



基于 DPI 不对称流量的同源同宿解决方案

潘洁¹, 高峰², 刘栋², 董昭¹, 侯慧芳¹

(1. 中国移动通信集团设计院有限公司, 北京 100080;

2. 中国移动通信集团采购共享服务中心, 北京 100053)

摘要: 在当前互联网应用分析领域, 为了获得更加准确的基础数据, 流量采集时需要保证每次会话流量的完整性。通过对流量不对称的起因及影响进行分析, 结合业务分析的实际需要, 提出同源同宿流量归并基本思路 and 解决方案。

关键词: 不对称; 同源同宿; 散列算法

中图分类号: TP913

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016306

Homologous homoclinic solutions based on DPI traffic asymmetry

PAN Jie¹, GAO Feng², LIU Dong², DONG Zhao¹, HOU Huifang¹

1. China Mobile Group Design Institute Co., Ltd., Beijing 100080, China

2. China Mobile Procurement Shared Service Center, Beijing 100053, China

Abstract: In the field of current internet application analysis, in order to obtain more accurate basic data, traffic collection needs to ensure the integrity of each session traffic. The cause of traffic asymmetry and the impact analysis, combined with the actual needs of business analysis, homologous homoclinic flow merge the basic ideas and solutions were proposed.

Key words: asymmetry, homologous homoclinic, hash algorithm

1 引言

在互联网网络工程建设中, 出于负载均衡、路由安全保护等考虑, 运营商一般按网状、多出口的结构进行设计和建设, 导致互联网中较多完整端到端会话(session)的上行流、下行流分布在不同的物理链路中, 当不同的链路接入不同的流量采集设备时将会造成会话数据不完整, 因此需要在数据分析之前对非对称流量进行同源同宿处理。同源同宿处理是深度分组解析 (deep packet inspection, DPI) 应

用的重要基础。

2 网络流量不对称原因

目前造成会话不完整的主要原因是网络路由不对称, 这是由网络中路由器的工作特点决定的: 路由器在接收到数据分组时, 会根据路由协议对下一跳路由度量值(cost)进行计算, 并选定一个最优路由进行数据转发, 源、宿两端的路由器在路由选择时可能各自选择一条最优但不重合的路由。最典型的情况是两个相邻路由器间有多条等价的

最优路由, A→B 发送数据时选择第一条路由; B→A 发送数据时选择另一条路由, 从而造成路由不对称。当前大部分城域网采用双出口的结构建设, 两个出口流量负荷尽量均等。网络结构和数据流向如图 1 所示。

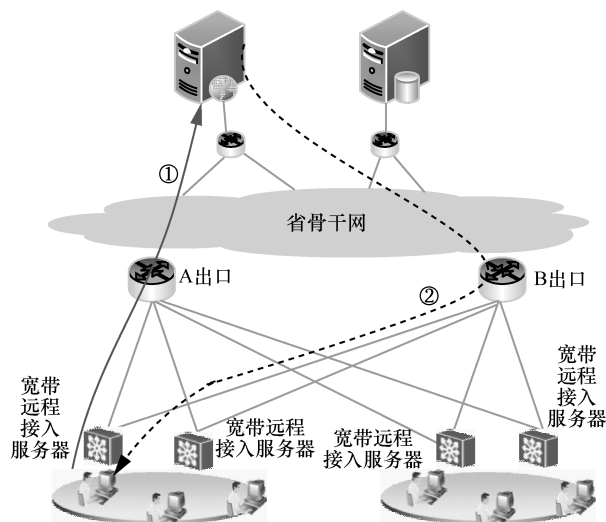


图 1 多节点出现的流量不对称

图 1 以 HTTP Web 访问行为为例。当用户访问远端的 Web 服务器时, 其上行请求数据由 A 路由器进入骨干网(路由如图 1 中左侧实线①), 服务器做出响应时其数据经骨干网另一入口并由 B 路由器返回到用户(路由如图 1 中右侧虚线②)。此时部署在 A 出口的采集设备只能获得上行数据, 部署在 B 出口的采集设备只能获得下行数据, 各自将数据传递给分析系统时二者无法直接关联, 数据分析结果自然不全面、不准确。

对于同一个路由器, 如图 1 中 A 路由器转发端口有 4 个, 通过该路由器的上行数据可能从 1# 端口发出, 下行数据虽然回到 A 路由器但可能从 4# 端口接收, 若采集设备按照端口分开部署, 则各自所采集的数据仍然不全面, 如图 2 所示。

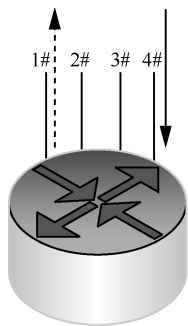


图 2 单节点出现的流量不对称

在网络节点配置路由器的时候可以采用基于数据分组(packet)或者数据流(flow)的负载均衡方式进行数据转发, 当前各运营商主要采用基于数据流的方式进行负载均衡, 而基于分组的方式则更加复杂, 故本文均按照基于数据流的负荷分担部署方式分析。

3 流量不对称影响分析

当存在流量不对称时, 基于数据流的分析应用将受到严重影响。

(1) IDC/ISP 信息安全监控

当监控上行(相对终端用户来说, 访问 IDC 资源的请求为上行)链路的采集系统发现宽带用户访问了违规网站时, 信息安全系统必须在下行链路中对违规关键字内容信息进行封堵, 并告知上行链路对此 URL 进行封堵, 建立访问日志、封堵日志。若流量不对称, 上下行数据流关联不上, 则上行发现违规时无法直接匹配下行封堵, 封堵日志中“访问 URL”字段无法填充。

(2) 僵尸蠕检测控制

当监控上行链路的采集系统发现用户访问了有僵尸蠕代码的网站时, 系统需要在下行链路对代码数据进行扫描、还原形成病毒样本, 若判定为病毒则需要告知上行链路封堵访问 URL。若流量不对称, 则根本无法触发还原和有效封堵。

(3) 流量流向分析

当监控的流量不对称时, 理论上可以对入网流量和出网流量进行单独统计, 然后进行全域的合并归类分析。但这样会给分析系统造成繁重的运算压力, 网络规模较大时, 需要配置庞大的存储计算资源。故一般在采集系统部分就将上下行流量进行关联, 然后报送分析系统。

(4) 网络感知分析

网络感知需要获取一个会话中完整的信息, 以确定 TCP 建立连接的过程中各个环节的时延。当监控的流量不对称时, 只能获取请求报文的信息, 无法获得回应报文信息, 因此 TCP 时延无法计算。

目前部分运营商基于业务需要和成本考虑, 在省网出口部署的 DPI(深度分组解析)一般只监控上行链路, 故无需对下行流量进行检测, 也无需进行同源同宿处理。在此部署环境下, 可开展的业务应用包括用户行为分析(PV 等)、HTTP get 报文获取、信息推送等, 而对流量有关的统计、控制等不太关注。



4 同源同宿的基本原理

为了解决流量不对称造成的影响,就要对链路流量进行同源同宿处理。

同源同宿是基于散列算法来实现的。散列算法又叫散列表(hash table),通过把关键码值(key value)映射到表中一个位置来快速查找。这个映射函数叫做散列函数,存放记录的数组叫做散列表。同源同宿的散列算法中关键码值是每条流的五元组信息(源 IP 地址、源端口、目的 IP 地址、目的端口、协议号),同样的五元组得出的散列算法结果是一样的。当对上行流量、下行流量进行标记,将下行的源 IP 地址、源端口地址、目的 IP 地址、目的端口对调之后,就能获得同样的散列值。在流量汇总输出时,只需将同样散列值的流量通过同一端口输出,即能实现上、下行流的关联归并。在实际的散列值计算时常用除留余数法,对所有散列值用总的输出口数进行取模,得到的余数决定了由哪个端口输出。

举例说明。设置输出组里包含 N 个输出端口,某个会话(以五元组信息来标记一个会话,上下行需要标记、调换)通过散列算法算出散列值为 A ,用 A 对 N 进行取模运算,算出值是 a 。那么这个会话(相同五元组信息的所有数据分组)将会从 N 个端口中的第 a 个输出。

汇聚分流设备在同源同宿处理时将接收的所有流量通过散列算法将数据分组归并、关联后转发到不同的输出端口,达到将同一个会话的数据从同一个输出端口输出的效果。由于散列算法是基于所有流量的运算,所以对同源同宿处理的不同设备、插板必须取定相同的散列算法以保证同步。

同时汇聚分流设备在同源同宿处理时还要同步支持多个相关功能,包括:

- 支持按照五元组的任意组合(如一元、二元、三元、四元)进行散列计算;
- 能调整不同端口的输出量权重,从而调整不同端口的输出流量比例;
- 在某输出端口出现故障(link down)时能够自动将负载分担到其他几个端口;
- 正确匹配并转发 IP 碎片报文,使属于同一数据分组的碎片能发送到同一端口。

不同汇聚分流器厂商的散列算法在上述基础上略有不同。

5 同源同宿解决方案

同源同宿解决的目的是将会话的上下行数据集中在一台设备采集,保证数据采集的关联性和完整性。目前处理同源同宿主要有 4 种环境:单板、单机箱、单机房多机箱/跨机房、异厂商间的同源同宿。本文按照双向并接部署方式进行举例说明。

5.1 单板同源同宿

当链路数量较少、可以接入分流设备同一块单板时,只需在该接口板上配置散列算法,将多条链路中同一散列值的流数据在同一端口输出即可,如图 3 所示。

5.2 单机箱同源同宿

当链路数量较多、需接入分流设备多块单板时,一条会话的上行流数据可能在 1# 接口板接入、下行流数据在 3# 接口板接入,则此时两块单板配置相同的散列算法,通过交换板的调度将不同接口板接入的同一散列值的流数据全部集中在一个端口输出,如图 4 所示。

5.3 单机房多机箱/跨机房同源同宿

5.3.1 单机房多机箱

当监控链路数量继续增加,需多个分流机箱接入时,

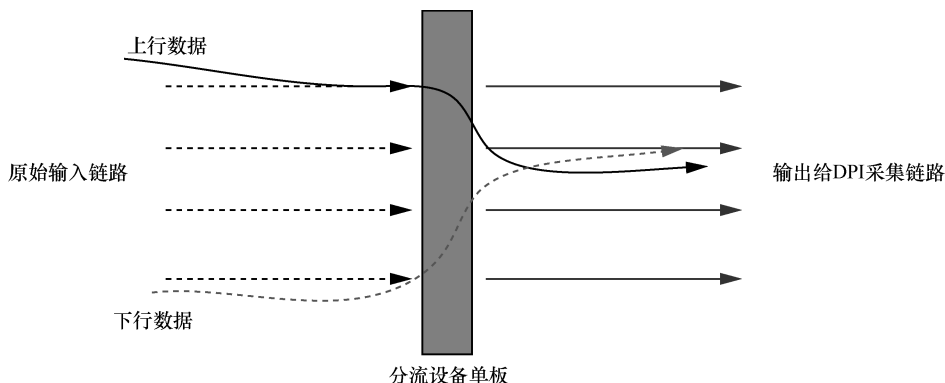


图 3 单板同源同宿解决方案

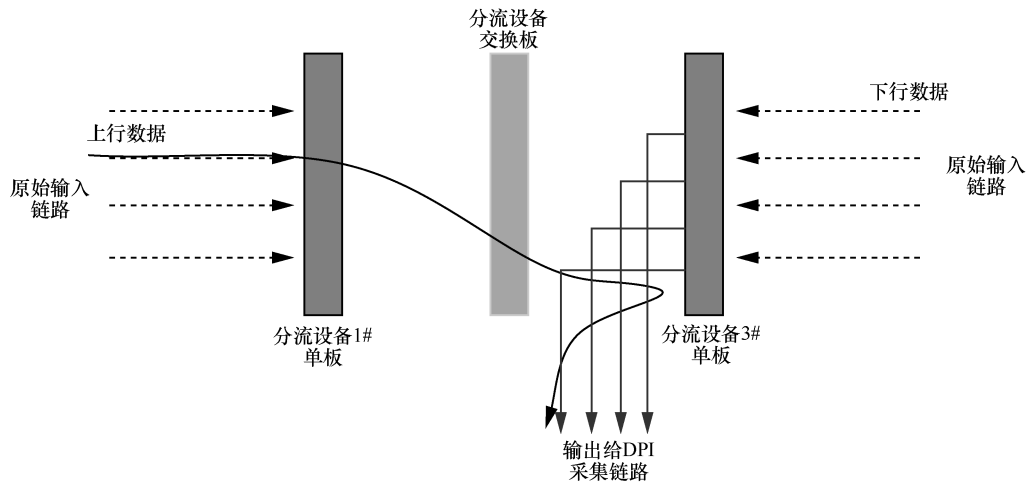


图 4 单机箱同源同宿解决方案

将全部分流设备配置完全一致的散列算法和端口输出方式，将相同散列值的数据输出给同一个采集设备进行处理，DPI 采集设备再做一次同源同宿（类似单板）处理，如图 5 所示。

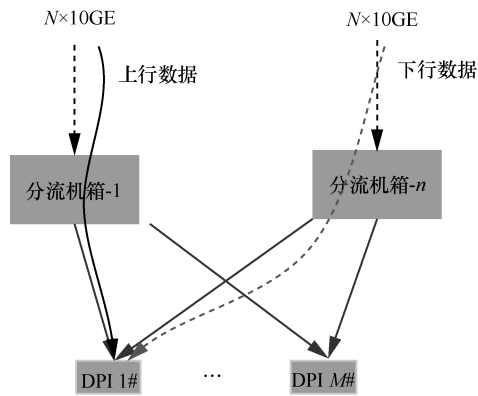


图 5 单机房多机箱同源同宿解决方案(单级)

当单个机房内总流量超大时，采用上述方式将导致 DPI 采集系统端口受限，无法单台设备接入全部分流机箱时，需要采用二级分流模式。第一级分流设备接入链路流量然后采用同样的算法输出到第二级分流设备，第二级分流设备会将第一级同样算法的流量再进行同源同宿输出到 DPI 设备，如图 6 所示。

5.3.2 跨机房

当监控链路分布在不同机房，而又需要全部链路做上下行数据关联时，即需要跨机房的同源同宿。可将两个机房的接入流量完全复制到一个机房后，在单个机房完成同源同宿处理和采集。但往往由于监控链路过长，采用此种方式调度需要消耗大量传输资源。故可根据实际上下行链路情况，将本机房较少的上行流量转接到另一机房，从而保证单个机房均能监控到上行和下行数据，进而进行同源同宿处理。具体方案如图 7 所示。

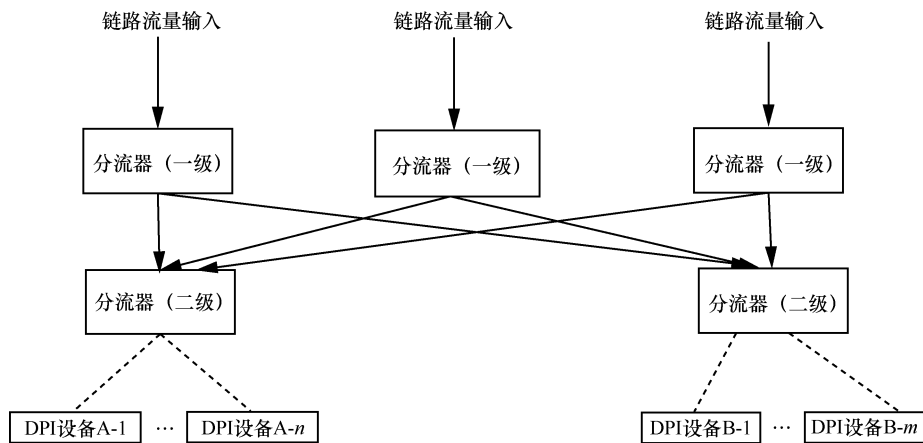


图 6 单机房多机箱同源同宿解决方案(多级)

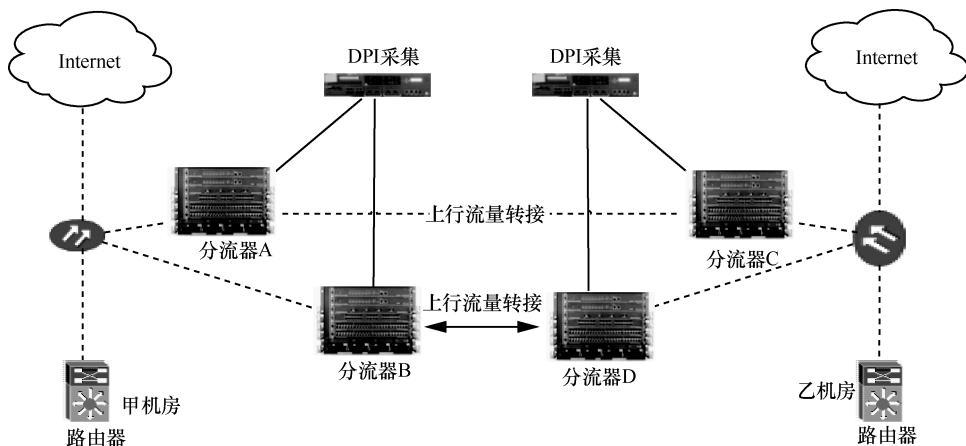


图7 跨机房同源同宿解决方案

5.4 异厂商 DPI 设备同源同宿

对于不同厂商之间的汇聚分流设备处理流量归并,最彻底的解决方案是运维部门将城域网路由器的流量策略归并好,实现上行和下行走同一链路。但这个方案主要取决于现网路由规划和实际运维条件,如果无法改动路由方式,则需要在各个分流器厂商之间打通流量归并的接口来解决。有两种思路可以选择:

- 采用流量散列策略把同源地址段的流量散列到同一厂商的设备里,保证同一用户的上下行流量经由同一厂商设备进行处理;
- 采用分级部署归并方案,在两个厂商的设备之后再部署一级服务器作为流同步服务器进行异厂商信息互通。

5.4.1 散列方式的异厂商设备流量归并

本方案 and 同厂商之间的同源同宿类似,基于源地址段把用户的上行和下行流量汇聚到一台设备上。具体的处理方式是 A 厂商的设备把源地址段 B 的上行流量转发给 B 厂商设备进行处理, B 厂商设备把目的地址段是 A 的下行流量转发给 A 厂商设备进行处理。这样保证源地址段 A 的用户一定在 A 厂商设备上处理、源地址段 B 的用户一定在 B 厂商设备上进行处理,如图 8 所示。

此方案的优点是当两个厂商支持相同的散列策略时,可以把上下行流量汇聚在同一块业务板进行处理,可以实现非常好的协议识别效果和准确的 XDR 话单处理。流量控制和智能镜像功能也都非常完整。

此方案的挑战在于进行流量归并的两个厂商是否能支持相同的散列算法。如果两个厂商不能支持相同的散列策略,就无法实现同一用户的流量归并。这些散列策略包

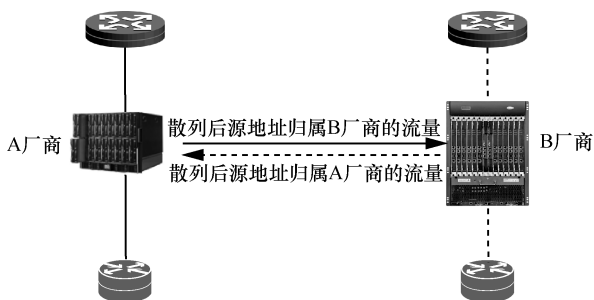


图8 基于散列方式的异厂商同源同宿解决方案

括源地址/目的地址散列方式,散列目标链路的数量和负载分担选择算法。如异厂商之间无法在前面的 3 个参数达成一致就无法实现有效的流量归并。

5.4.2 分级部署方式流量归并

分级式部署方案在异厂商无法按照散列方式进行流量归并时提出了另一种部署方案。通过引入一级流同步服务器来实现两个厂商之间的信息互通,从而提高协议识别的准确率和话单输出准确率。该方案如图 9 所示。

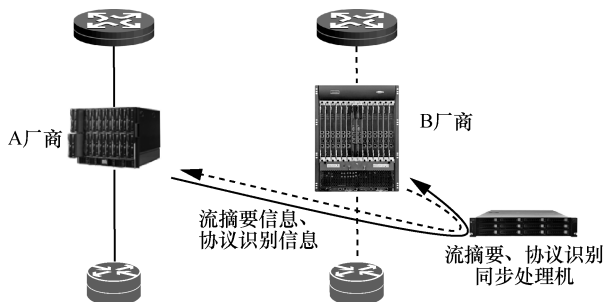


图9 基于分级流同步的异厂商同源同宿解决方案

流同步模块在两个 DPI 厂商设备之间同步自己设备上的每一个流的摘要信息和协议识别信息,从而提高协议

识别准确度和话单。这个同步需要两个厂商按照统一的格式开发消息结构。流同步服务器负责把两个厂商的流信息进行实时和把 DPI 设备进行同步。这涉及两厂商的深度开放,难度更大。

6 结束语

同源同宿处理是深度分组解析 (DPI) 应用的重要基础,处理的好坏将严重影响到协议识别、流量统计、XDR 话单质量,进而影响运营商的各类数据分析应用,因此在相关 DPI 项目中需投入大量精力来解决同源同宿。

本文对多种场景下的同源同宿方案进行深入剖析,力求以更简化的网络结构和更少的成本投入达到更理想的处理效果,为运营商的互联网流量分析系统建设提供参考。

参考文献:

- [1] 谢希仁. 计算机网络 (第 6 版)[M]. 北京: 电子工业出版社, 2013.
XIE X R. Computer network (6th edition)[M]. Beijing: Electronic Industry Press, 2013.
- [2] 张诗超. OptiWay T 技术白皮书[R]. [S.l.:s.n.], 2013.
ZHANG S C. The white paper OptiWay T technology[R]. [S.l.:s.n.], 2013.
- [3] LAWRENCE B. TCP/IP 详解卷 1: 协议[M]. 范建华, 胥光辉, 张涛, 等译. 北京: 机械工业出版社, 2000.
LAWRENCE B. TCP/IP volume 1: agreement [M]. Translated by

FAN J H, XU G H, ZHANG T, et al. Beijing: Mechanical Industry Press, 2000.

[作者简介]



潘洁,女,中国移动通信集团设计院有限公司工程师、高级咨询设计师,主要研究方向为数据网络。



高峰,男,中国移动通信集团采购共享服务中心高级工程师、采购四部副经理,主要研究方向为数据网络。

刘栋,男,中国移动通信集团采购共享服务中心高级工程师、项目经理,主要研究方向为数据网络。

董昭,男,中国移动通信集团设计院有限公司高级工程师、网络所副所长,主要研究方向为数据网络。

侯慧芳,女,中国移动通信集团设计院有限公司工程师、咨询设计师,主要研究方向为数据网络。