



宽带私网双栈地址溯源的云化实现

张承立, 何肖嵘, 宋蔚芳

(中国电信股份有限公司上海分公司, 上海 200122)

摘 要: 在 IPv4 地址耗竭, 加快推进 IPv6 规模部署的关键时刻, 运营商为保证平滑过渡, 并兼顾后期纯 IPv6 的目标, 采用 IPv4/IPv6 双栈的方案, 同时实施 IPv4 地址的私网化改造, 解决 IP 资源紧缺的问题, 实现纯 IPv6 的全网部署。通过研究私网双栈的溯源模型, 结合私网双栈溯源的电信类业务应用场景, 以及溯源系统在传统固网 AAA 架构基础剥离后的云化改造, 探索固网宽带私网双栈的地址溯源云化实现方法, 为后续固网 AAA 整体云改提供参考。

关键词: IPv6; 双栈; AAA; NAT; CGN; RADIUS; 时钟同步

中图分类号: TN925

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021115

Privatize modification of broadband dual IP stack on cloud platform

ZHANG Chengli, HE Xiaorong, SONG Weifang

Shanghai Branch of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: As IPv4 address became exhausted even sooner, China Telecom had been speeding up the IPv4/IPv6 dual IP stack construction before the final large scale application of IPv6. Meanwhile, emphasis of IPv4 address privatize modification was placed on following schedule at this very moment to ensure the smooth transition and public IPv4 address providing. By studying the trace-back model of IPv6 and private IPv4 dual stack IP address in typical telecom business scenarios, the way to separate IP address trace-back model from traditional broadband AAA platform and migrated into cloud deployment were acquired, thus provided a reference for the AAA entirety reconstruction.

Key words: IPv6, dual IP stack, AAA, NAT, CGN, RADIUS, clock synchronization

1 引言

2011 年 4 月, IANA (Internet Assigned Numbers Authority) 宣布 IPv4 地址池剩下最后一个 A 类地址, 全球各大运营商都将互联网 IPv6 的

演进提上日程。IPv4 向 IPv6 的迁移过程中, 由于 IPv6 无法兼容 IPv4, 产生了隧道及双栈两种过渡方案。因现网用户和业务大多承载在 IPv4 网络上, 在 IPv6 终端及应用未大量普及前, 运营商为保证平滑过渡, 并兼顾后期纯 IPv6 的目标,

往往采用 IPv4/IPv6 双栈的方案,同时实施 IPv4 地址的私网化改造,解决 IP 资源紧缺的问题,最终实现纯 IPv6 的全网部署。

鉴于国家网络安全和信息化委员会对现网运维的相关要求,CGN、BRAS、AAA 等网络设备,需要保存 6 个月以上的用户日志,然而私网双栈过渡期间,地址类型将涵盖 IPv4 公网、IPv4 私网、IPv6 地址,并涉及 IP 地址端口映射,地址类型的多样性及兼容模式与各网络设备交互量成倍增长,对 IP 地址溯源的传统物理架构下,各设备大量离散数据的分布式处理、硬盘频繁读/写产生的物理硬件损坏率的提升,以及突发请求下系统性能的灵活扩展,提出了挑战。而溯源云化实现方式,较好地解决了这些问题。

2 宽带私网双栈 IP 溯源模型

IP 地址双协议栈技术指在一台设备上同时启用 IPv4 协议栈和 IPv6 协议栈,这台设备既能和

IPv4 网络通信,又能和 IPv6 网络通信。固网宽带私网双栈,作为 IPv6 过渡期的运营商实施方案,关键是引入了 NAT444,通过两次 IPv4 的地址网络地址转换(network address translation, NAT)映射,将用户侧的 IPv4 私网地址对应到电信侧的 IPv4 公网地址端口范围,大幅提升了 IPv4 公网地址复用率,延长了 IPv4 的生命周期,同时也为 IPv6 的最终部署争取了宝贵的时间。

IPv4 两次 NAT 映射分别发生在电信末端的接入设备家庭网关及上层多业务边缘路由器(multi-service edge router, MSE),如图 1 所示。通过 NAT 映射,首先将一个公网地址 65 536 个端口按每段 4 096 个分割为 16 段端口范围,单个用户宽带认证通过后,最终分配到可使用的 IPv4 公网资源为一个公网地址加 4 096 个端口,这一做法实现了 IPv4 公网地址 1:16 的复用,相当于直接将运营商现有的 IPv4 资源扩容了近 16 倍。但不同的用户可同时使用相同的 IPv4 公网地址,加上

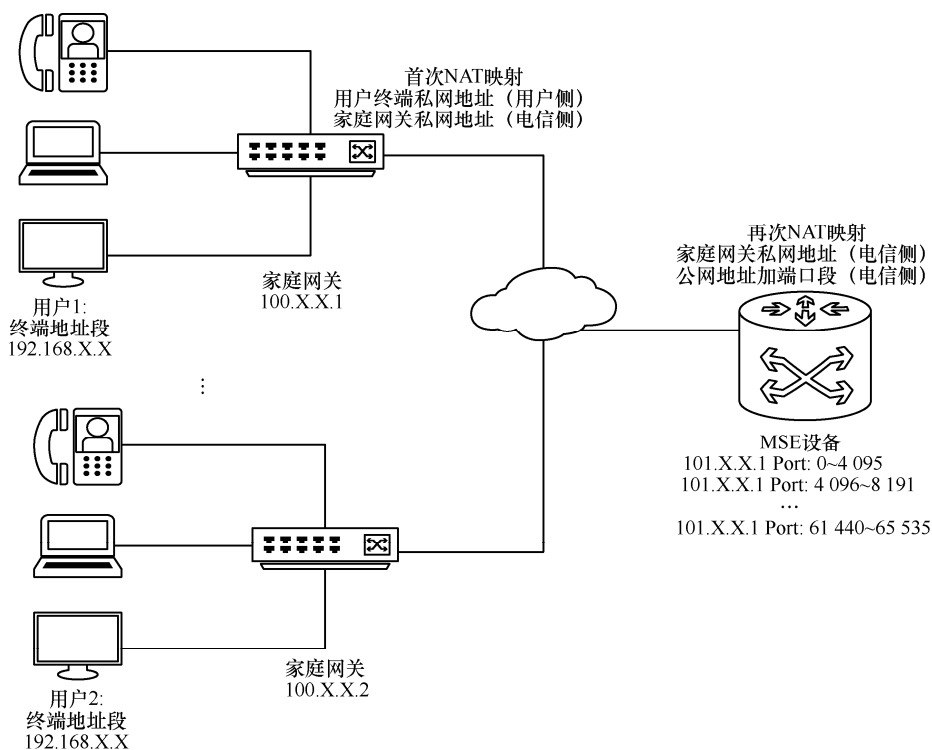


图1 IPv4 私网化的两次 NAT 映射



IPv6 地址, 大幅增加了溯源的复杂度。

鉴于同一 IPv4 公网地址可分配不同用户, 以及不同 MSE 设备可以统一配置模板下发相同的电信侧私网地址, 原有 IPv4 公网地址+时间戳的溯源将无法唯一定位上网账号, 新 IPv6+IPv4 私网双栈的溯源模型如图 2 所示。

客户终端设备 (customer premise equipment, CPE) 在上网认证、授权和计费 (authentication authorization and accounting, AAA) 时, 对于私网双栈的场景, MSE 会下发电信侧的 IPv4 私网地址为 100.X.X.X 的地址及 240E 开头的 IPv6 地址。

对于溯源系统而言, 需要采集处理如下信息。

- 远程认证拨号用户服务 (remote authentication dial-in user service, RADIUS) 日志, 包含用户私网 IP 地址、接入设备 NASIP。
- 运营商机 NAT (carrier-grade NAT, CGN) 日志, 包含 CGN 标识、私网 IP、映射后公网 IP 地址、分配的编口范围。

匹配步骤如下:

- 通过 RADIUS 日志、NAT 日志的私网 IP、时间戳进行匹配查找;
- 根据 NAT 日志的 CGN 标识与 RADIUS 的 NASIP 是否相同, 判断是否是同一 NAT 域;
- 采集 MSE 设备 CGN 板卡通过 Syslog 推送的 IP 地址映射记录, 从而获得公、私网地

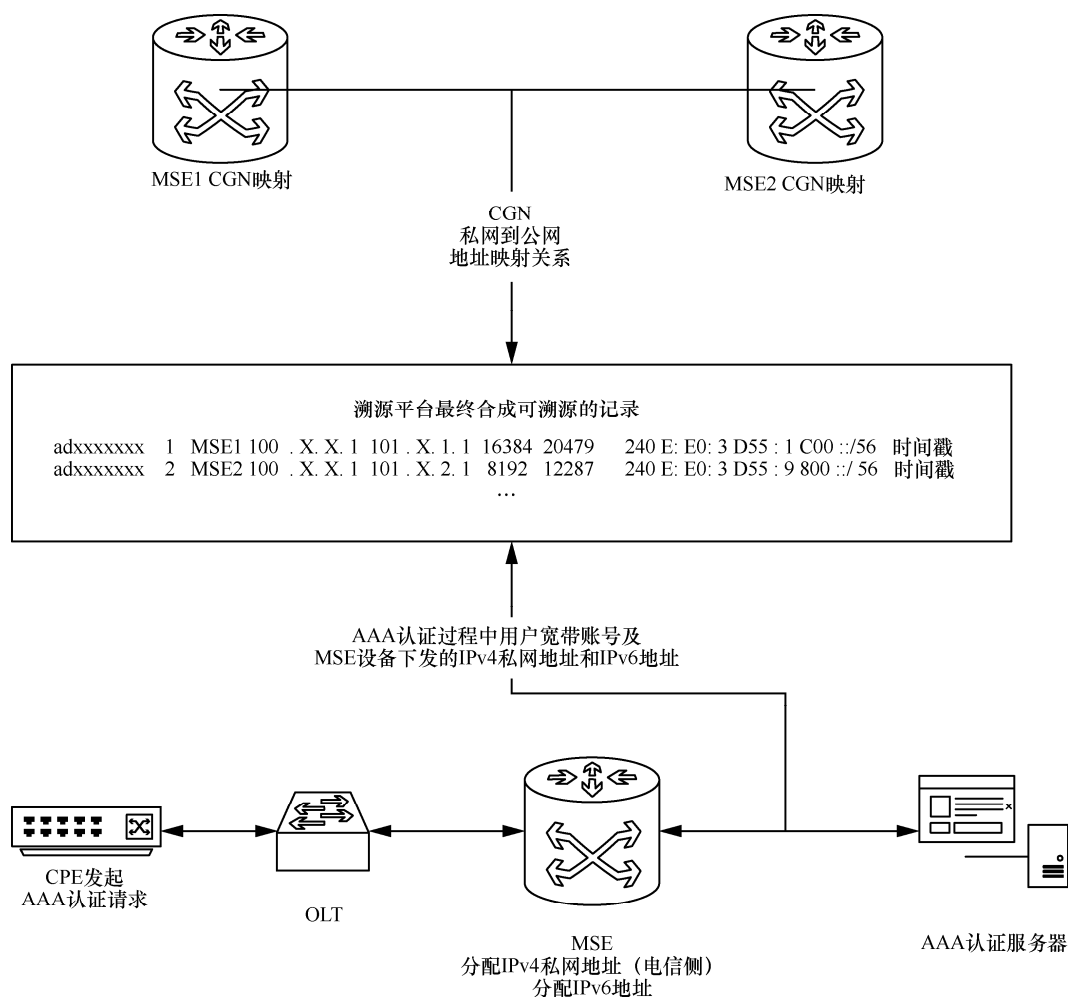


图2 私网双栈的溯源模型

址及端口段的对应关系;

- 将上述信息合成, 最终将宽带账号、上联MSE、IPv4 私网地址、IPv4 公网地址、端口范围、IPv6 地址以及上下线的时间等信息, 合成成为一条记录。

得到合成记录后, 溯源的查证通过 IPv4 公网地址+端口 (或 IPv6 地址)+时间戳, 反向对应到具体上网账号的过程, 如图 3 所示。

事实上, 地址、端口、时间戳 3 个要素便可以溯源到唯一的上网账号。而溯源在时间上的一致性, 通过网络侧设备、平台侧设备与多级时钟源同步完成。

通常, 时间同步网由各节点时钟和传递同步定时信号的同步链路构成。同步网的主要功能是准确地将网络中的各级时钟与基准时钟建立并保持信号同步, 满足通信网整体在时间上的高度一致性。各级时钟源同步确保溯源时间一致如图 4 所示, 图 4 中一级 TS 时钟源, 可由自主运行的铯原子钟组或铯原子钟与卫星定位系统 (GPS 和/或 GLONASS 及其他定位系统) 组成, 作为同步网的全网时钟基准。二级 TS 时钟源可作为区域时钟的基准, 通过向一级 TS 时钟源同步或可同时接收 GPS 等作同步校准。三级 TS 时钟源则可在二级时

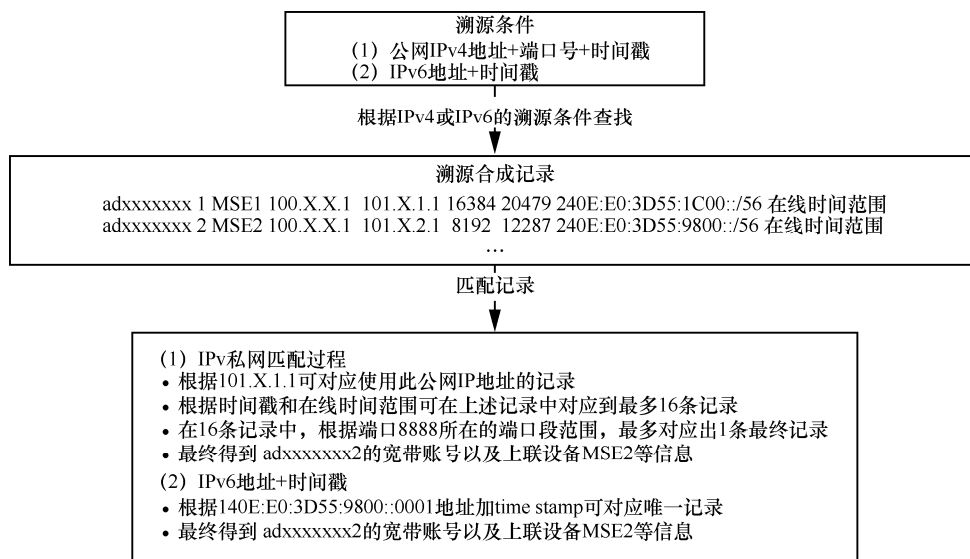


图3 私网双栈溯源过程

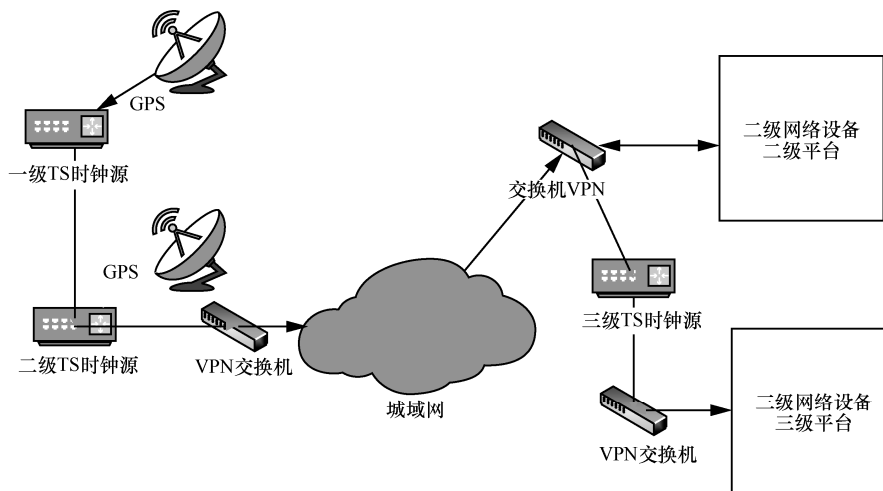


图4 各级时钟源同步确保溯源时间一致



钟源的基础上扩展,以满足区域内的设备及相关业务的时间同步需求。

本文中的溯源平台及相关网络设备,与三级TS时钟源同步完成用户上网请求时RADIUS报文的时间戳、MSE上公转私NAT Syslog报文时间戳、溯源平台的本地时间戳等保持一致。溯源合成记录中所取时间是相关报文时间戳的值,这个值由时钟同步的网络设备插入,与用户侧的终端设备时间是否准确并没有关系。可确保从不同的用户设备、不同的区域、不同的网元线路上网,只要提供正确的IP地址、上网端口及时间戳3个要素便可以精准地溯源唯一上网账号。

值得注意的是,由于网络传输存在一定的时延,不同的网络设备厂商对IPv4及IPv6报文处理方式和顺序不同,现网的运行过程中,IPv4、

IPv6、CGN NAT映射的请求报文在实际到达溯源平台时会存在乱序的情况,便需以时间同步为基础,通过缓存进行二次关联,尽可能满足在报文乱序情况,溯源记录的完整性。例如基于Spark的内存时延关联、二次关联合成技术。当采集的NAT日志找不到匹配的RADIUS上下线信息时,支持临时写入缓存队列,延迟一个可配置的时间段(比如30s内)再次加载匹配。尽可能降低因设备差异、路由差异、采集时延等导致的公转私溯源记录匹配失败场景。Spark内存延迟匹配技术示意图如图5所示。

3 私网双栈溯源的业务应用

为提升用户体验,实现单点登录或自动获取用户账号等信息,运营商在应用侧会通过IP地址信息溯源定位当前用户,如图6所示。

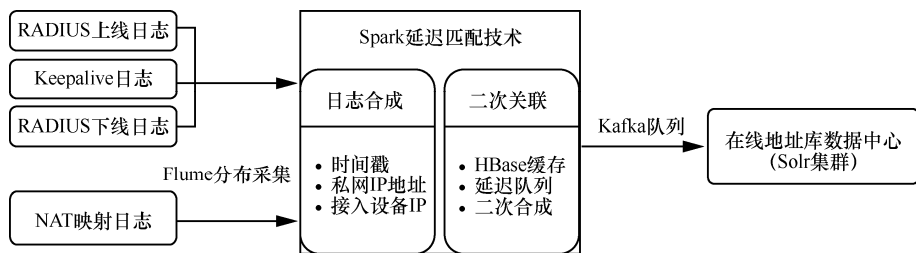


图5 Spark内存延迟匹配技术示意图

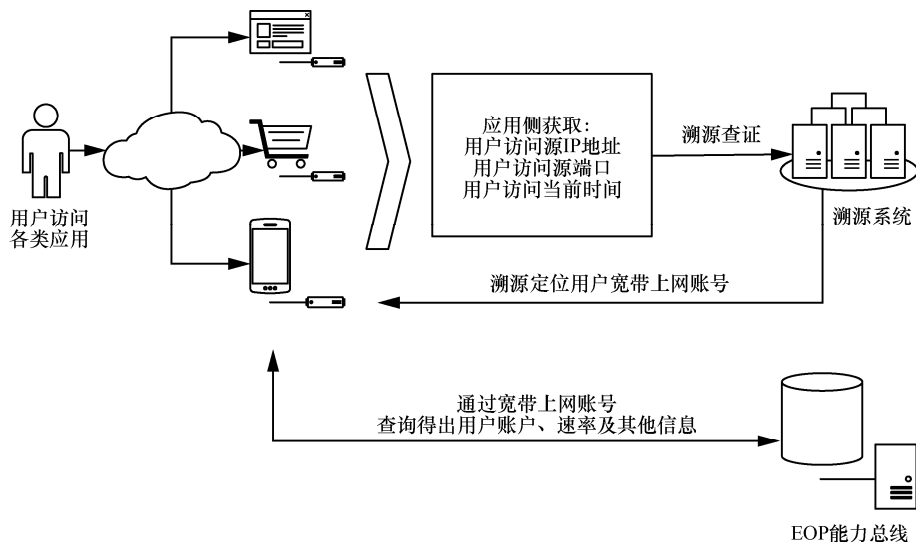


图6 私网双栈溯源的业务应用场景

如 IPTV 用户在订购电信产品时, 通过 IP 地址溯源用户账户信息, 并与机顶盒 SN 所登记的用户账户对比, 实现用户身份的一致性校验, 满足用户无感订购的需求, 同时加强订购过程的安全性。用户通过网上营业厅进行测速时, 无须输入账号, 点击开始测速, 便可以得出该用户签约速率是否测速达标。其通过溯源, 定位用户当前的上网账号, 进而返回该用户在运营商处的签约速率, 与实际测速速率进行比较。

在溯源系统提供了私网双栈的溯源能力后, 应用需确认如何获取当前用户的 IP 地址和端口信息。需要注意的是, 在使用 nginx (一种高性能的 HTTP 和反向代理 Web 服务器)、防火墙 Server NAT 等场景下, 应用侧需保持用户侧的源地址和源端口信息, 否则无法准确溯源到实际用户。

4 私网双栈溯源云化改造

溯源系统理论上可以利用 RADIUS 报文, 在 AAA 认证计费平台的基础上构建紧耦合的溯源功能。考虑运营商的 AAA 平台建设较早, 大多为传统的 IT 架构, 而私网双栈溯源在处理公私网映射、上下线报文入库、中间计费报文更新、IPv4 和 IPv6 上网记录的合成、海量的历史溯源记录存储及对外提供的溯源大数据查询等能力, 都更贴近云的特性。

在 IPv4、IPv6 利用私网双栈平滑过渡的同时, 可考虑溯源系统的云化改造, 充分利用云存储和云计算的优势, 为固网 AAA 整体云化提供实践基础。

4.1 RADIUS 镜像报文采集

为保持原有 AAA 架构不变, 且不影响现网业务的正常使用, 可采用镜像报文采集的方式, 如图 7 所示, 获取 RADIUS 的所有认证计费报文, 并由 MSE 的 CGN 板卡发送 NAT 映射 Syslog 报文给溯源平台, 以便获取溯源所需的必要信息。固网 AAA 镜像流量采集如图 7 所示。

4.2 溯源记录合成

采集数据处理示意图如图 8 所示, 使用 Flume 分布式数据采集服务框架, 可以通过分布式部署、平滑扩展支持多节点、多数据源的数据采集, 充分利用现网 AAA 平台的冗余架构, 保证数据的完整性, 并能对采集的数据进行指定预处理 (例如按条件过滤或格式规整), 适合于 RADIUS 镜像流式数据以及 CGN 映射数据的处理。通过分布式计算能力, 建立溯源关键数据的索引, 提供快速查询的能力; 利用云存储完成历史溯源数据的海量存储, 满足运营商的溯源及统计分析要求。

4.3 溯源查询接口

当溯源记录合成后, 根据用户当前是否在

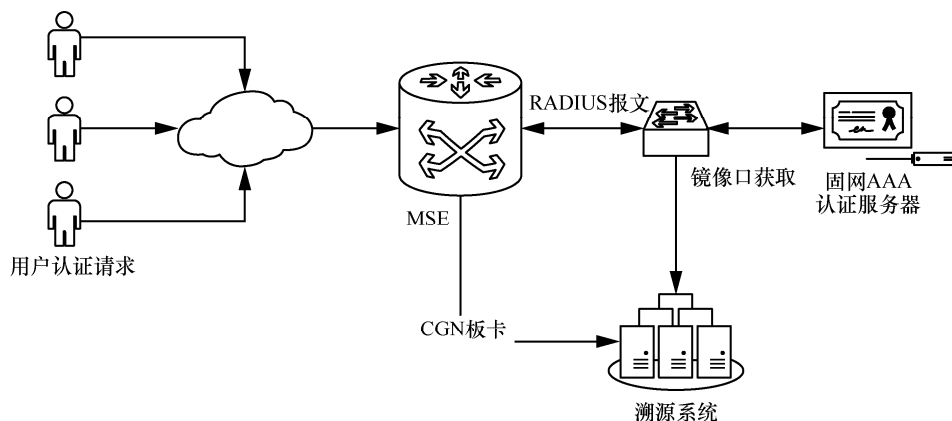


图 7 固网 AAA 镜像流量采集

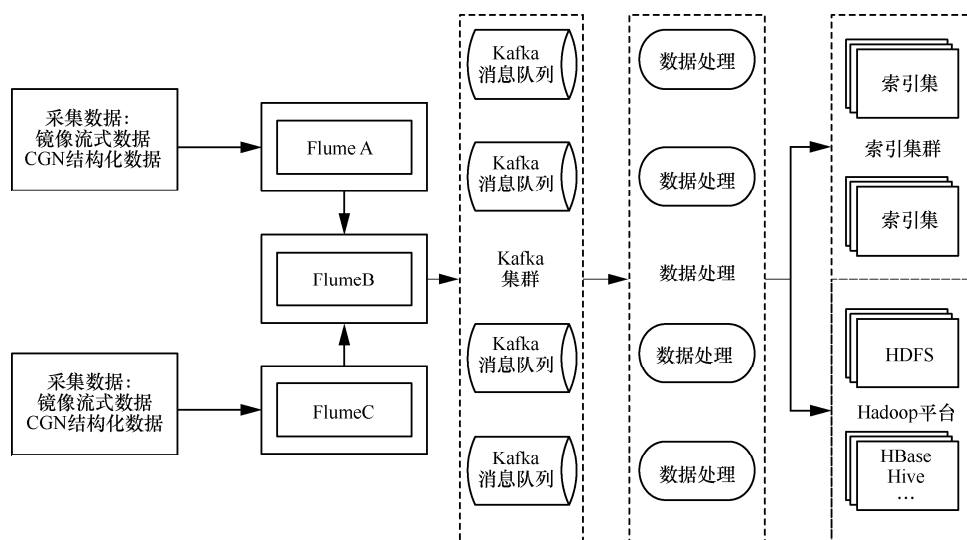


图8 采集数据处理示意图

线（是否收到断线请求报文），可生成在线溯源库及历史溯源库，针对不同的应用场景通过统一的溯源查询中心能力接口，提供外部应用完成溯源查询的请求，并通过能力开放平台（enabler open platform, EOP）等，获取业务所需的账号关联信息，溯源查询接口如图9所示。

溯源查询能力也可注册能力开放平台，从而参与各类电信级业务能力编排，更快速、灵活地支撑业务产品所需的相关功能。

5 溯源云化实现的性能及价值分析

为验证溯源云化实现的性能，模拟云虚拟机单台配置为8核CPU、32GB内存；数据库组成MySQL 2×2集群的条件，400万宽带用户基数溯源云化实现的部分性能指标见表1、表2、表3。

云化后若使用通用的共享虚拟磁阵进行数据库读/写，可能会存在与其他共享同一虚拟磁阵的应用发生I/O争用的问题，且普通硬盘共享磁阵的I/O在测试中只能稳定在5000TPS（transaction

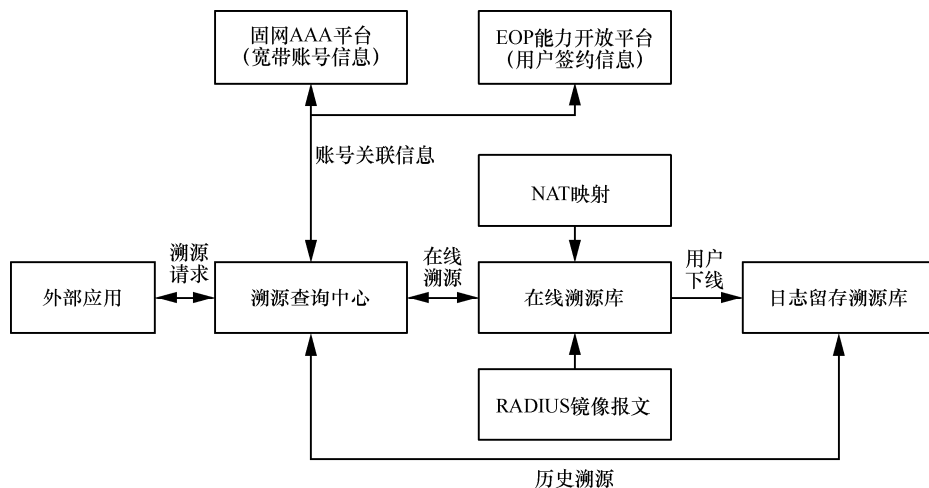


图9 溯源查询接口

表 1 单台虚机 RADIUS 计费采集的性能

场景	处理能力/TPS	响应时间/ms
RADIUS 计费开始报文采集	5 375	1
RADIUS 计费结束报文采集	5 375	1
总计	10 750	1

表 2 单台服务器上 Syslog 日志采集的性能

场景	处理能力/TPS	响应时间/ms
CGN 端口分配日志采集	2 662	1
CGN 端口回收日志采集	2 658	1
总计	5 320	1

表 3 数据库合成溯源记录增、删、改、查的性能

模式	增	查	改	删	组合
2×2 集群	处理能力: 31 072 TPS 响应时间: 13 ms	处理能力: 44 000 TPS 响应时间: 10 ms	处理能力: 33 200 TPS 响应时间: 13 ms	处理能力: 47 639 TPS 响应时间: 7 ms	处理能力: 40 911 TPS 响应时间: 26 ms

表 4 云化虚机与传统物理硬件平台对比

对比项	传统物理硬件	云化虚拟机
硬件成本	独立硬件, 采购及建设成本较高	虚拟共享, 平均成本相对低
软件成本	传统架构, 在处理流式大数据量时没有较大的优势	云分布式架构, 有较多的开源软件为基础实现二次开发成本较低, 且云存储及云计算较适合大量的流式数据处理
处理性能	独享硬件相对处理性能更高、更稳定	共享硬件相对虚拟化及调度等开销, 相对稍低, 但通过架构优化后相差不大
扩展能力	动态扩展能力相对较弱	动态扩展能力强
部署周期	算上采购周期等, 实施周期较长	实施周期短
维护成本	自行独立维护硬件成本高	云中心集中维护硬件成本低
存储成本	相对存储成本高, 存储空间小	云存储更能满足溯源半年以上的大数据量的存储要求

per second) 的读/写能力, 通过优化为 SSD 的硬盘或建立内存数据库可较好地解决这一问题。

云化虚拟机与传统物理硬件平台比对分析见表 4。

6 结束语

私网双栈的现网部署, 切实解决了 IPv4 地址资源耗尽问题, 为互联网应用迁移至 IPv6 提供了无缝过渡的可能。而与传统 AAA 平台解耦后的私网双栈溯源系统, 使现网认证无须改造, 也能支持私网双栈的宽带业务, 并通过溯源系统云化部

署, 在云计算和云存储的支撑下, 提供更好的溯源大数据处理, 能更灵活地满足运营商在运维及网络安全要求下海量溯源数据留存的需求。

参考文献:

- [1] IETF. Remote authentication dial in user service (RADIUS)[EB]. 2020.
- [2] IETF. Network access servers requirements: extended RADIUS practices[EB]. 2020.
- [3] RFC. Address allocation for private internets[R].1996.
- [4] 王晨曦, 古锐, 肖亚群, 等. 基于“IPv6+”的智能 IP 网络方案[J]. 电信科学, 2020, 36(8): 66-80.
WANG C X, GU R, XIAO Y Q, et al. IPv6+ based intelligent

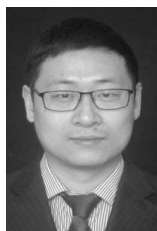


IP network solution[J]. Telecommunications Science, 2020, 36(8): 66-80.

- [5] 何林, 况鹏, 王士诚, 等. 基于“IPv6+”的应用感知网络 (APN6)[J]. 电信科学, 2020, 36(8): 36-42.

HEL, KUANG P, WANG S C, et al. Application-aware IPv6 networking[J]. Telecommunications Science, 2020, 36(8): 36-42.

[作者简介]



张承立 (1981—), 男, 中国电信股份有限公司上海分公司工程师, 主要从事宽带固网 AAA 系统信息化及架构设计等工作。



何肖嵘 (1976—), 男, 中国电信股份有限公司上海分公司工程师, 主要从事云网融合、核心网、光宽端到端等工作。



宋蔚芳 (1979—), 女, 中国电信股份有限公司上海分公司高级工程师, 主要从事云网融合、平台云化、虚拟化技术等工作。