



解决 NFV 环境下拨号上网地址 占用问题的 NAT444 方案

颜永明,许斌,徐海峰,叶克闯

(中国电信股份有限公司上海分公司,上海 200085)

摘要:虚拟化技术的兴起,打破了网络设备的传统形态。电信运营商期望在通用硬件架构上部署 NFV,以降低 CAPEX 和 OPEX 成本,提升运营效率,实现快速提供新服务的能力,提升运营商收入。但是,NFV 虚拟化技术有其独有特点,若在政企宽带拨号上网类业务中使用与传统网络设备相同的部署和配置方式,会引发公网 IPv4 地址的大量消耗,提出在 NFV 环境中引入 NAT444 作为解决方案,并对部署、溯源方式进行了研究。

关键词: NFV; PPPoE; NAT444; spine-leaf; 溯源

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016298

NAT444 solution to the dial-up address occupation in NFV environment

YAN Yongming, XU Bin, XU Haifeng, YE Kechuang

Shanghai Branch of China Telecom Co., Ltd., Shanghai 200085, China

Abstract: With the boom of NFV technology, the traditional network hardware was being replaced. For telecom operators, NFV was expected to deploy on the universal server, which could reduce the CAPEX and OPEX, improve operational efficiency, achieve rapid deployment of new service, and ultimately increase the operator's revenue. However, NFV had its unique characteristics, which would consume a large amount of public IPv4 addresses if we used the same deployment and configuration as traditional network hardware in the scenarios of enterprise PPPoE dial-up network. NAT444 was introduced in NFV environment as a solution, including the deployment and traceability system.

Key words: NFV, PPPoE, NAT444, spine-leaf, traceability

1 引言

随着软件定义网络(SDN)相关技术的不断发展,网络功能虚拟化(network function virtualization, NFV)备受关注。NFV 使用 x86 平台通用硬件结合虚拟化技术,用于实现诸如交换机、路由器、防火墙、负载均衡器等传统网络设

备功能。使得网络设备功能不再依赖于传统的专用硬件,在降低网络设备成本的同时,可根据实际业务需求实现自动部署、业务链、弹性伸缩、故障隔离等能力。

为充分发挥 NFV 低成本、部署灵活等方面的优势,电信运营商正着手将 NFV 技术应用于上网专线和宽带拨号业务场景下的用户网关设备,以实现 vCPE (virtual

customer premise equipment, 虚拟用户驻地网络设备)功能, 满足上网用户三层终结、路由转发、NAT(network address translation)等需求。同时, 结合前沿的 spine-leaf 网络架构, 可进一步将 vCPE 用户网关功能上移至电信局端, 在通用 x86 服务器上运用虚拟化技术, 集中部署多个 NFV 用户网关, 充分发挥一体化部署、运营优势, 进一步减少在用户侧的投资和运营成本。但是, 与传统专用硬件网关设备不同, 在局端通用服务器上部署的 NFV 网关设备具有不关机、不下线的特点, 这将导致通过 PPPoE 拨号获得的公网 IPv4 地址被长期占用, 抵消了原有拨号技术的优势, 造成大量公网 IPv4 地址被消耗。因此, 急需找到一种可行的方案加以解决。此外, 在解决网关 NFV 虚拟化所引发问题的同时, 仍需满足国家相关管理部门对于网络信息安全的要求, 即对网络访问行为必须做到有迹可寻, 能够溯源到发起网络访问的组织/个人。本文针对上述难点展开研究, 提出引入 NAT444 技术以解决此类问题, 并给出具体方案。

2 研究对象

考虑到家庭宽带拨号用户数量庞大, 家庭网关需同时满足上网、IPTV、语音等诸多业务, 且面向用户侧物理接口类型、数量多样, 涉及 PPPoE、IPoE 等多种协议, 桥接、路由混合应用场景复杂。受当前 NFV 功能和性能限制, 现阶段在电信局端运用 NFV 技术, 为家庭宽带拨号用户大规模集中部署 vCPE 尚不成熟。因此, 本文选择将研究对象聚焦于业务场景较为单一、用户总量较小、业务需求较为迫切的政企宽带拨号上网客户。

3 传统政企宽带拨号的实现

传统政企宽带拨号上网通常采用 PPPoE 拨号技术。与固定 IP 宽带接入方式相比, PPPoE 方式具有高可控性并在性价比上拥有优势。通常, 由用户侧网关设备向运营商局端 MSE 或 BRAS 设备发起 PPPoE 拨号, 使用用户名和密码进行认证成功后, MSE 根据 AAA 系统授权, 为用户 WAN 口动态分配公网 IPv4 地址, 用户内网则使用私有 IPv4 地址, 经过 NAT/PAT 转换后实现 Internet 访问。用户下线后(用户主动挂断、异常死机、网络中断等), MSE 释放 IP 地址等相关资源。

NFV 技术出现之前, 企业用户大多使用传统的硬件路由设备作为内网的三层网关和 PPPoE 拨号发起设备, 在需要

时可以主动下线或关闭网关电源, 此时释放的公网 IPv4 地址可被再次分配给其他拨号用户使用。由于拨号上网产品整体消耗的公网 IPv4 地址数量更少, 计费方式灵活多样, 因而与使用固定 IP 地址的专线上网业务相比, 在资费上极具吸引力, 深受政企中小用户的欢迎。

4 NFV 环境下的政企 PPPoE 宽带拨号

4.1 业务场景需求

随着大众创业、万众创新国家战略的推进, 针对创客园区/办公楼这类租户较多、资费敏感、带宽需求较小的上网需求日益旺盛, 电信相关部门提出了结合业界前沿的 spine-leaf 云网络组网架构以满足新一代 NFV 环境下政企宽带拨号上网业务的需求, 用户可通过连接电信机房云资源池中的 NFV 三层网关发起 PPPoE 拨号后访问互联网应用。而在电信云资源池网络中则采用了 spine-leaf 架构, 如图 1 所示。

其中, 政企用户通过企业/办公楼的用户汇聚交换机或城域接入网 OLT 等设备构建的二层网络, 上联至 spine-leaf 架构中的边缘汇聚交换机(BLeaf)。而内部汇聚交换机(iLeaf)负责接入网络业务控制器、计算服务器、存储单元等云资源池和 NFV 设备。核心交换机(Spine)负责对业务流量进行交互和引导。整个架构采用 VxLAN overlay 技术实现二层网络连通。而多业务路由器(MSE)则作为网络资源的三层出口控制上联城域网。

NFV 环境下, 政企宽带拨号上网用户可直接通过 PON 二层通道连接到 BLeaf 节点, 三层终结在与 iLeaf 节点相连的 NFV 网关上, NFV 网关取代了传统场景下位于用户侧的政企宽带拨号上网网关, 并负责向 MSE 发起 PPPoE 拨号, 通过 AAA 服务器的认证授权后实现 Internet 访问。

4.2 NFV 网关引发的问题及解决思路

(1) 问题 1

位于 x86 通用服务器上的 NFV 网关基于虚拟化技术, 每台 NFV 负责三层终结一家政企用户, 同一台 x86 服务器上运行多台 NFV 网关, 一般不会关机, 与传统用户侧硬件拨号网关相比, NFV 网关会产生不下线, 不释放 PPPoE 拨号所获公网 IPv4 地址的情况, 这将大量消耗运营商的公网 IPv4 地址。为减少 NFV 不下线对公网 IPv4 地址的消耗, 在目前 IPv6 尚未能取代 IPv4 的情况下, 一种可行的方法是进一步加大私有 IPv4 地址的应用, 即在 NFV 网关发起拨号验证成功后, 为其 WAN 口分配私有地址, 减少

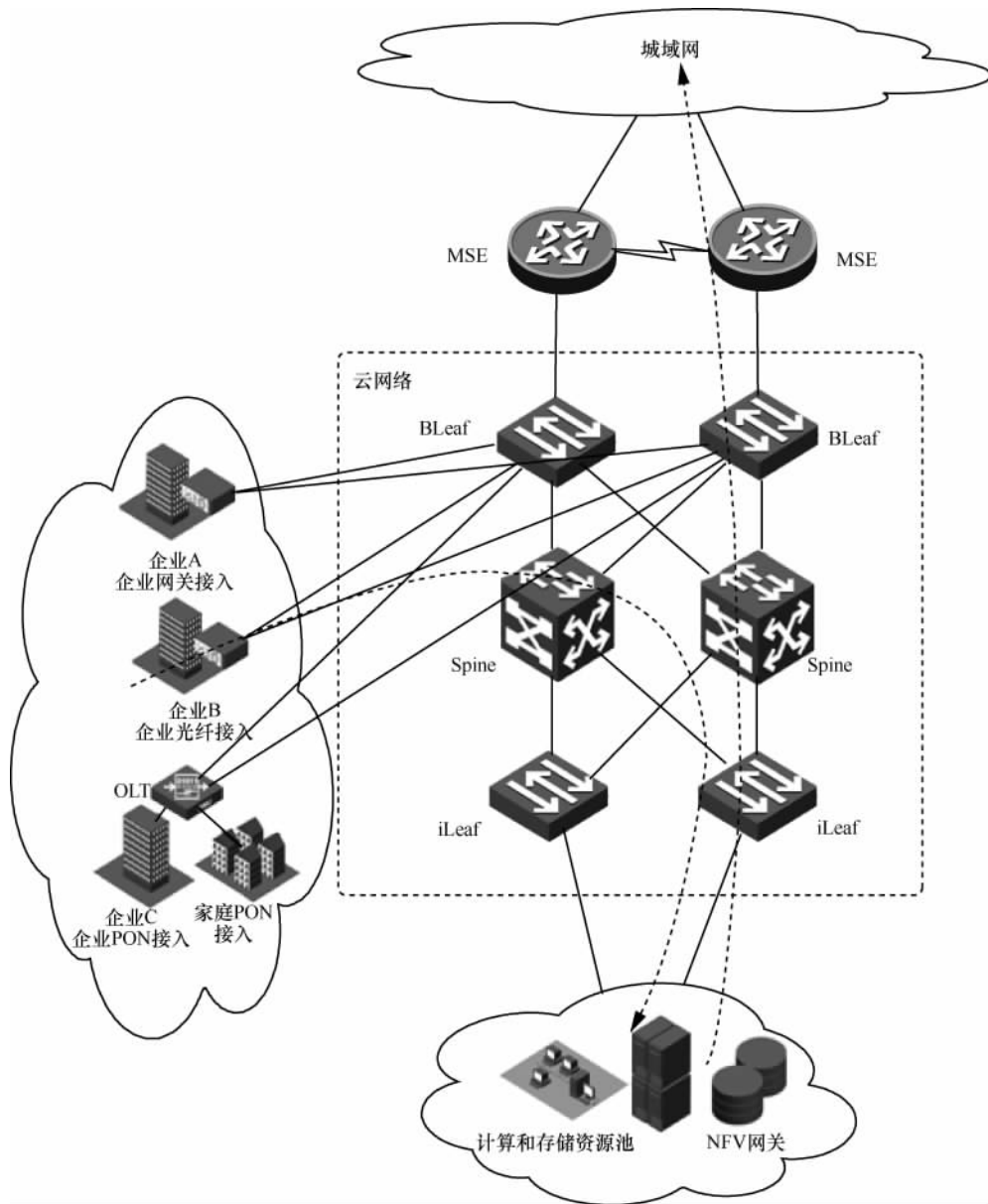


图1 spine-leaf 架构

NFV 拨号网关大规模部署后对公网地址的消耗。

(2)问题 2

业务场景中很多创客团队对于上网资费非常敏感,希望电信运营商能提供与其他用户共享一个公网地址访问 Internet 的业务,从而进一步降低创业成本,但考虑到网络安全和个性化配置,每家政企用户仍希望使用独立的 NFV 网关。可行的解决方法是在 NFV 网关拨号成功后,为其 WAN 口分配私有地址,并在 MSE 上针对多家政企用户 PAT(port address translation)翻译成一个公网 IPv4 地址,但需解决溯源问题。

(3)问题 3

在 spine-leaf 架构下的云组网业务模式中,部分 NFV 网关可交由用户自行配置、运维,这对于用户来说更加灵活,此场景下不应强制用户在其内网统一使用中国电信规划的 IPv4 网段,因此,不能在 NFV 上采用纯路由方式转发数据分组,而应通过在 NFV 上将 LAN 网段私网地址 PAT 翻译成 WAN 口拨号获得的运营商统一规划私网地址,来屏蔽用户内网地址信息,这样既可减少电信对用户内网网段分配的限制,又能确保中国电信 MSE 设备上地址不重复和统一规划。

综合问题 1~3 及其解决思路, 总体方案初现端倪, 引入 NAT444 技术最为合适。通过两次 PAT 可解决 NFV 环境下政企宽带拨号上网的地址问题。结合 PPPoE 账号、NAT444 地址和端口块的绑定, 还可满足国家相关管理部门对于用户访问行为溯源的要求。

5 NFV 环境下政企宽带拨号上网方案

5.1 NAT444 总体方案

由于拨号上网用户产品资费较低, 且 NFV 与传统物理网关不同, NFV 的 PPPoE 拨号如第 3.2 节所述会处于长在线或不下线状态, 使用公网 IPv4 地址作为 NFV 网关的 WAN 口地址会造成公网 IP 地址的大量消耗, 可考虑在 NFV 环境下引入 NAT444 技术。当 NFV 网关拨号成功后, MSE 设备通过 PPPoE 协议为 NFV 网关的 WAN 口分配私网地址。经过 2 次 PAT 实现用户上网, 第 1 步是在 NFV 网关上实现用户私网到电信私网的 PAT 转换, 第 2 步是在 MSE 设备上实现电信私网到电信公网的转换。即先在 NFV 网关上对用户 LAN 网络的私有地址做一次私有地址 PAT, 翻译后的 IPv4 私有地址由电信运营商统一规划, 并通过 PPPoE 拨号分配给用户, 在 MSE 上针对翻译后的私有地址再做一次公有地址 PAT。

5.2 NAT444 的实现

在运营商侧需部署 NAT444 运营级地址转换设备 (CGN)。在 MSE/BRAS 上终结 PPPoE 拨号后将电信统一规划的私网 IP 地址分配给每个用户的 NFV 网关的 WAN 口。NFV 网关用户侧的 LAN 口则可由用户随意配置私网网段。当用户访问互联网时, 由运营商侧部署的 NAT444 运营级地址转换设备 (CGN) PAT 与 NFV 网关的 PAT 组成两级地址转换, 自下而上可形成“用户自定义私有网段、电信统一分配私有地址、公网地址”这 3 部分地址空间。

由于 NAT444 分配给用户的是私网地址, 可能会出现同一时间段内多个用户使用同一个公网地址访问 Internet 的情况, 无法直接通过公网 IP 地址反查到访问网络的组织/个人, 而 NAT444 提供的端口分配方式, 在 CGN 中可采取分配端口块的方式划分用户, 为每个使用私网地址的用户分配一段固定数量的端口号, 能确保利用“公网 IP 地址+端口号”能够反查到用户账号, 从根本上解决用户溯源的问题。另外, 可采用 NAT444 设备支持端口段长度的灵活配置方式, 配合后台支撑系统根据用户拨号账号下发个性化的 NAT444 配置, 满足不同规模客户的多种业

务需求。

CGN 网关可以进行 NAT444 转换的 IP 地址映射工作, 但是其设备特性决定其无法大量记录用户行为。这就要求在后台必须有 Syslog 溯源系统、AAA 服务器、测速系统、行为分析系统等支撑平台协同工作。NAT444 溯源系统如图 2 所示。

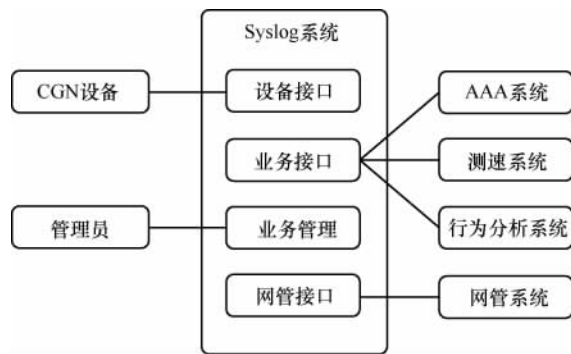


图 2 NAT444 溯源系统

5.3 部署方案

实现 NAT444 功能的 CGN 可以选择在 PPPoE 服务器设备侧通过插卡或者旁挂来实现, 目前上海电信城域网中选择采用了插卡式 CGN, 因此本文也基于这种 CGN 设备形态开展研究。NAT444 功能可以选择部署在城域网的 BRAS 设备侧, 也可以部署在 spine-leaf 网络的 MSE 设备侧, 下文将针对 2 种可选的组网方式进行讨论; 其中方式 1 是利用城域网原有的 NAT444 平台进行组网部署; 方式 2 是在 LSN 网络内新建 CGN 和 AAA 系统。

(1) 方式 1

城域网 BRAS 侧, 共享集中式+CGN 插卡式 NAT444 组网方案, 如图 3 所示。

- 采用 spine-leaf 搭建 underlay 网络资源, 租户连入 spine-leaf 后通过 VxLAN 实现在 overlay 网络上二层连通到局端的 NFV 网关, 局端 NFV 网关直连 BLeaf, BLeaf 则通过光纤与城域网 BRAS 相连。为节省 CGN 板卡投资, 上海电信城域网在 IPv6 改造项目中已采用共享集中方式, 在城域网选择若干台 BRAS 部署完成了 CGN 板卡。因此, 可选择直接利用这几台城域网 BRAS, 为 NFV 环境下的政企宽带拨号业务 PPPoE 拨号和 NAT444 转换提供服务。
- 位于局端的用户 NFV 网关向集中共享式 BRAS 发起 PPPoE 连接, 用户名和密码由已有的城域网 AAA 系统认证。认证通过后, 城域网 BRAS 的 CGN

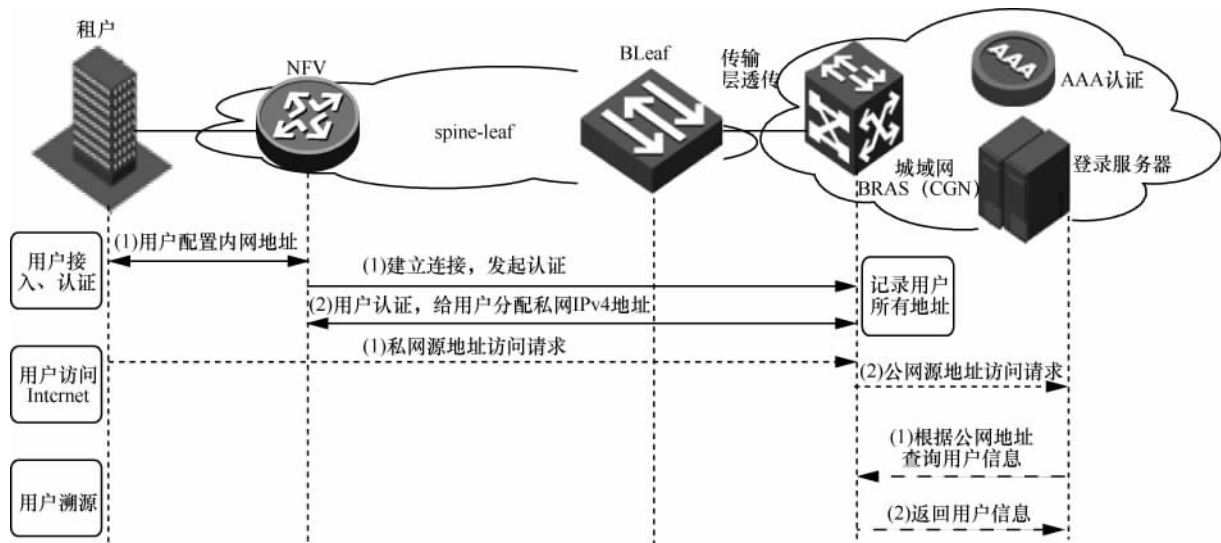


图3 共享集中式+CGN 插卡式 NAT444 网络架构示意

板卡通过 NAT444 功能向 NFW 分配私网地址,并将系统生成的用户账号与地址、端口块的映射关系发送给统一的 Syslog 服务器保存。

- 为解决用户溯源和访问记录问题,可采取分配端口块的方式划分用户,建议为每个使用私网地址的用户分配连续的 2 048 个端口号。系统 NAT444 转换前是私网地址,转换后是共享公网地址和端口块,并可通过查询公网地址的端口块行为日志信息来溯源用户上网行为。

(2)方式 2

spine-leaf 网络中自建分布式+CGN 插卡式 NAT444 组

网方案,如图 4 所示。

- 采用 spine-leaf 搭建 underlay 网络资源,租户连入 spine-leaf 后通过 VxLAN 实现在 overlay 网络上二层连接到局端的 NFW 网关,局端 NFW 网关直连 BLeaf, BLeaf 则上联至 spine-leaf 连接的本地 MSE, NFW 网关向本地 MSE 发起 PPPoE 拨号,由分布式 AAA 系统(可利用本地云平台虚拟环境构建)进行认证。
- AAA 认证通过后,本地 MSE 的 CGN 板卡通过 NAT444 功能向 NFW 分配私网地址,并将系统生成的用户地址、端口块的映射关系发送给远端的 Syslog 服务器保存。

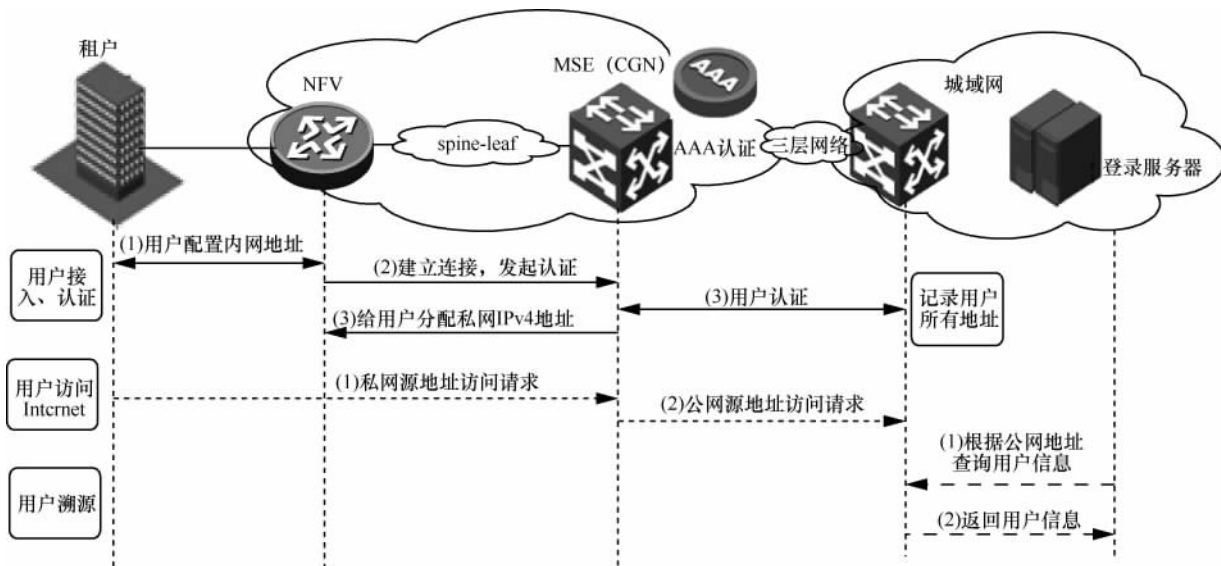


图4 自建分布式+CGN 插卡式 NAT444 网络架构示意

- 针对用户溯源和访问记录问题,与方式 1 相同,将系统生成的用户账号与地址、端口块的映射关系发送给全网统一的 Syslog 服务器保存。

两种组网方式的对比见表 1。

共享集中式+插卡式 CGN 方式,无需新增设备,可以直接利用城域网已有的 NAT444 平台和 AAA 系统,设备投资较小,但是 BLeaf 需通过直拉光纤连至城域网 BRAS,应采用多光纤冗余上联,否则一旦光纤出现问题,用户的 PPPoE 拨号将受到影响,无法上网。自建分布式+插卡式 CGN 方式,需要在每个节点新增设备,投资较大,但是不会改变现有城域网结构,每个节点都需要有各自的 CGN 和 AAA 服务器,通过本地 MSE 直接终结 PPPoE 拨号和 NAT444 功能和私网地址,可以保证用户 PPPoE 连接的稳定性。

在规划 spine-leaf 节点布点数量较少的情况下,可选择方式 1。但从长远发展角度来看,应采用方式 2,通过分散部署 CGN,能解决日后城域网 BRAS 产生的瓶颈问题,分散式的 AAA 系统搭建可充分利用 spine-leaf 节点中的云计算资源,以节省投资。

5.4 公网地址端口块的划分

需要指出的是,NAT444 的业务架构在特定情况下可能造成同一公网地址发起大量源端口号访问和会话请求的现象,并集中访问某个公网网站或者应用,这类正常行为会被误认为是 DoS 攻击或端口扫描,进而触发网站/应用的防火墙或者安全策略将此类行为定义为非法访问,并对源公网 IP 地址进行限制甚至阻断,例如,某些创客开发团队会集中连接租用的同一远端云服务器实施开发工作。所以在划分公网地址和端口号段的时候,应充分考虑用户的上网行为和所承载的业务类型,合理分配相关网络资源数量。

考虑到每个公网 IP 地址的用户数以及用户所访问的业务(包括网页、下载、视频、会议等),针对小型政企用户,建议默认情况下可将 NAT444 端口块的长度设为 2 048 个,且采用动态地址端口映射方式,这样可以实现 1:32 地址转换,即 1 个公网 IP 地址可同时供 32 家小型政企客户使用。针对中大型政企用户,则应充分了解用户上网行为习惯和个性化需求,通过标识用户 AAA 系统中的账户字段属性,按照与用户商定的端口块长度,下发用户业务个性化配置。在必要时,还可向用户提供多个公网 IP 地址解决源端口数量不足或触发 DoS 防护的问题。从业务角度来看,可以针对这些对端口块长度和公网 IP 地址的额外需求,设立不同的产品,差异化收费。

6 结束语

随着 SDN/NFV 技术的不断发展,在不久的将来,以 SDN 和 NFV 体系架构和技术引发的网络革命将转化为电信运营商竞争能力和盈利能力的核心,但随之而可能会在实现中产生了一些专用硬件设备环境下不曾遇到过的问题。本文就运营商在 NFV 环境下的政企宽带拨号上网场景进行了研究,提出了在 NFV 环境下利用 NAT444 技术,一方面可解决了 NFV 不下线产生的公网 IPv4 地址紧缺问题;另一方面,满足了国家相关管理部门的要求,实现了对上网用户行为的溯源管理工作。上海电信作为应用 NFV 技术最早的省级运营单位之一,通过自身实践,为类似场景提供了一套可行、可供借鉴的解决方案。

参考文献:

[1] 中国通信标准化协会. 运营级 NAT44 设备技术要求: YD/T 2544-2013 [S/OL]. (2013-04-25)[2016-08-01]. http://www.ptsnet.cn/standard/std_query/show-yd-4139-1.htm.

表 1 共享集中式与自建分布式组网的对比

比较项	共享集中式+插卡式 CGN	自建分布式+插卡式 CGN
CGN 部署位置	集中部署在城域网 BRAS 上	分散部署在各 spine-leaf 节点内的 MSE 上
部署难度	无需新增设备,利用已有城域网平台	需新增 CGN 设备,独立新建 AAA 服务器
对现网影响	插有 CGN 板卡的城域网 BRAS 需连接至各 spine-leaf 网络,且需同时服务于已有公共客户和新增拨号用户,增加了配置复杂度	城域网结构不变,由本地 MSE 负责终结 NAT444 功能,并提供私网地址终结,公网出口
可靠性	CGN 需要维护大量会话,设备 AAA 认证系统拉远,链路中断或遭受攻击会造成业务全阻	单节点 CGN 需要维护的会话较少,单点故障影响范围小,AAA 本地认证,业务可靠性较高
维护量	较小	较大
投资	设备投资小,传输链路消耗大	设备投资较大



China Communications Standards Association. Draft: the technical specifications for carrier grade NAT (NAT44): YD/T 2544-2013 [S/OL]. (2013-04-25)[2016-08-01]. http://www.ptsn.net.cn/standard/std_query/show-yd-4139-1.htm.

- [2] 宋盈, 马睿. 运营级 NAT44 设备的部署方案研究与测试[J]. 现代电信科技. 2013(3): 71-77.

SONG Y, Ma R. The deployment research and test of carrier grade NAT44 device [J]. Modern Science and Technology of Telecommunications, 2013(3): 71-77.

- [3] 张届新, 傅志仁, 吴志明, 等. VxLAN 在云数据中心组网的应用[J]. 电信科学, 2015, 31(9): 163-169.

ZHANG J X, FU Z R, WU Z M, et al. VxLAN application in cloud datacenter network[J]. Telecommunications Science, 2015, 31(9):163-169.

[作者简介]



颜永明(1978-),男,中国电信股份有限公司上海分公司信息网络部 IDC 浦东运营中心副经理、高级工程师,主要研究方向为数据网络、云组网等。



许斌(1988-),男,中国电信股份有限公司上海分公司信息网络部 IDC 浦西运营中心技术室经理、助理工程师,主要研究方向为数据网络、网络安全等。



徐海峰(1974-),男,中国电信股份有限公司上海分公司信息网络部 IDC 浦东运营中心技术工程师、助理工程师,主要研究方向为数据网络、云组网等。



叶克闯(1991-),男,中国电信股份有限公司上海分公司信息网络部 IDC 浦西运营中心技术工程师、助理工程师,主要研究方向为数据网络、云组网等。