组播笔记二

一、RPF 校验——反向路径转发

当路由器收到组播流量，路由器先对组播流量进行拆包，查看源 IP 地址，再看单播路由表中有无去往组播信源（该流量中的组播信源）的路由，如果没有，则 RPF 校验失败，路由器直接丢包。

如果有去往组播信源的单播路由 ,则看该单播路由的出接口是否与接收组播流量的入接口一致，如果一致，则 RPF 校验通过，如果不一致，则没通过。

任何信源对于某台路由器来说，有且仅有一个 RPF 接口，绝不会有两个 RPF 接口。

如果出现负载均衡，不同的路由对应不同的出接口，则比较出接口的 IP 地址，IP 地址大的接口就为 RPF 接口。

如果负载均衡的路由对应相同的出接口，但是不同的下一跳时，则 IP 地址大的下一跳就为 RPF 接口。

一台路由器上拥有 RPF 接口，但是在接收组播流量时，希望修改 RPF 接口，则可以采用以下方法进行，而不用改单播路由

①最牛方案——写组播路由 ,但这个不是路由 ,因为在组播中没有路由的概念 ,叫组播路由，但其实不叫路由

Ip mroute x.x.x.x y.y.y.y　RPF 接口　上一跳地址(去往信源的上一跳的 IP 地址)

其中 x.x.x.x 为信源 IP 地址，y.y.y.y 为掩码，

②启用 MPBGP，

利用 MPBGP 建立 IPV4 组播邻居 ,传递 IPV4 组播路由 ,与上面提到的 IP mroute 一样的，实质不是路由。

PIM 的 Dense-mode 工作过程:（源树、最短路径树——SPT）

使用在组播流量接收者特别多的环境中，每个信源对应一棵树，对应于 Push-mode。

每台路由器在收到组播流量后，只要 RPF 校验成功，就会转发给自己的所有 PIM 邻居和 IGMP 的组员。每台路由器在转发组播流量的时候，会形成组播表项，这些表项是以(S,G)的形式记录的，其中 S 为信源的 IP 地址，G 为 IGMP 组地址。

例如 :IP 地址为 10.1.1.1 的信源给 IGMP 组 224.1.1.1 发送组播流量时，每台路由器都会形成(10.1.1.1,224.1.1.1)的组播表项，这个表会占用特别多的资源。

因此，当网络中的组播信源越多，则 SPT 树（源树）就越多，表项就越多，资源消耗就越多，路由器的负担就越重。

所以，在实际中，不使有 PIM 的 Dense-mode。

在现实网络中，并不是每个广播域内都有 IGMP 的组员，也就是说，并不是每台最后一跳路由器上都连接着组播流量的接收者。

因此，当发自信源的组播流量到达那些没有 IGMP 组员、又没有下游 PIM 邻居的最后一跳路由器时，该路由器应当告诉自己的上一跳 PIM 邻居路由器，请不要给自己转发组播流量。此时该路由器会发送一个修剪报文给上一跳 PIM 邻居，让其暂时不要给自己转发组播流量。

但上一跳 PIM 邻居路由器仅仅会停止转发组播流量 3 分钟，3 分钟后，又会向该最后一跳路由器转发组播流量，即修剪报文的作用时间仅仅为 3 分钟。

为什么是 3 分钟呢？

因为当路由器转发组播流量后，其会产生组播路由表项，该表项的存在时间为 3 分钟，所以路由器的修剪周期约为 3 分钟，同时也可以这样认为，**不管一台路由器是否转发组播流**

量，其组播路由表都被维持着。

因为，当路由器发出修剪报文后，就不会接收到上游 PIM 邻居转发的组播流量，经过 3 分钟，其刚要删除组播路由表时，上游的 PIM 邻居路由器又会转发组播流量下来。

因此，在 PIM 的 Dense-mode 中，由于不管路由器是否转发组播流量，其都会一直维持着组播表，这样会造成内存资源的消耗，因此，在实际当中，PIM 的 Dense-mode 很少使用。

运行 Dense-mode 的路由器，收到组播流量并完成 RPF 校验后，其能从以下两种接口将组播流量转发出去：

①连接 PIM 邻居的接口

②接收过 IGMPV Report 报文的接口

如果有运行 Dense-mode

如果运行 Dense-mode 的路由器发现没其它的 PIM 邻居，且又没有任何接口接收过 IGMP 的 Report 报文，其就会向上游 PIM 邻居，发送一个 Prune 报文，上游的 PIM 邻居收到这个 Prune 报文，告诉上游邻居暂时不要给自己发组播流量，而上游路由器通过某个接口收到这个 Prune 报文后，并不会立即停止转发流量，而是要做出一番判断，判断：

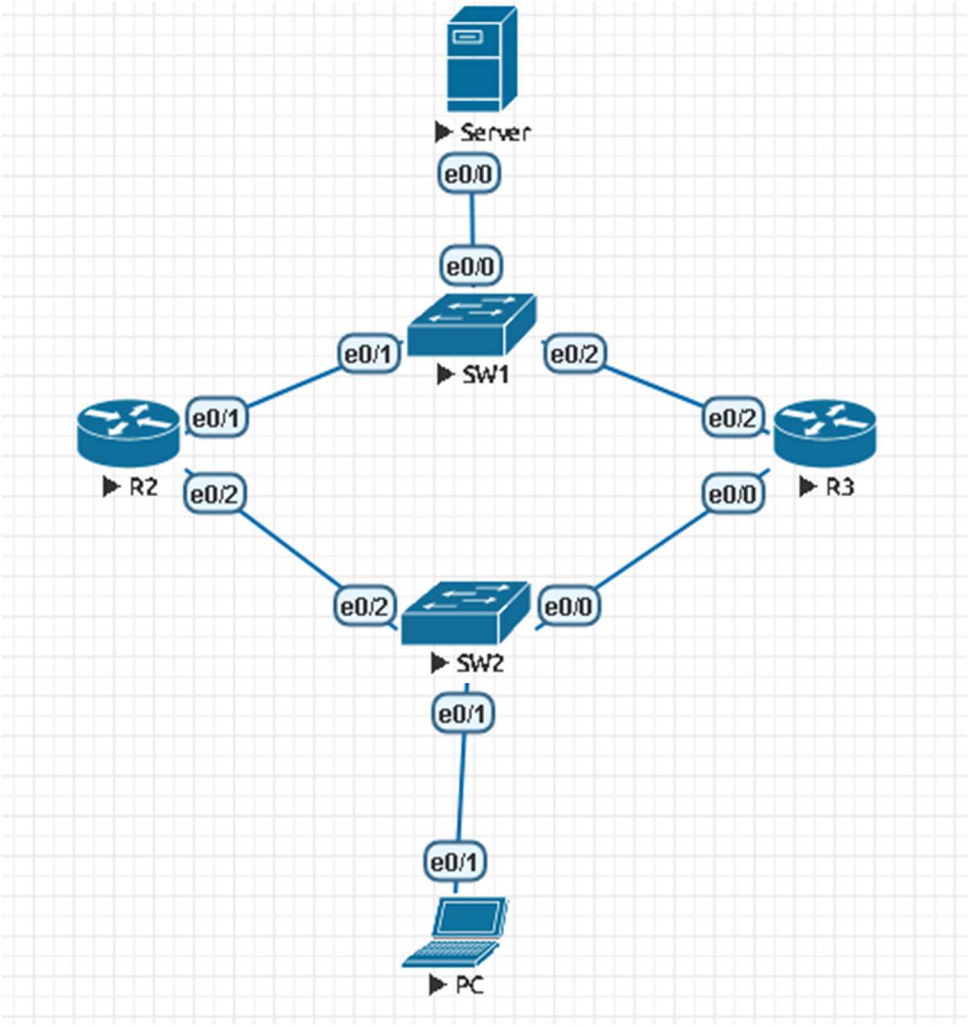
是否连接在这个接口上的所有 PIM 邻居都发回了 Prune 报文，这才停止转发组播流量。即

当路由器的某个接口上连接有两个 PIM 邻居，但其中一个未发回 Prune 报文，则路由器不会停止从向该接口转发报文，因为必须等两个 PIM 邻居都发回 Prune，这才监时停止转发报文，停止 3 分钟，即路由器会在该接口上开启一个 Prune 标记，并且开启一个 180S 的倒计时器，计时器过期后，再重新从该接口转发报文。

虽然路由器此时停止向该接口转发流量，但是其组播路由表中的(S,G)表项仍然还要存在 3 分钟，结果当 3 分钟过完时，Prune 计时器的 3 分钟也过了，路由器又开始转发组播流量，所以，不管是有无 Prune 功能，该路由器的组播表中的(S,G)表项是不会清失的，即不吃凉

粉，还占板凳。

在现实中包含 MA 网段的网络中，如下图的拓扑中



当信源 S 发送组播流量后，R2、R3 都会从 R1 处收到组播流量，由于 R2、R3、R4 这块是 MA 网段，当 R2 转发的组播流量到达交换机时，交换机会进行泛洪，即又会发至 R3 处；而 R3 同样将从 R1 处收到的组播流量发至交换机，交换机也会泛洪，同样也会到达 R2 处，这样就可能在 SW2 处产生重复报文(PC 可能会收到相同的重复报文)，**为了避免这种重复报文的出现，引入了 Assert 机制，在 R2、R3 中选择一台设备，让其可以转发组播流量，而另一个不可转发流量，这样避免了重复报文。**

Assert 机制在 PIM 的 Dense-mode 中很流行，其工作过程为：

R2、R3 通过接收(发送)相同组播流量的接口，在本拓扑中为 R2 的 e0/2 和 R3 的 e0/0，发送一个 Assert 报文，其中包含:

①R2、R3 去往信源路由的管理距离，越小越优先

②R2、R3 去往信源路由的度量值，越小越优先

③R2 的 e0/2 和 R3 的 e0/0 的 IP 地址，越大越优先

然后通过比较 Assert 报文中的以上信息，得出哪个接口是 Assert 机制中的获胜者，获胜者有且只有一个，经过比较后，**只有 Assert 的胜利者可以通过该接口转发组播流量，而失败者是不允许转发组播流量的。**

通过 Assert 机制，避免了重复报文的出现，即 SW2 仅仅可以通过 R2 或 R3 中的某一个接口收到组播流量，假设 Assert 的胜利者为 R2 的 e0/2，SW4 再泛洪，当 R3 的 e0/0 收到 SW4 泛洪的组播流量后，肯定没法通过 RPF 校验，R3 会将该组播报文丢弃，这样就保证 SW4 总是收到一份组播流量，避免了重复报文的出现。

而一旦 Assert 机制中的获胜者挂掉，即如果 R2 的 e0/2 挂掉，则 R3 的 e0/0 立刻转发组播流量。

Graft:嫁接

在 PIM 的 Dense-mode 中 ,当路由器没有 IGMP 的接收者 ,也没有其它的 PIM 邻居 ,此时，该路由器会向其上游发送一个 Prune 报文，告知上游暂停转发组播流量 3 分钟，

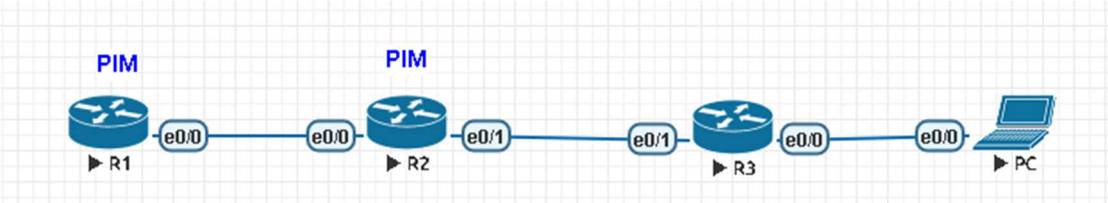
而当这台路由器刚向上游发送了 Prune 报文，其又发现下游有了 IGMP 的接收者或 PIM 邻居，此时，下游的 IGMP 接收者或 PIM 邻居，必须等待 3 分钟（最多 3 分钟）方才可以收到组播流量，这就与 PIM 的 Dense-mode 的优点（转发组播流量特别快）相悖。

为了解决这种 Prune 后突然又要转发组播报文而出现的长时间等待问题 ,引入了 Graft 机制，其工作原理如下：

当路由器突然发现**自己拥有了 PIM 邻居，或者突然拥有了 IGMP 的接收者**，并且自己的组播表项中有相应的(S,G)的记录，那么路由器就会向上游的路由器发送一个 Graft 报文，让上游路由器立即给自己转发流量，从而节省了因 Prune 报文而引发的 3 分钟的等待时间。

并且在不同的环境下，Graft 报文由不同的 PIM 路由器发的:

①情景一



R1 上游连接着组播信源，R1、R2 都启用了 PIM 的 Dense-mode，而 R3 上未启用 PIM 协议，此时问 PC 能否收到组播流量？

明显不能收到组播流量，因为 R2 收到 R1 的组播流量，其没有对应的 PIM 邻居，也没有 IGMP 组接收者，故其不会转发组播流量，并会向 R1 发回一个 Prune 报文，则 R1 不会再向 R2 转发组播流量。

如果 R3 突然启用 PIM 的 Dense-mode，请问 PC 能否立即收到组播流量(注意是立即)？如果收到组播流量，是否是因为 Graft 作用引发的?哪台 PIM 路由器发出了 Graft 报文？

此时 PC 肯定会收到组播流量，并且是立即收到，因为 Graft 特性的影响，但是**此时的 Graft 报文是由 R2 发出的，而不是由 R3 发出的。**

当 R3 上开启 PIM Dense-mode 之后，其组播表中没有(S,G)的表项的，因为发送 Graft 报文发送必须有个前提，即其组播表中必须要有对应的(S,G)表项，故 R3 仅仅只能向 R2 发出 hello 报文，用于建立 PIM 邻居，而不会发送 Graft 报文。

而路由器 R2，突然由于 R3 启用 PIM 协议，其会意识到自己有了 PIM 邻居，立即向 R1 发一个 Graft 报文，R1 收到 Graft 报文后，立即向 R2 转发组播报文，从而 PC 会立即收到组

播流量。

②情景二

在此环境中，如果 PIM 路由器 R3 之后，突然连接了组成员 PC，此时，PC 肯定会立即收到组播流量，并且该 Graft 报文是由 R3 发出的。

因为 R1 收到信源的组播流量后，会向 R2 转发组播流量，R2 也会转发组播流，R3 就会收到该组播流量，并在组播表中生成(S,G)的表项，这一现象是由于 PIM Dense-Mode 的转发特性——**收到流量完成 RPF 校验后，会向其 PIM 邻居和组成员泛洪而引起的**。

当 R3 收到组播流量，并在其组播表中形成(S,G)的表项后，其会向 R2 发出 Prune 报文，则 R2 会停止向 R3 转发组播流量。

当 R3 后的 PC 加组后，R3 会意识到有了组播流量接收者，故立即向 R2 发出 Graft 报文，R2 收到后，会立即向 R3 转发组播流量，PC 会立即收到组播流量，此时 R3 能向 R2 发出组播流量的原因——R3 的组播表中有(S,G)的表项。

PIM 的 Dense-Mode 的配置:

①运行 PIM 协议的路由器之间的互连接口

②PIM 路由器连接信源的接口

③PIM 路由器连接接收者的接口。