域网络 双链路自动切换

1 应用背景

随着信息技术的迅猛发展,企业级窄带广域网络单条线路上的应用不仅包括各种业务应用系统、办公自动化系统、电子邮件、WEB 网站信息发布服务等,还包括网络电话(VoIP)、视频会议和可视电话等多媒体应用。这些业务相互融合,相互集成,构造了在 IP 网络上的多媒体应用。以单条线路架构的窄带广域网络应用面临着诸多问题:如何合理分配网络资源,保证数据业务正常运行的同时,又能保证视频会议图像稳定传输,网络电话通过质量清晰;如何保证多媒体应用效果的同步和一致;如何保证广域网络上关键应用端到端的服务质量(QoS);如何避免单条链路或广域网络设备的单点故障,保障业务系统不受影响。

在资金允许和业务量不断增加的情况下,可以考虑采用广域网络双链路、双路由器与局域网络双交换机组成广域网络冗余机制,保证业务系统可靠安全运行。

2 广域网络冗余机制的实现目的

保证两条链路的高可用性,双链路互为生产线,分别承载不同业务;双链路互为备份,当一条链路或网络设备出现故障时,该条链路上运行的应用系统可以自动切换和移植到正常链路上。当链路或网络设备恢复正常后,应用系统也自动回切。

保证两条链路带宽的同时复用,实现业务动态负载均衡或任务分担,保证带宽的有效利用。将所有的应用系统在两条链路上进行负载分配,例如第一条链路承载数据业务系统,第二条链路承载网络电话和视

频会议系统,实现静态负载分担,通过策略路由进行应用分流,避免各种业务在同一链路上传输、争用网络带宽而造成网络拥塞现象。也可以根据两条链路的负载流量,平均分配多媒体业务应用系统在两条链路上运行,实现双线路负载均衡。或者在静态负载分担基础上,增加动态负载策略,提高广域网络冗余机制下每条链路和网络设备的利用率。

不仅能够保证每条线路上的应用系统的 QoS,也同时保证出现故障的线路上的应用系统,自动切换到正常链路上后的各项应用系统的 QoS。

3 广域网络冗余机制的拓扑结构及组成

广域网络冗余机制是指由双链路、双路由器、双网络交换机组成的网络。其拓扑结构如图 1 所示。

中心点采用两台 3 层交换机 S1 和 S2 作为核心交换机,它们之间以多条链路采用 TRUNK 技术进行合并为带宽的逻辑链路连接,提供了线路冗余功能。在三层网络设备上设置合理的 VRRP/HSRP,根据应用类型划分不同的 VLAN, S1 作为 VLAN1 的 Active (S2 Standby), S2 作为 VLAN2 的 Active (S1 Standby), 依此类推。路由器 R1、R2 的分别配置两块以太网卡,通过 Tracel、Trace2、Trace1、Trace2 连接连到 S1 和 S2。R1 和 R2 上启动 OSPF、策略路由等路由协议的设置,使双线路、双网络设备组成广域网络冗余机制系统能够自动识别任意一点的故障,并进行动态切换,从而确保了网络的可靠使用。VRRP/HSRP 实现了设备的相互切换,路由协议实现了路由的自愈与迂回,二者的相互结合,确保了

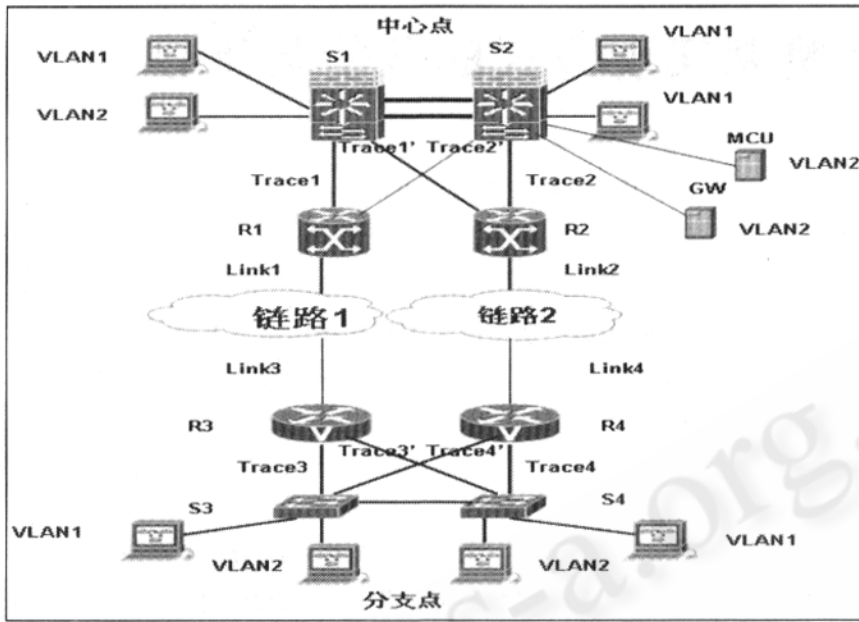


图 1 双线路拓扑结构图

广域网络冗余机制的实现。分支点的配置情况与中心点基本一致。

4 广域网络冗余机制实施方案比较

4.1 策略路由

策略路由是通过对数据包中的源 IP 地址或目的 IP 地址的识别与控制,使数据按规则中要求的路径上进行传输,实现应用分流。在三层交换机上,针对数据业务 VLAN1, VRRP/HSRP 的 ACTIVE ROUTER 设在 S1 上, S2 作为该 VLAN1 的 STANDBY ROUTER。视频业务 VLAN2, VRRP/HSRP 的 ACTIVE ROUTER 设在 S2 上, S1 作为此 VLAN2 的 STANDBY ROUTER。

针对数据业务,在 S1 和 S2 交换机上连接数据业务网段的 VLAN 端口上设置规则是该网段(源 IP)与分支点(目的 IP)的数据业务网段通信时,其路由网关(下一跳地址: NEXT HOP)分别指向与 S1、S2 连接的 R1 路由器的端口。针对多媒体业务,在 S1 和 S2 上连接多媒体业务网段的 VLAN 端口上设置规则是该网段(源 IP)与分支点(目的 IP)的多媒体业务网段通信时,其路由网关(下一跳地址: NEXT HOP)分别指向与 S1、S2 连接的 R2 路由器的端口。

在 R1 上对数据业务进行策略路由配置,在连接 S1 和 S2 的 LAN 端口上进行设置,与分支点数据业务间的

通信时,其网关指向分支点 R3。在 R2 上对多媒体业务进行策略路由配置,与分支点多媒体业务间的通信,其网关指向分支局的 R4。网络运转正常的情况下,数据业务走 S1→R1→Link1→Link3→R3→S3,多媒体业务走 S2→R2→Link2→Link4→R4→S4。

如果网络中任意一点的线路发生通信故障,系统会自动识别,并根据实际情况,调整路由策略。当中心点的 VLAN2 不能与 S2 进行正常通信时,该 VLAN2 的 VRRP/HSRP 会自动切换, S1 将成为该 VLAN2 的 ActiveRouter。多媒体业务的信息传输将首先发往 S1,并按照 S1 上的策略路由规则,经 R2 传输到分支点。

当 S2 与 R2 之间通信中断时, VLAN2 的 ActiveRouter 也将随之变更。VLAN2 的 VRRP/HSRP 在监测本 VLAN2 与 S2 通信状况的同时,还可监测 S2 与 R2 之间的通信状况,在任一连接出现通信故障时, VLAN2 的 VRRP/HSRP 都会进行自动切换,此时 S1 将成为该 VLAN2 的 ActiveRouter。多媒体业务的信息传输将首先发往 S1,并按照 S1 上的策略路由规则,经 R2 传输到分支点。

当 R2 发生故障,不能与 S1、S2 通信, VLAN2 的 VRRP/HSRP 监测到 S2 与 R2 之间失去连接, VLAN2 的 ActiveRouter 进行切换, S1 将成为 ActiveRouter,多媒体业务信息首先发给 S1。根据策略路由本应再转发到 R2 上,但因为 R2 故障,导致 S1 无法与策略路由器中定义的下一跳地址进行正常通信,因此, S1 将按照路由表中的路由,将多媒体业务信息转发给 R1。直到 R2 故障恢复后, S2 发觉到与 R2 能够进行正常通信时, VRRP/HSRP 进行切换,信息仍将按定义中的规则进行信息传输。

Link2 不能正常通信时,中心点的多媒体信息经 S2 到达 R2 后,由于 R2 无法与策略路由中定义的下一跳地址(R4 端口地址)进行通信,其将按本地的路由表为准进行信息传输。传输的路径为 S2→R2→S1→R1→Link1→link3→R3→Trace4→S4

策略路由只能实现静态的负载均衡,不能实现动态的负载均衡;当不使用视频会议时,数据业务流量也

不能自动转到 Link2、Link4,造成了网络资源浪费。如果采用二层交换机,在链路出现故障时,造成 R3 和 R4 之间的路由信息循环传输,导致目的网络不可达,无法实现链路的故障自动切换。因此必须有策略路由的三层交换机作支持,这样会增加设备投入成本。

4.2 动态负载均衡

链路的负载均衡是指路由器学习到两条以上、优先级相同(OSPF COST 值)路由时,应用将会同时在两条或多条线路上进行传输,即为线路的负载均衡。当其中一条线路出现通信故障时,系统会自动识别并继续使用另外线路,一旦故障线路恢复通信,应用仍将使用所有的正常线路传输。COST 值可以手工改动,在路由器上将通过不同线路学习的、到达同一目的地的 OSPF 路由的 COST 值改为一致,即可实现线路的负载均衡。

分支点的 R3 和 R4 间启动了 VRRP/HSRP,分别以 2M 的 Link3 和 Link4 上连到中心点,R3 作为 ACTIVE 路由器,R4 作为 STANDBY 路由器,分支局的 VLAN1 欲与中心点 VLAN1 进行数据业务通信时,数据将首先发到 R3,分别通过 R1 和 R4 学习路由信息,选择到中心点的路由。由 R1 学习的路由 $COST1 = 108/108$ (Trace1 速率为 100M) + $108/2 \times 106$ (Link1 速率为 2M); 由 R4 学习的路由 $COST2 = 108/108$ (Trace2 速率为 100M) + $108/2 \times 106$ (Link4 速率为 2M) + $108/108$ (Trace4 速率为 100M)。与 R1 连接的 R3 WAN 端口、与 R4 连接的 R3 LAN 端口,通过手工分别设置 $cost1 = cost2$,这样 R3 通过各方向学习的路由 COST 值始终一致,业务系统可以分别沿 R3→Link3→Link1→R1→S1(S2), R4→Link4→Link2→R2→S1(S2) 两条线路进行传输,实现负载均衡。

R1、R2、R3 和 R4 对数据转发的负载均衡方式有两种:一种是以数据到达的目的地为标准,进行负载均衡。分支点的 LAN1 192.168.101.0 的多台 PC 欲与中心点的同一目的地通信时,R3 路由器将来自 192.168.101.1 的数据转发给 R1,而将来自 192.168.101.2 数据转发给 R4,R4 向转发到 R1。依此类推,192.168.101.3 的数据经 R3 到 R1,192.168.101.4 的数据经 R3 到 R4 再到 R2 等等,从而实现线路的负载均衡,这种方式对路由器资源的消耗较小。另一种是“Per Packet”方式,是以 PACKET 为标准,对数据中的每个 PACKET 进行负载均衡传输,R3 将数据的 PACKET1 转发到 R1 上,PACKET2 则被转发到 R4,向上再经 R2 到达目的地,PACKET3

转发给 R1,PACKET4 转发到 R3,依此轮循转发。此方式对路由器资源的消耗较大,到达同一个目标的数据包被分成两条路径走。

如果 Link1→Link3、Link2→Link4 为不同类型的两条链路,比如分别为 SDH 和 ATM 时,由于协议开销和延迟不一样,必然导致这每个数据包在到达目的地重新组合时,造成语音通话质量不连续或视频会议的抖动和视频画面与语音的“唇音不同步问题”,影响多媒体应用效果。当网络拓扑发生变化(比如某条链路出现故障的时候),OSPF 路由会发生收敛,这时需要发送大量的 LSA 报文,将会消耗更多的网络带宽;OSPF 路由协议收敛需要 5-10 秒才能重新形成稳定的路由表,造成一定时间的网络服务中断。

4.3 BLGP (Gateway Load Balancing Protocol)

考虑到广域网络的安全,在 S1 与 R1、S2 与 R2、S3 与 R3、S4 与 R4 之间加设防火墙,针对协议端口制定安全策略。视频会议系统和 Voip 支持协议--H.323 标准与防火墙的安全策略存在着矛盾,可以将视频终端 MCU 直接连在路由器的局域网口,Voip 的语音模块内置到路由器中。视频和网络电话缺省地固定一条链路上运行,数据业务在两条链路上参与负载分配。传统的 VRRP/HSRP 的处于 Active 的三层设备工作,而 Standby 的三层设备不工作,造成设备一部分工作,另一部份闲置。虽然可以使三层设备针对不同的 VLAN 设置为 Active,但是不同 VLAN 之间的数据流量互不相同,因此,策略路由方式不能实现真正意义的负载流量均衡分配。为克服以上不足,CISCO 公司推出了 BLGP (Gateway Load Balancing Protocol)。以分支点为例,BLGP 可以使在 GLBP 中的三层设备 R3、R4 在一个网段 192.168.101.0 内,都处于 Active 状态,共同参与负载均衡分配。R3、R4 通过优先级,选出 R3 为主,R4 为辅,虚拟为一个特殊的网关地址(vIP),并产生 R3、R4 的虚拟 MAC 地址,vMACr3 和 vMACr4。不同应用不必划分不同 VLAN,每个客户端以网关地址 vIP 向 R3 发出请求,R3 和 R4 根据双链路流量情况,分别将 R3、R4 的 vMAC 以 ARPReply 方式反馈给相应的客户端,并作为客户端的 ARP 地址,客户端按照 ARP 地址找到 vMAC 地址对应的 R3 或 R4,建立并确定数据传输路径,实现数据业务的负载均衡。

在不开视频会议和网络电话时,业务数据流量可

以平均占用两条线路 2Mbps 的资源,充分利用双链路的网络资源,保证业务系统可靠运行。在开视频会议和网络电话时,仍然可以保证在该链路有 2M - 768K (视频) - 48K (4 路语音) = 1184K 带宽资源被数据业务使用。实现了链路的静态负载分担的基础上的动态负载均衡,要求必须使用 CISCO 的网络设备。

采用 IP Precedence 优先权技术,将视频、关键业务、网络电话、HTTP、电子邮件、文件 FTP 由高至低划分 6 个优先级,分别提供不同级别的 QoS 服务。在有限的线路带宽,保证服务级别高的应用正常可靠运行。

4.4 智能交换机

在 S1 与 R1、S2 与 R2、S3 与 R3、S4 与 R4 分别设置七层智能交换机,如 Radware 设备,为避免单点故障,两台 Radware 运行 VRRP 协议,互为热备。通过 Radware 的智能负载均衡技术,智能的将所有出入网络的流量以会话为单位,“均衡”地分配到两条专线上。Radware 的健康检查机制监视对应节点的 IP 地址(3 层机制)、远端服务器服务端口(4 层机制)、远端服务器的应用(7 层机制),在该链路出现故障时,仍然能够

快速地切换到另一条正常的链路上去。Radware 根据不同的源地址,目的地址,以及应用的不同选择特定的链路,根据链路的负载情况,网络延时,实现多链路的实时动态的负载均衡。

5 总结

特性 项目	策略路由	负载均衡	BLGP	智能交换机
负载均衡	静态	动态	动态	智能
三层网络交换支持	要求	不要求	不要求	不要求
线路自动切换	能	能	能	能
设备投入成本	比较高	较低	较低	高
异种网络设备互连	支持	支持	不支持	支持
端默认网关	多个	多个	一个	多个

参考文献

- 1 Tere Parnell, 广域网建立与管理指南, 张侃等译, 机械工业出版社, 1998 年 5 月。
- 2 Srinivas Vegesna. IP 服务质量, 人民邮电出版社, 2001 年 8 月。