网关负载均衡

实践教程

腾讯云

文档目录

实践教程
轻松实现第三方虚拟设备与 GWLB 的适配

最近更新时间：2024-12-13 17:58:52

网关负载均衡（Gateway Load Balancer，GWLB）是运行在网络层的负载均衡，通过 GWLB 可以帮助客户部署、扩展和管理第三方虚拟设备，如防火墙、入侵检测和预防系统、分析、可视性等，操作更简单，安全性更强。本文将为您介绍如何轻松高效的实现第三方虚拟设备与 GWLB 的适配工作。

第三方虚拟设备适配

可支持的第三方虚拟设备

网关负载均衡 GWLB 在网络层处理业务流量，并且与设备的状态无关。此种设定，可以兼容第三方虚拟设备，只要该设备可以支持 GENEVE 封装-解封装和原始数据包。

适配操作

为了与 GWLB 配合使用，第三方虚拟设备需要确保完成以下改造：

- 支持 Geneve 协议与 GWLB 交换流量。Geneve 封装是 GWLB 和设备之间数据包透明路由以及发送额外信息（又称元数据）所必需的。
- 支持编码/解码 GWLB 相关的 Geneve TLV（Type-Length-Value）对。
- 响应来自 GWLB 的 TCP 健康检查。

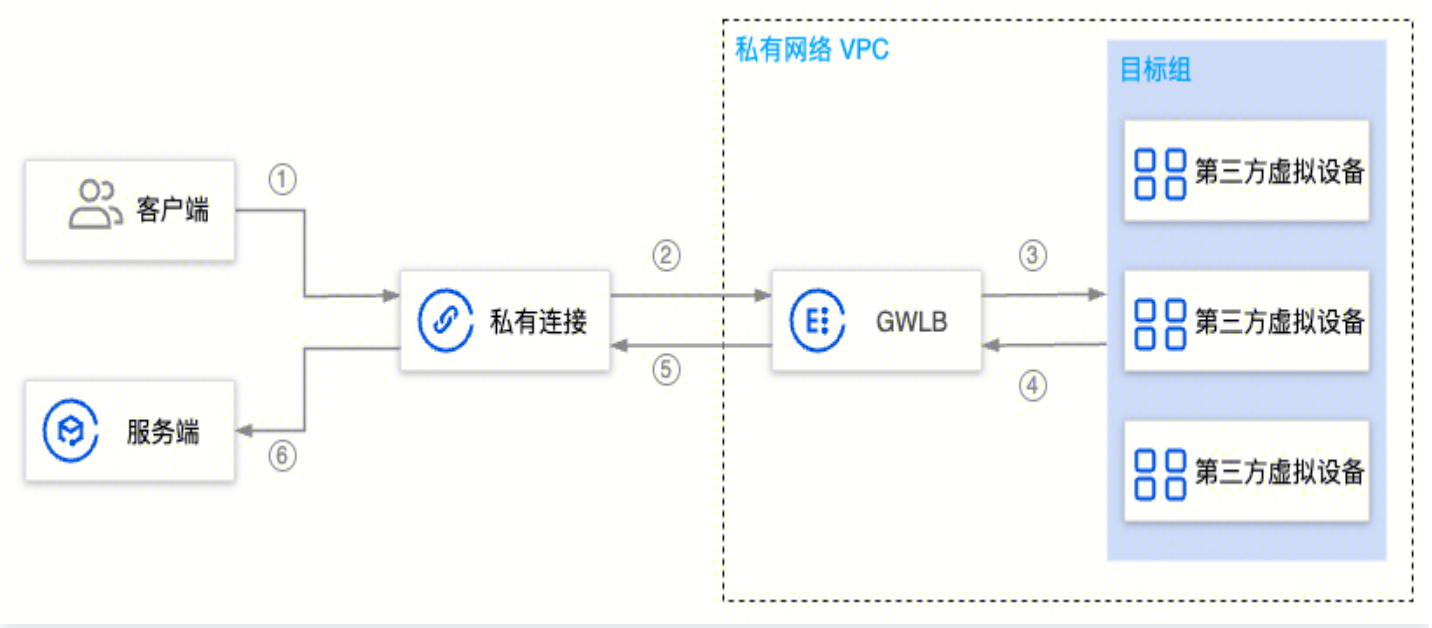
为什么第三方虚拟设备需要支持 GENEVE 封装?

第三方虚拟设备（GWLB 的后端子机）支持 GENEVE 封装是保持原始数据包完整和 GWLB 实现透明转发的基本要求，是 GWLB 提供的一项关键功能。
数据包上的源/目的 IP 与 GWLB 或设备上的 IP 不同，所以，基于这些 IP 的一般 VPC 路由将导致数据包绕过 GWLB 或第三方虚拟设备。因此，需要使用 GENEVE 协议将原始数据包封装到新的 L3 数据包中，这是 GWLB 和第三方虚拟设备之间转发数据包的唯一可行方案。
另外，为了兼容有重叠 CIDR 的多租户设备（此时第三方虚拟设备需要知道流量的来源）。GWLB 可以通过使用类型-长度-值 (TLV) 三元组将每个数据包的额外信息（如 GWLBE/VPCE ID、流 Direct、流 Cookie、附件 ID）发送到设备来跟踪流量并避免用户流量混杂。
GENEVE 协议 (RFC 8926) 非常灵活，允许传递这些额外信息。这种可扩展且、可定制的3层封装机制允许支持广泛的用例并简化客户体验，因为它们不需要对源设备和目标设备进行任何更改。请参阅本文档后面的 GENEVE TLV 格式 。

GWLB的工作原理

技术架构

如下图所示，GWLB 可以连接到另一个 VPC 中的网关负载均衡型私有连接的终端节点上。



GWLB 分为前端和后端。连接到私有连接的一侧称为 GWLB 前端。连接到目标设备的一侧称为 GWLB 后端。在后端，GWLB 充当负载均衡作用，用于将流量路由到多个等效目标设备中的一个。GWLB 确保流向目标设备的双向流量粘性，并且如果所选设备变得不健康，还会重新路由流量。本文重点介绍后端功能 – GWLB 和目标设备之间的通信。

从源发送到目标的数据包中不包含 GWLB IP 作为目标 IP 地址，但由于路由表配置，它们将被路由到 GWLB。为了实现透明转发行为（即保持原始数据包内容不变），GWLB 使用 Geneve 封装原始数据包，并向第三方虚拟设备发送数据包或接收来自第三方虚拟设备的数据包。第三方虚拟设备还需要解封装 Geneve TLV（Type-Length-Value）对。

GWLB 是一种数据包输入/数据包输出服务。它不维护任何应用程序状态，也不执行 TLS/SSL 解密/加密。这些功能由设备本身执行。

当 GWLB 收到新的 TCP/UDP 流时，它会使用三元组弹性哈希（源 IP、目的 IP、传输协议）从目标组中选择一个健康的设备。随后，GWLB 将该流的所有数据包（正向和反向）路由到同一设备（粘性）。对于非 TCP/UDP 流，GWLB 仍使用三元组（源 IP、目的 IP、传输协议）来做出转发决策。

负载均衡调度算法

弹性哈希

GWLB 支持基于源 IP、目的 IP、传输协议的三元组对称哈希算法来做流量调度，会将三元组相同的流量调度到相同的后端服务器中。

负载均衡流粘性

目标组中新增或者删除实例：所有的流量将会全部重新 Hash。

目标组中的某个实例的健康状态从正常变为异常：故障节点的流量将重新 Hash 到其他健康节点。

目标组中的某个实例的健康状态从异常变为正常：故障节点恢复，属于节点的流量重新回到该节点。

健康检查

GWLB 根据用户定义的时间间隔定期运行健康检查。GWLB 通过向设备发送 TCP/PING 数据包来执行这些健康检查。设备需要响应 TCP/PING 数据包，如下所述：

- TCP：建立连接即视为通过。
- PING：若 Ping 成功，且在响应超时时间内，后端服务器未返回 port XX unreachable 的报错信息，则表示服务正常，判定健康检查成功。

第三方虚拟设备必须在 GWLB 超时内完成整个检查。这些检查假设正确响应 TCP/PING 数据包（通常来自其控制平面）的设备也能够通过其数据平面将数据包转发到目的地。

数据转发

步骤 1：当私有连接（GWLBE）从源接收到数据包时，它会使用底层 PrivateLink 技术将数据包发送到 GWLB。数据包停留在 VPC 网络上并到达 GWLB。
步骤 2：GWLB 使用传入数据包的三元组（源 IP、目的 IP、传输协议）并选择特定设备作为目标。此外，GWLB 会生成一个流 Cookie。然后，GWLB 将流条目及其对应的流 Cookie 存储在其流表中。然后，GWLB 使用 Geneve 标头封装原始数据包，并以类型、长度、值三元组（也称为 TLV）的形式嵌入元数据。
步骤 3：GWLB 将封装的数据包转发到特定设备。GWLB 会在该流的生命周期内将该三元组流在流量的两个方向上都转发到该特定设备。
步骤 4：设备必须配置一个可以接受 UDP/IP 数据包的 IP 接口。转发到设备的所有数据包都通过该 IP 接口转发。
步骤 5：设备使用 Geneve 标头封装原始数据包，并嵌入最初为该流接收的相同元数据。
步骤 6：从设备接收到数据包后，GWLB 会移除 Geneve 封装。然后，GWLB 通过将传入（内部）数据包和在 Geneve 中提取的元数据来进行验证、查询、转发。当转发查询失败时，GWLB 将丢弃传入数据包。
步骤 7：数据包使用底层 PrivateLink 技术遍历到 GWLBE。数据包停留在 VPC 网络上并到达 GWLBE，GWLBE 使用路由表下一跳将其传送到目的地。

GWLB 数据格式

下面的数据包格式显示了使用 Geneve 封装的设备接收到的数据包。有关 Geneve 标头的说明，请参阅 Geneve 协议 (RFC 8926)。

```
Outer IPv4 Header:
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version|  IHL  |Type of Service|          Total Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Identification        |Flags|      Fragment Offset    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Time to Live |Protocol=17 UDP|         Header Checksum       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Outer Source IPv4 Address               |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Outer Destination IPv4 Address              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Source Port = xxxx      |     Dest Port = 6081 Geneve     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          UDP length           |          UDP Checksum           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Outer Geneve Header:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=0|Opt Len = 8|O|C|    Rsvd.   | EtherType = 0x0800 or 0x86DD   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Virtual Network Identifier (VNI) = 0       |    Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Outer Geneve Options: Tencent Gateway Load Balancer TLVs
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Option Class = 0x0167 (Tencent)|   Type = 1    |R|R|R| Len = 2 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                      64-bit GWLBE/VPCE ID                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Option Class = 0x0167 (Tencent)|   Type = 2    |R|R|R| Len = 2 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|              64-bit Customer Visible Attachment ID            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Option Class = 0x0167 (Tencent)|   Type = 3    |R|R|R| Len = 1 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  3-bit flow-dir  |                29-bit Flow Cookie           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Customer payload follows…
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|   Customer payload identified by EtherType in Geneve header   |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Source Port：使用三元组 hash 选择 GENEVE 源端口。

GWLBE/VPCE ID：这是分配给 GWLBE 的 64 位终端节点服务字符串 ID（如终端节点服务字符串 ID 为 vpce-12345678，则这里就是12345678，服务供应商使用该 ID 时需要将该字段进行前缀拼接成最终的 ID），设备可以使用此标识符将数据包与其配置的私有连接进行关联。此 GWLBE/VPCE ID 可用于确定流量的源

VPC。每个 GWLBE 只能属于一个 VPC。将 GWLBE 映射到源 VPC 是提供应用服务供应商的管理软件的责任。设备必须“按原样”将此 ID 传回。
Customer Visible Attachment ID：当前没有使用，填充为0。
Flow Dir：标识流方向，RS 必须“按原样”将此 Cookie 传回。
1：外网访问 VPC 网络，该场景下，geneve 协议承载的原始 IP 报文中，源 IP 为外网 IP 地址，目的 IP 为 VPC 网络的 EIP
2：VPC 网络访问外网，该场景下，geneve 协议承载的原始 IP 报文中，源 IP 为 VPC 网络的 EIP，目的 IP 为外网 IP 地址
3：保留
4：VPC 访问 VPC 流量
Flow Cookie：流 Cookie 是 GWLB 在转发时生成的29位流 ID。RS 必须“按原样”将此 Cookie 传回。

第三方虚拟设备响应流程

1. 将原始数据包封装在 Geneve 报头中。
2. 交换外层 IPv4 报头中的源 IP 地址（第三方虚拟设备 IP 地址）和目标 IP 地址（GWLB IP 地址）。
3. 保留原始端口，不交换外层 IPv4 报头中的源端口和目标端口。
4. 更新外层 IPv4 报头中的 IP 校验和。
5. Geneve 报头中所有字段原样返回，不进行修改。
6. Geneve Option 中的所有 TLV 字段（GWLBE ID、Flow Direct、 Flow Cookie、Attachment ID），原样返回，不进行修改。

> ⚠ **注意：**
> 第三方虚拟设备返回给 GWLB 的报文长度不能大于接收报文的长度。

测试验证

当第三方虚拟设备支持 Geneve 协议、GWLB TLV 的编码/解码并响应健康检查，即可进行测试验证。
您可以从单个设备开始作为最简单的测试用例。查看设备是否响应健康检查，在设备上打开数据包捕获来查看数据包流，验证数据包是否为预期格式。

产品高可用说明

最近更新时间：2024-10-21 16:43:01

网关负载均衡 GWLB 的高可用是从系统架构、产品配置等多维度来保障的。您可以根据业务场景和需求，选择跨地域容灾、同地域跨可用区容灾等多种功能方案。

GWLB 集群高可用

网关负载均衡 GWLB 实例采用集群部署，支持消除服务器单点故障，提升系统冗余，保证服务稳定。所有 GWLB 实例均具备集群高可用。
腾讯云网关负载均衡主要基于腾讯自研的 GWLB 网关，GWLB 网关具有可靠性高、扩展性强、性能高、抗攻击能力强等特点，单集群可以处理 Tbps 级别的流量，可以支持每秒数百万级的请求。轻松应对各种流量分发的场景。

单 GWLB 实例高可用

网关负载均衡 GWLB 提供负载均衡服务，SLA 为99.99%。网关负载均衡在多个可用区部署集群，同一个 GWLB 实例，会同时下发到多个可用区。客户端访问该 GWLB 实例时，访问流量会自动选择时延最低的可用区集群，然后转发到后端服务器。
如果某个可用区的 GWLB 集群不可用，网关负载均衡可在非常短的时间内（约30s）自动切换到其他可用区并恢复服务。如果此时仅 GWLB 集群故障则对客户的访问流量无影响；如果此时整个可用区故障（后端服务器也同时故障），则 GWLB 实例通过健康探测得知此后端服务器异常，则不会将流量转发至故障可用区的后端服务器上。

后端服务高可用

网关负载均衡 GWLB 通过健康检查来判断后端服务的可用性，避免后端服务异常影响前端业务，从而提高业务整体可用性。
开启健康检查后，无论后端服务器权重是多少（包括权重为0），都会进行健康检查。您可在目标组详细信息中的**目标组内实例**页签，查看后端服务器的“健康状态”。或者在目标组详情页面的监控页签中查看不健康 RS 个数。关于健康检查的详细机制，请参见 健康检查概述 。