

# 负载均衡设备中最大带宽、最大期望带宽、剩余带宽、当前带宽、繁忙比等概念的理解

四层服务器负载均衡 孔梦龙 2021-09-05 发表

## 问题描述

不区分SLB和LLB，都是一样的，可以参考理解，

## 解决方法

【带宽算法】：带宽算法是链路均衡的一种常用动态算法，综合当前各个链路的当前带宽，各链路能提供的最大带宽与各链路占用带宽的权重，保证被分担的数据流根据当前各物理链路的带宽占用情况进行分发。

【最大带宽算法】：即报文总是分发给当前空闲带宽最大的链路。保证被分担的数据流分发给当前带宽最大的链路。支持按照Inbound方向带宽、Outbound方向带宽或总方向带宽进行带宽统计。此处应注意最大带宽和最大期望带宽的关系。

如果使用的算法是：带宽算法或者是最大带宽算法，则有剩余带宽的概念

剩余带宽 = link的最大带宽 - 当前带宽

1) 【link最大带宽】：rate-limit bandwidth命令用来配置链路所允许的最大带宽，命令分为：链路型、实服务器、虚服务型，可参考命令手册

//缺省情况下，链路的最大带宽不受限制 注：通俗的理解，也就是说人为的干扰这个链路的最大的带宽，假设这个链路的实际的物理口是1000M的，人为的规定使用LB功能的时候，这个链路是800M，为了给不适用LB功能的流量省出点空间。

2) 【当前的带宽】：就是目前真实流量的带宽，本机上跑LB功能的流量的带宽。

缺省情况下，链路的带宽由负载均衡模块自行统计（LB自己涉及）；

也可以配置链路的带宽由接口统计：bandwidth interface statistics enable，RS的当前带宽不再使用LB自行统计的带宽，而是使用RS IP地址所在的接口的带宽作为RS的当前带宽，LB自行统计带宽，是指的LB分发给该link的报文，每秒做一次，然后下一秒分配连接时，就根据前一秒的这个值来确定链路带宽。（通俗得来讲就是本机上跑LB功能的流量的带宽，当然是基于接口的）。

【最大带宽】一般是指在上述的两种算法（带宽算法或者是最大带宽算法）中的特定的说法，在所有的算法中还有一个最大的期望的带宽。最大带宽，沿用上面的例子，这么理解，最大带宽就是告诉流量本链路能跑的带宽的大小是800M，不管设置了什么样的参数，流量多一点都不行，LB这个模块就通不过。

【最大期望带宽】是一个概念、阈值、参考，主要是用在繁忙比上的，比如：链路是1000M，现场设备设置的最大期望的带宽是800M，繁忙比是70。这样的话，实际上设备上跑的LB的流量达到560M的时候，设备就认为这个链路或者服务器繁忙的了，就不再向这个服务器分发新得流量了。

上面两者举一个例子，手机套餐流量，每月60G封顶，30G以后限速，其中60就是最大带宽，30就相当于最大期望带宽和繁忙比的乘积。

那么两者的关系呢？

(1) 数值：在带宽算法和最大带宽算法中，剩余带宽为链路的最大期望带宽（max-bandwidth）与当前带宽的差值。若未配置最大期望带宽，剩余带宽为链路所允许的最大带宽（rate-limit bandwidth）与当前带宽的差值。

(2) 繁忙比：如果未配置最大期望带宽（不受限），则繁忙保护带宽上限受实服务器允许的最大带宽rate-limit bandwidth (real server view)命令控制。

理解了上面的概念以后，再说说普通的【繁忙比】：

(1) 繁忙比，就是一个比例，是一个数字，但是这个数字怎么发挥自己的作用呢？就要和最大期望带宽配置使用了，假设某链路是1000M，最大期望的带宽是800M，繁忙比是70。这样的话，实际上设备上跑的LB的流量达到560M的时候，设备就认为这个链路或者服务器繁忙的了，就不再向这个服务器分发新得流量了。但是已有的会话，继续转发。

(2) 繁忙恢复比：就是当：（LB跑的流量/最大的期望的带宽）<繁忙恢复比，此时LB认为这个链路不繁忙了，就开始重新给他分发流量；所以繁忙比就分为：总繁忙比、总恢复比、总上行繁忙比、总上行繁忙恢复比、总下行繁忙比、总下行繁忙恢复比；越精细越优先。

理解了上面的概念以后，再说说普通的【最大期望带宽】：

(1) 可参考上面的理解，如果未指定inbound和outbound参数，则最大总期望带宽，总带宽等于入带宽与出带宽之和。

