工程与应用

大型多POD新型城域网的分布式出口路由模型实践

方鸣[1]，朱磊[1]，岳志超[2]，翁雯倩[3]
（1. 上海电信网络操作维护中心，上海 200120；
2. 上海电信智能云网操作维护中心，上海 200120；
3. 上海电信数字集成部集成业务管理部，上海 200120）

摘　要：分析了多POD新型城域网在面临互联网南北向流量依旧占据半数以上带宽的情况下，互联网出口的路由设计所遇到的问题，进而提出了一种分布式出口的组网架构。该方案的原则是在最大限度上不改变现有城域网、主干网内涉及城域互联网业务的路由策略，而是以此为基础叠加新型城域网所需的分布式出口的策略。同时，采用各区域POD统一的策略逻辑，从而不因为POD数量的多寡而增加配置的复杂性。在框架设计和路由逻辑设计上，对该组网架构进行了详细介绍，并基于实践进行了较为全面的效果评估。特别说明了三级汇聚对互联网业务的精确引导、新老城交互影响的缺省路由、Super-spine带来的逃生通道能力，为其他各省新型城域网的部署提供借鉴和依据。

关键词：CR；新型城域网；eBGP；vBRAS CP/UP；一级汇聚/二级汇聚/三级汇聚
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2025091

0　引言

时至今日，新型城域网已从一个概念，迅速转变为组网实践。在各大电信运营商，百万级宽带用户的新型城域网在全国各地推广扩展。新型城域网具有模块化、标准化的组件进行构建，实现了网络的弹性扩展和业务的快速部署。大型城域网往往以50万宽带用户为单位，将网络分解为多个标准化的区域POD模块，在POD内部署支撑家宽、政企业务所需的网络资源，以及边缘云的计算存储资源；再通过一个城域级的超级POD打通各区域POD之间的网络连接，并部署覆盖全市的云计算、存储资源。

当然在新型城域网实践中也发现，将传统的扁平化城域网解耦为多POD的架构后，虽然能更为灵活地适配边缘云的资源，快速部署云内应用，但是如果在一段较长的时间内，云内应用尚不足以替换互联网应用的情况下，多POD反而面临南北向互联网流量在POD之间大量对穿的问题。如果将这部分南北向流量，从各区域POD汇总到城域超级POD，再引导到互联网，并将大量占用POD间原规划用于东西向云间业务的中继带宽，并极大堆叠对城域超级POD内SPINE节点的转发压力。

因此，需要找到一个既能维持新型城域网多POD的云网架构，又能快速有效疏导互联网流量的解决方案。

1　分析遇到的问题

1.1　传统城域网的互联网业务路由发布和流量引导

在传统城域网内，由网络边缘设备，如宽带

收稿日期：2024−11−12；修回日期：2025−02−17

接入服务器（BRAS）、服务路由器（SR）、多业务边缘（MSE）等，提供互联网接入服务。由城域网核心路由器（CR）连接所有边缘设备，并提供城域汇总出口的能力。城域网CR连接国家级主干网的C主干网设备；国家级主干网建有独立的、由国际层路由器I设备构成的国际层，与其他国家、跨国企业的主干网络进行互联，从而构成国际互联网的实体。

城域网内以网络边缘设备为源，如图1所示，将IP网段逐级汇聚，输出至国家级主干网络，从而引导全球互联网流量进入城域网。具体过程如下。

（1） 在网络边缘设备，以用户为单位分配业务级网段，如家庭宽带用户，分配/32的一段IPv4网段，和一段/56的IPv6网段；政企专线用户，分配/30的一段IPv4的LINK网段+*N**/24的多段IPv4的LAN网段，以及/64的一段IPv6的LINK网段+*N**/56的多段IPv6的LAN网段。

（2） 在城域网核心路由器CR设备，将业务级网段，进行两级汇聚；二级汇聚网段，一般将业务级网段汇聚为/18以内的2的次方个/24网段，以及/48以内的IPv6网段；一级汇聚网段，一般是将二级汇聚网段，进一步汇聚为/17或/16的IPv4网段，以及/40的IPv6网段。

（3） 城域网核心路由器CR，通过eBGP（跨域边界网关协议）的动态路由，同时向国家级主干网的C设备输出一级汇聚网段和二级汇聚网段；C设备侧对路由进行标识，设置不同的路由有效范围；其中二级汇聚网段仅在各地C设备构成的国内区域内有效，一级汇聚网段则除了在国内区域有效，还会在I设备构成的国际交换区域有效，既一级汇聚路由还会通过eBGP输出给其他国际、企业的网络。

城域网内互联网业务流量引导过程如图2所示，国际互联网资源从国外流入后由国家级主干网I设备（识别一级汇聚路由）——国家级主干网C设备（识别二级汇聚路由）——城域网CR设备（识别业务网段路由）——城域网边缘设备（识别业务路由）的路径，到达最终消费流量的用户。

1.2 新型城域网遇到的问题

与协议单域、网络扁平化的城域网不同，新型城域网采用模块化架构，可按照业务需求实现乐高积木式的弹性伸缩。这种弹性主要是基于新

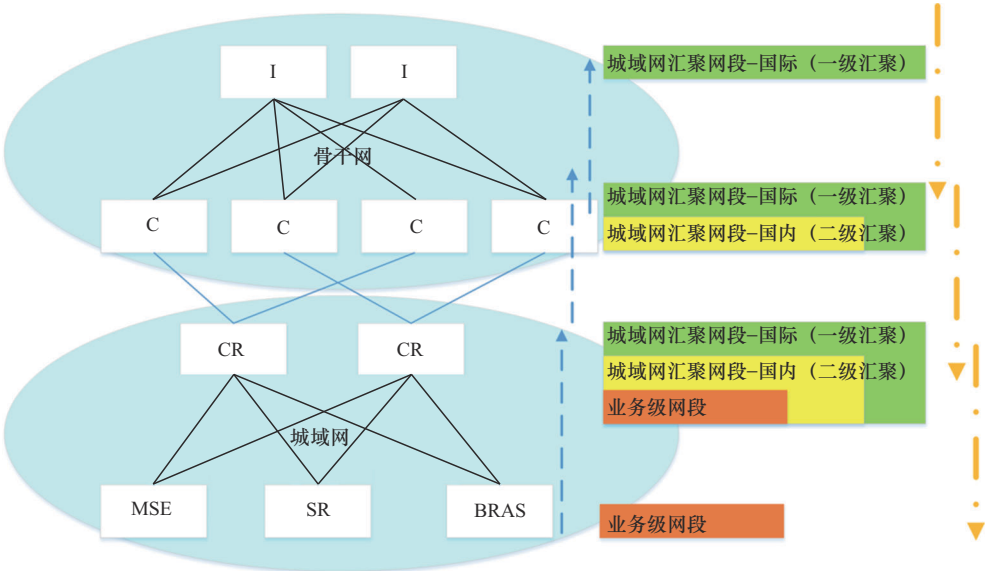

图1 传统城域网的互联网业务路由

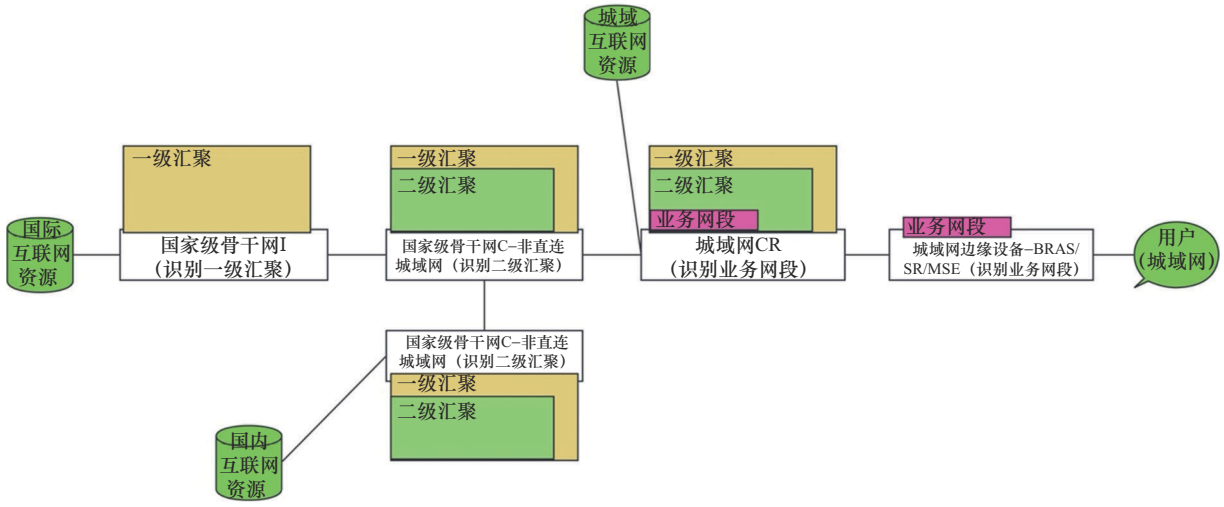

图2　城域网内互联网业务流量引导过程

型城域网内的业务网段在不同POD模块之间动态调度来实现的，也由此造成新型城域网内的IP业务网段呈现离散化和不断动态迁移的表征。

以图3为例，一个大型城域网部署1~*n*的若干个区域POD覆盖全市业务，以及1个POD0的super-pod串联各区域POD。在该网络中，互联网业务网段部署主要分为2B和2C两种模式。2B政企专线互联网业务直接部署在区域POD的接入

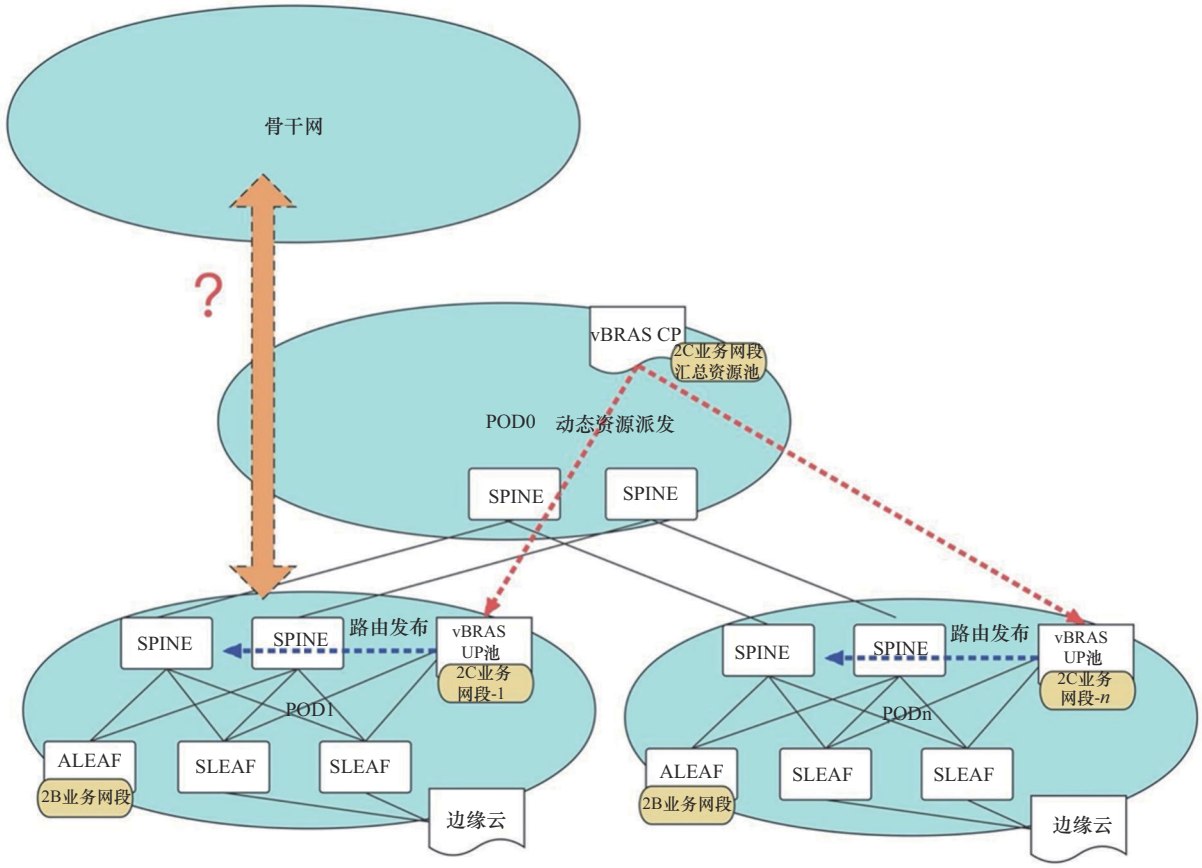

图3　新城城域网的互联网业务路由分布

LEAF节点，一般较为分散，也支持携地移机，以满足客户业务需要长期、固定使用IP地址的需求。2C家庭/政企拨号互联网业务则部署在POD内的vBRAS UP池上。城域网内，2B和2C的流量比例大致在1:10左右，2C业务占据大部分的带宽。因为2C引入了转控分离的、池化的新型vBRAS功能组件，相比2B其调度能力更为强化。

NFV网络虚拟化的宽带接入服务器vBRAS具有CP和UP两个组件，其中CP负责控制平面（Control Plane）的信令管理，UP则负责用户平面（User Plane）的媒体转发。

新型城域网内vBRAS调度2C业务网段的逻辑如下。

（1） 在vBRAS的CP，通过CP-UP管理接口调度整个城域网2C业务的汇总IP网段；需注意，CP的分配是资源层面的，并不体现在协议和媒体上。

（2） vBRAS CP根据各区域POD内UP池内的保护关系，将2C家庭拨号按照每用户为单位，均匀分配到成对或成组的UP设备上；同时，按照业务预设逻辑，为每台UP分配业务网段，该业务网段从CP的汇总IP网段中拆分；UP将业务网段在BGP路由协议中对外发布。

（3） vBRAS CP持续关注各UP的运行情况，并按照业务增减情况，增派或回收业务网段。

传统基于城域范围的一级、二级汇聚路由难以应对这种基于POD的、基于在线用户数量的动态调度机制。其遇到的问题有两个，一个是控制层面，如何避免POD内IP路由信息以日乃至小时为单位的频繁变化，影响与主干网之间原来相对稳定的eBGP邻居关系；另一个是媒体层面，互联网业务流量分布在每个POD内，保持原来城域网汇总后单一出口的架构代价极其高昂。

2 多POD分布式互联网出口的实践

2.1 基本框架设计

在维持新型城域网多POD模块化组网结构不变的前提下，需要为每个POD设计独立的互联出口连接。在如图4所示的连接关系中，POD的一部分出口连接城域网CR，确保本地化的互联资源能够继续就近疏导；另一部分出口连接到国家级主干网C，以访问国内和国际的互联网资源。这样做保留了城域网CR与主干网C原有的出口连接，一方面是考虑新老城业务的迁移是一个阶段性的工作，不可能通过一两次割接来完成；另一方面则是预留了城域网核心路由器CR未来向新型城域网POD0的super-spine角色转换的空间。

对于新型城域网各POD，可以根据各省互联网资源本地化率的情况，规划这两组出口中继的带宽分配。例如，按照一般大型城域网配套有较多IDC资源的情况，取本地化率中间值40%作为依据。单POD 50万2C宽带拨号用户，峰值在线率90%，用户平均峰值流量4 Mbit/s，则该POD的峰值流量为4 Mbit/s（单户）×50万（用户）×90%（并发）=1 758 Gbit/s；2B用户预测为2C的10%，峰值流量为175.8 Gbit/s。两者合计1 933 Gbit/s，其中773.2 Gbit/s可以本地化解决，余下的1 159.8 Gbit/s需要通过主干网到城域网之外寻找资源。折算带宽利用率阀值需低于80%并以2×100GE的倍数取整，则如图4中灰色实线所示，该POD需要配备到国家级主干网16×100GE的出口中继，到城域网10×100GE的出口中继。

注意，算法用到了互联网信息本地化率、宽带用户平均峰值带宽和峰值在线率、2C和2B流量比，这3个业务规划参数，以及中继利用率低于80%这个维护指标参数。另外，主流中继带宽规划依旧为100GE。

2.2 路由逻辑设计

多POD分布式出口设计从物理连接的角度看，需要确保各POD可以直接疏导城域内和跨域的互联网业务流量。但从路由策略的角度，如图5所示，则会面临POD与POD之间、POD与城域网CR之间、主干网与POD和城域网CR之

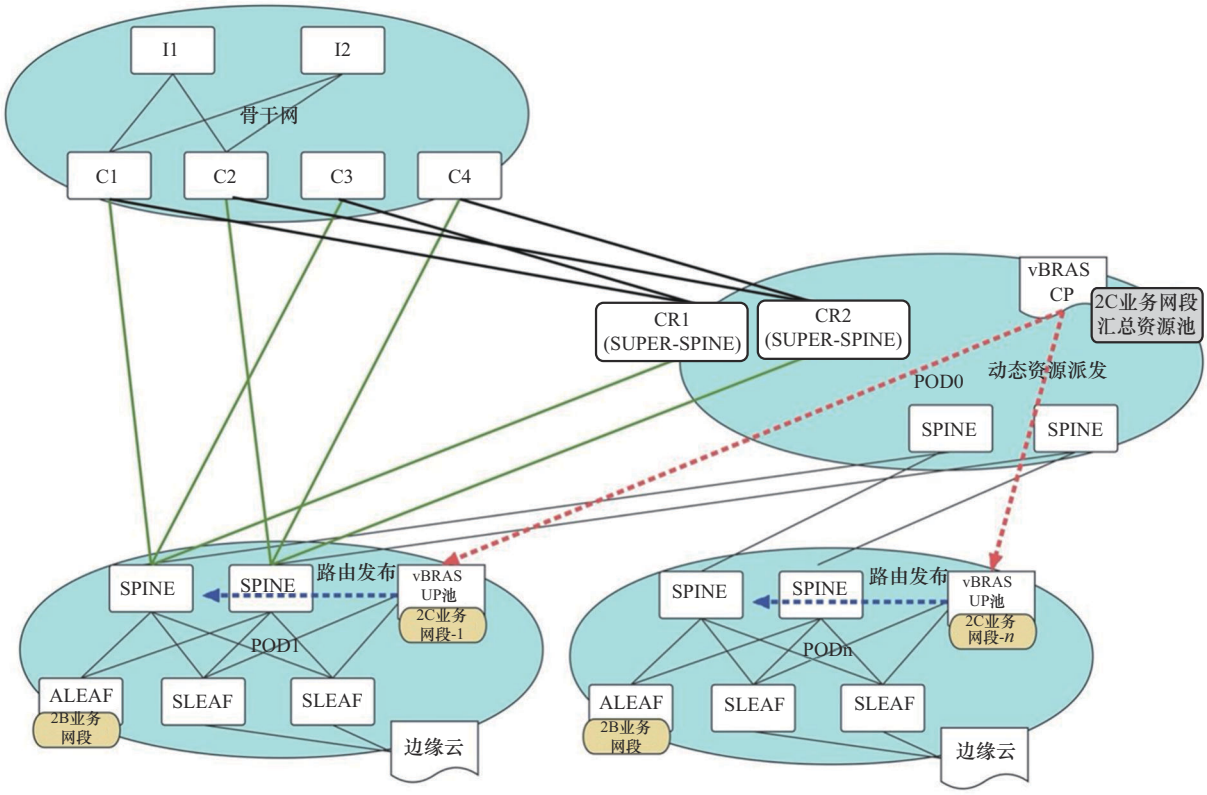

图4　多POD分布式出口的物理连接

间三角形的、交叉影响的复杂BGP邻接关系。在路由逻辑描述中，涉及路由类型的识别可以使用COMMUNITY等进行标识，也可以使用ip-prefix列表白名单，并不影响策略的效果，因此本文不再赘述。

接下来，逐项说明这三角形各组之间的路由逻辑。

（1）城域网CR与主干网C之间的逻辑

城域网继续保持原有的路由策略，即城域网CR将业务网段汇总为一级汇聚网段和二级汇聚网段，通过eBGP邻接分别输出至主干网的C路由器；主干网侧向城域网输出全球路由表，并将接收到的一级汇聚网段设置为全球有权，二级汇聚网段设置为仅主干网内有权。

（2）新型城域网SPINE与主干网C之间的逻辑-eBGP

这里引入了一套全新的路由逻辑。

首先，POD内需要规划准备输出给主干网的汇聚业务网段，命名为三级汇聚网段。对于2C的业务网段，设置CP分配给UP的业务IP资源具有较为完整的网段，可以作为新一级的汇聚网段。例如，家庭宽带拨号业务一般为每个用户分配/32一个IPv4地址和/56一段IPv6地址，那么设置CP在收到UP的IP网段请求时，缺省分配/24一段IPv4地址和/56一段IPv6地址，可以最多支持256个用户的互联网访问。对于2B的业务网段，可以选择重新设置汇聚网段，但需要比原有一级汇聚和二级汇聚更为精细；也可以选择不设置汇聚网段（该内容在第2.3节的“Super-spine的角色”进行详细描述）。

基于三级汇聚网段，新型城域网区域POD的SPINE节点作为与主干网建立出口的网元设备，仅向主干网C路由器输出本POD互联网业务的汇聚网段，并且仅接受主干网向POD发布的IPv4/

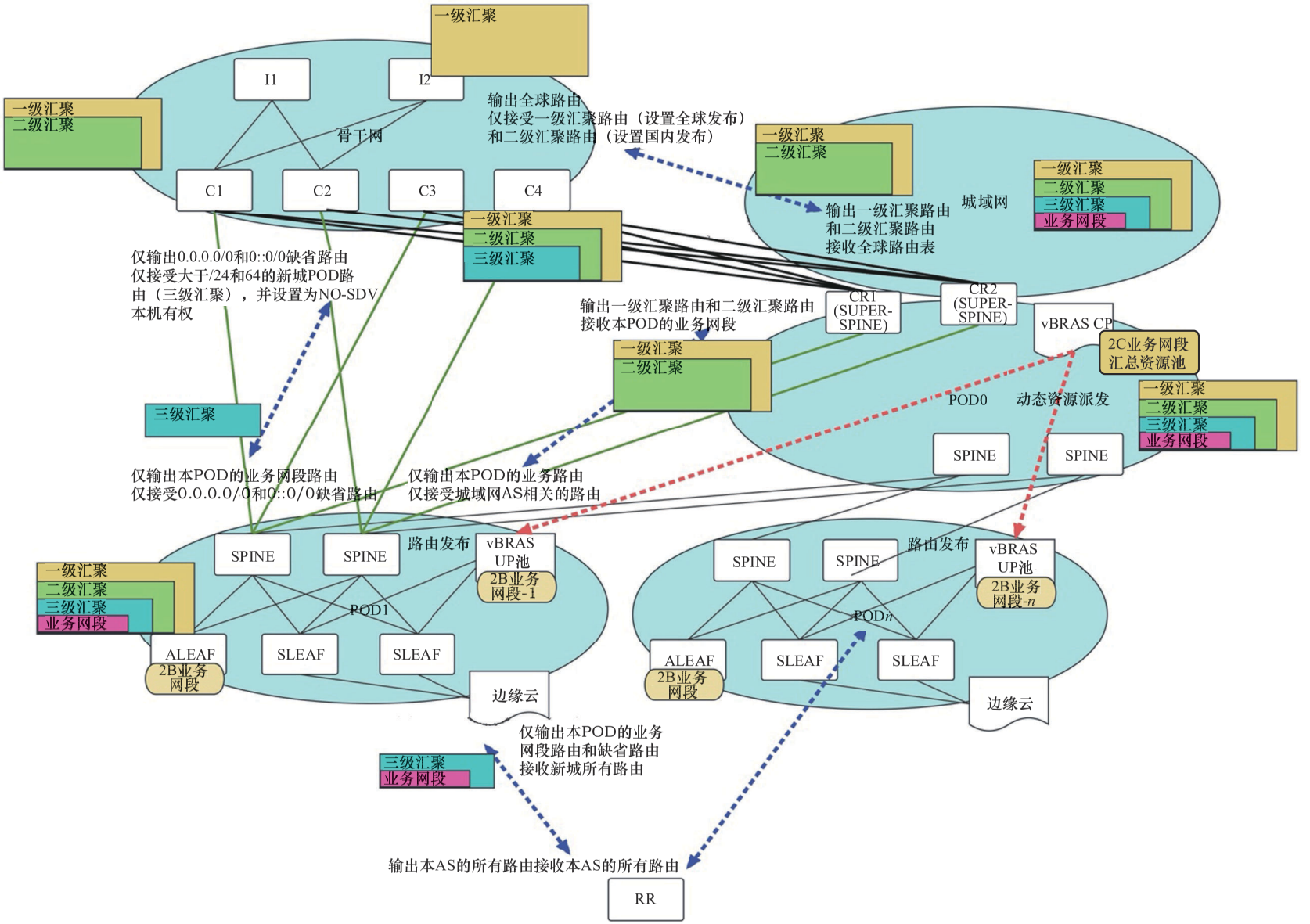

图5　多POD分布式出口的路由逻辑

IPv6 缺省路由。主干网 C 路由器则仅接收新型城域网的三级汇聚网段，仅输出 IPv4/IPv6 缺省路由。为简化主干网侧的逻辑复杂性，主干网 C 路由器不区分具体是哪个区域 POD，将新型城域网各 POD 的出口当作一个整体来配置策略。在主干网内，区别于一级汇聚网段和二级汇聚网段，设置三级汇聚网段，仅在连接 SPINE 的 C 设备本机有权，不再发布到主干网内其他网元设备。

（3）新型城域网 SPINE 与城域网 CR 之间的逻辑-iBGP 或 eBGP

如果城域网与新城城域网采用相同 AS，则区域 POD 的 SPINE 与 CR 之间为 iBGP 邻接关系；如果 AS 不同，则两者之间为 eBGP 关系。无论是 iBGP 还是 eBGP，策略上需要确保新型城域网 SPINE 输出本 POD 的业务网段及三级汇聚网段至城域网 CR，城域网 CR 则输出一级汇聚、二级汇聚网段至 SPINE 即可。

这组策略主要是作为城域网业务逐步向新型城域网迁移过程中的互通通道而使用的。

（4）新型城域网各 POD 之间的逻辑-iBGP

此外，这个三角形的 BGP 关系，也会影响新型城域网各 POD 之间的 iBGP 策略。

新型城域网各区域 POD 和 Super-POD 之间，通过 RR 交互各自的业务网段路由。这里增加的策略是需要过滤各 POD 从直接出口获取的外部路由，以免造成本该通过 POD 的出口疏导的互联网流量，泄漏到其他 POD 的流量异常对穿问题。

对应的设计是在 POD 内 vBRAS-UP、LEAF 与 RR 的 iBGP 邻接上，确保只输出本 POD/本机的业务网段的原有策略。而在开通了至主干网出口的 SPINE 上增补策略，除本 POD/本机的业务网段外，需要将从主干网获取的 IPv4/IPv6 缺省路由也输出至 RR，并过滤其他的所有路由。这组缺省路由是用于引导本 POD 内 vBRAS、LEAF 等其他网元的互联网流量流向 SPINE 节点。新型城域网 RR 保持与区域 POD 的网元原有策略不变，接收并转发各 POD 发布的所有路由；当然也可以考虑增补一个根据新型城域网 AS 号进行保护性过滤的策略，以规避 SPINE 上过滤策略缺失的异常情况。

2.3　实现效果评估

（1）新型城域网的互联网流量引导

比较图 2 与图 6 可以发现，受多出口策略影响的主要是城域网 CR、新型城域网区域 POD 的 SPINE，以及直连新老城域网的主干网 C 设备。

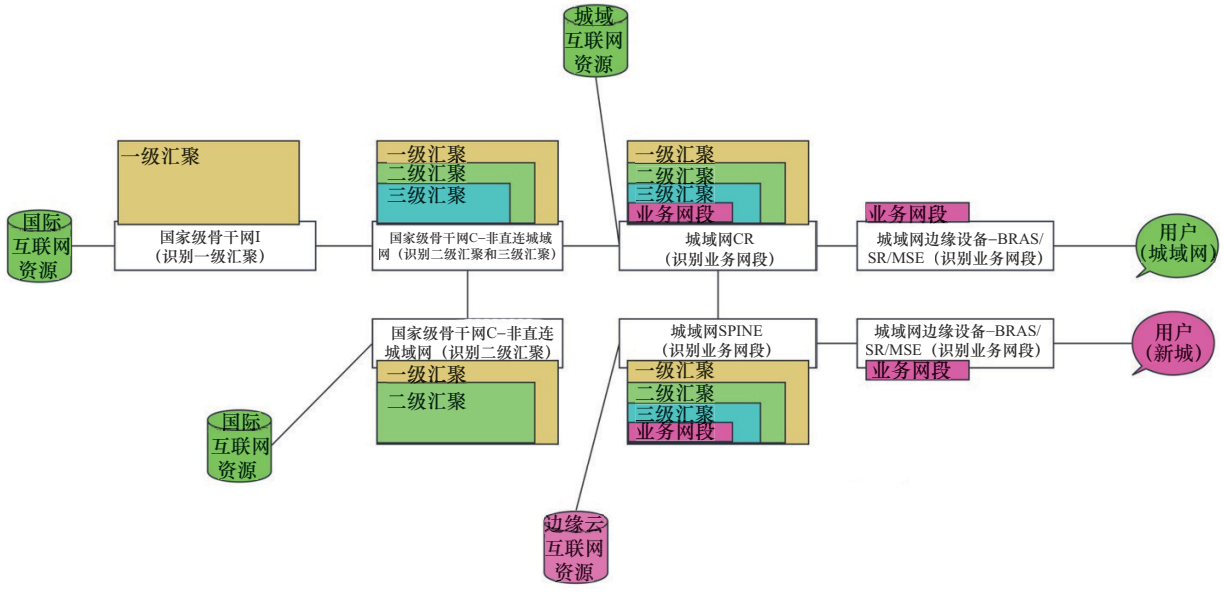

图 6　多 POD 分布式互联网出口部署后业务流量引导过程

主干网侧，国家级主干网的I国际路由器将维持只有该城域网一级/二级汇聚网段路由的现状。与城域网CR和新型城域网SPINE直连的C核心路由器则将拥有该城域网的一级/二级/三级汇聚网段的路由，增加了一段三级汇聚路由；非直连的C核心路由器也维持只有一级/二级汇聚网段路由的现状。因此对国家级主干网的整体影响较小。

城域网侧，CR核心路由器除原有一级/二级汇聚网段和业务网段外，还会增加新型城域网的三级汇聚网段。可以将这组三级汇聚网段沿用新型城域网与城域网交互业务网段的策略，控制其只在城域网内有效。因此，城域网内的路由影响也是较小的。

新型城域网内，区域POD的SPINE作为本POD出口设备，增加了三级汇聚的配置，并接收城域网发布的一级、二级汇聚路由，是整个策略中变动最大的网元。其他网元设备并不涉及多出口策略的配置。

互联网流量引导的过程调整为，国际互联网资源从国外流入后由国家级主干网I设备（识别一级汇聚路由）——国家级主干网C设备（识别二级汇聚路由和三级汇聚路由，其中三级汇聚路由精准定位到新型城域网区域POD）——新型城域网SPINE设备（识别业务网段路由）——新型城域网vBRAS/LEAF设备（识别业务路由）的路径，到达最终消费流量的用户。

（2）缺省路由的逻辑

之前的多POD分布式互联网出口设计主要描述了进入城域网方向的路由和流量引导关系，而城域网输出方向则需要依赖0.0.0.0/0和0:0/0的缺省路由的引导，参见图7的描述。

城域网原有架构下，CR核心路由器从与主干网的eBGP获取全球路由表，以及IPv4/IPv6缺省路由，并将缺省路由通过iBGP发布到城域网BRAS、SR、MSE等其他网元设备。多POD分布式出口架构下，城域网模式保持不变，CR同时将从主干网获取的IPv4/IPv6缺省路由，发布到新型城域网各区域POD的SPINE。

新型城域网内，各区域POD的SPINE从与主干网的eBGP获取IPv4/IPv6缺省路由，并通过新型城域网RR发布到本POD的LEAF、vBRAS-UP等其他网元。为简化RR的配置，设置RR与各区域POD网元的策略一致，因此本POD的SPINE发布的缺省路由，也会发布到其他POD的网元。但根据next-hop的IGP选择最短路径的原则，本POD的网元优选本POD SPINE提供的缺省路由。

另外，在本文设计中，如果城域网和新型城域网是同AS号的iBGP关系，则可以通过调整SPINE与CR之间IGP的metric值确保CR发布到本POD的缺省路由为次优；如果城域网和新型城域网是异AS号的eBGP关系，则可以通过调度MED值确保CR发布到本POD的缺省路由为次优。无论哪种情况，都可以确保其他POD的SPINE发布到本POD的缺省路由为最差选择。

（3）Super-spine的角色

新城城域网多POD分布式互联网出口的主备缺省路由设计，实则也提供了出口的容余保护能力。这里就要对城域网CR向新型城域网Super-spine迁移的角色定位加以说明。

新型城域网各区域POD的政企互联网专线业务承载在ALEAF节点上，这些客户因为域名、邮箱、VPN等互联网应用的关联，需要长期稳定持有IP地址，同时其办公、生产场所又经常面临搬迁而申请携地移机的情况，造成2B业务网段地址较为离散、难以汇聚的情况。

如图8所示，如果区域POD内存在部分2B业务网段无法汇总为三级汇聚的情况，那么从主干网到涉及的政企互联网专线用户的流量，将遵循二级汇聚的路由引导到城域网CR上，再通过CR与区域POD SPINE获取的业务网段路由，引导到最终用户。该用户输出至互联网的流量，则

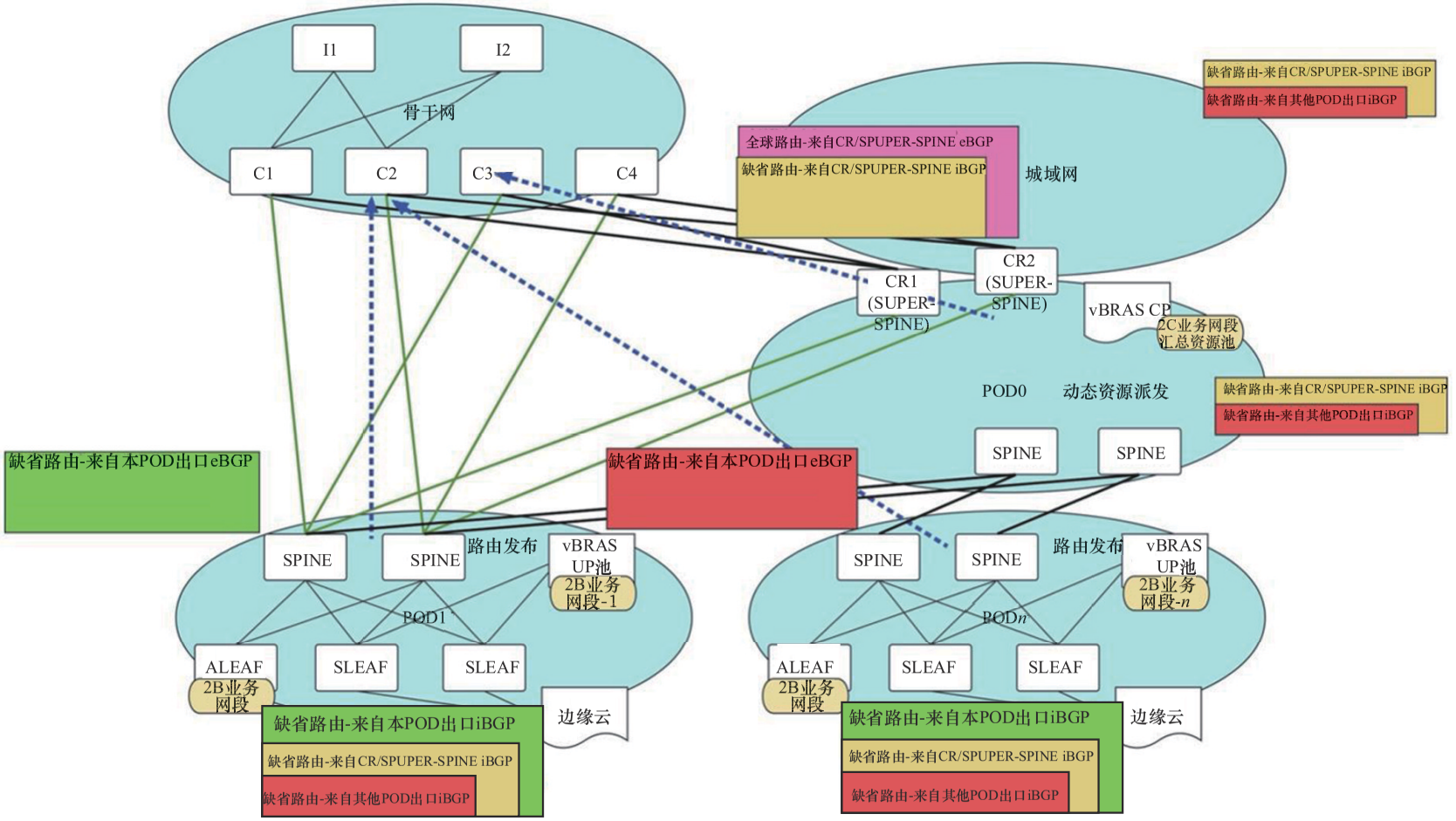

图7　多POD分布式互联网出口中缺省路由的逻辑

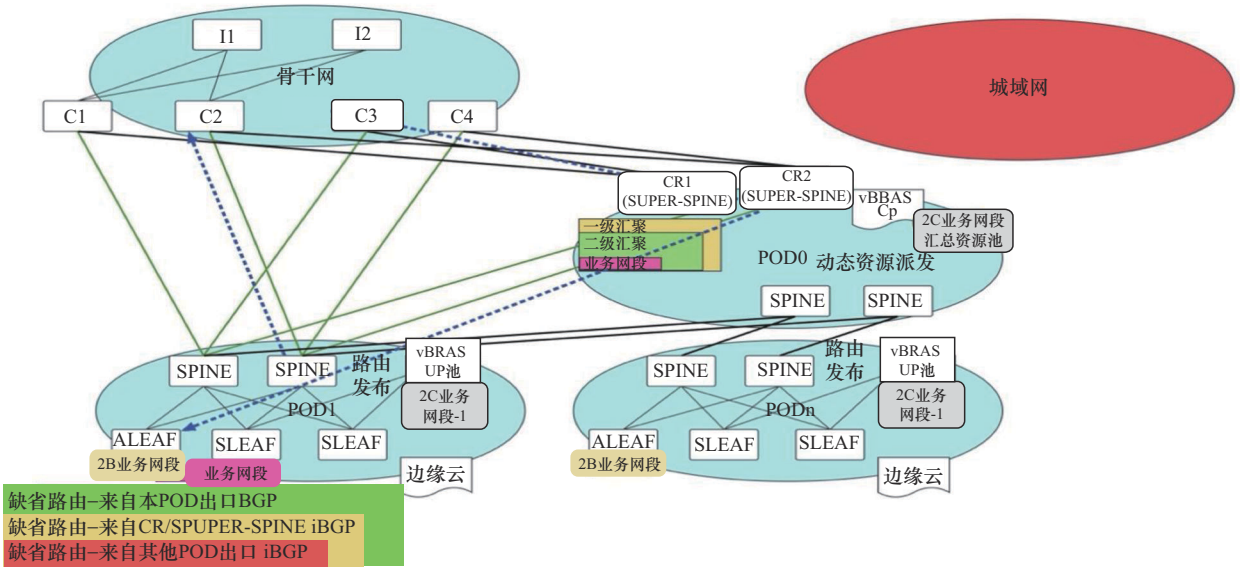

图8　CR向Super-spine的转换

受本POD的SPINE发布的缺省路由影响，通过SPINE直接送达主干网。

因此建议在城域网退网下线后，将城域网CR转换为Super-spine，加入POD0的模块，充当政企互联网专线业务的专用出口设备。

Super-spine还为各区域POD的互联网出口故障提供了一条逃生通道。如图9所示，POD1在连接某一台主干网C设备（C4）的出口发生故障，eBGP邻接关系中断，则该C设备的三级汇聚失效。受一级和二级汇聚的引导，访问POD1内业务网段的流量依旧会到达C4。此时输入流量不能直接到达POD1的SPINE，转而被引导至Super-spine。Super-spine上有POD1发布的业务网段路由，可以将流量引导回POD1的SPINE，到达最终用户，互联网业务不会阻断。

上述场景由于区域POD的互联出口仅仅是局

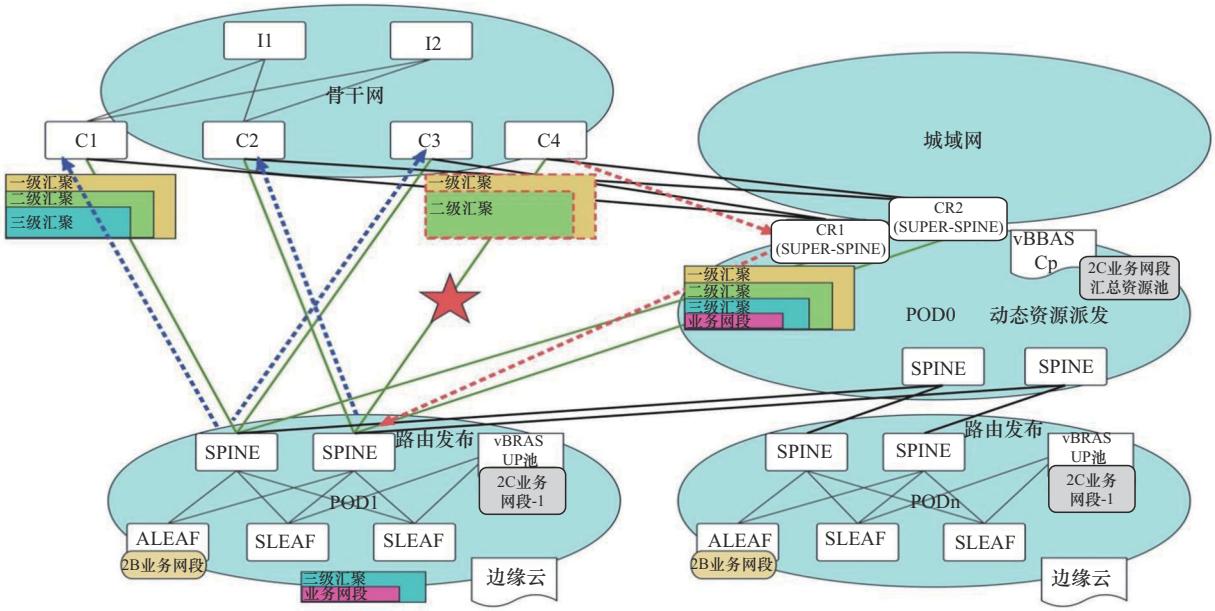

图9　Super-spine提供的逃生通道

部阻断，因此POD输出至互联网的流量，依旧可以从其他出口中继送往主干网。如果在更极端的情况下，由于传输或网络攻击造成区域POD至主干网的所有eBGP邻接关系完全阻断，则除了互联网输入POD的流量会绕转Super-spine，POD输出互联网的流量也会收到次选缺省路由的引导，从Super-spine迂回到主干网。

当然，无论是在哪种冗余保护场景下，都要预先规划Super-spine至主干网出口的带宽。并且其带宽减去实际2B业务流量后的冗余量，不能小于单个区域POD的出口带宽能力。

3　结束语

新型城域网是近三四年才兴起的城域IP组网新概念，具有云网融合、灵活智能的巨大优势。但在南北向流量依旧占据过半信息主体的情况下，其多POD的乐高积木式架构往往让网络设计者难以决策。本文提供了多POD分布式出口的包括框架及路由的组网模型，并根据实践对效果进行了整体评估，希望能抛砖引玉，通过本次研究成果为行业内各地新型城域网的组网实践提供借鉴，近一步完善和推广新型城域网。

参考文献：

[1] 陈运清, 雷波, 解云鹏. 面向云网一体的新型城域网演进探讨[J]. 中兴通讯技术, 2019, 25(2): 8.

[2] 尹远阳, 孙嘉琪, 卢泉, 等. 基于SDN的IPRAN网络智能路由管理系统应用研究[J]. 移动通信, 2016(20): 61-65, 69.

[作者简介]

方鸣（1975– ），男，上海电信网络操作维护中心（NOC）主任工程师、高级IP网络工程师、集团高级专家，主要研究方向为大型IP城域网组织及优化、MPLS VPN技术、IP网络安全技术、IP网络的网管实现。

朱磊（1977– ），男，上海电信网络操作维护中心（NOC）高级工程师、高级专家，主要研究方向为IP承载网技术、宽带接入技术、城域网优化演进。

岳志超（1995– ），女，上海电信智能云网操作维护中心（ICNOC）承载网络技术支撑师、中级网络工程师，主要研究方向为IP核心汇聚网络架构建设及优化、基于SRv6、EVPN技术的灵活多出口政企专线网络与三线BGP用户网络调优与实现。

翁雯倩（1974– ），女，上海电信数字集成部集成业务管理部挂职副主任、一级网络技术（IP）专家，主要研究方向为结合SDN/NFV、SRv6等新技术打造新一代连接双线、IP和传输网管能力向政企客户价值变现等。