问鼎OSPF（2）-初涉路由显神通，民主治军四方拥

**绿油油的草儿，碧澄澄的溪~**

**轻飘飘的风儿，迎面吹~**

**哞嗨嗨的牛儿，青幽幽的堤~**

**乐陶陶的人儿，哼着歌~寐~**

“白龙马儿，蹄儿朝西，驮着唐三藏，跟着三徒弟~西天取经不容易，一走就是几万里~”那个乐陶陶的人好不自在。

“主公，这个人真的是那人吗？怎么如此吊儿郎当？”说话的人方脸大耳，膀阔腰圆，一看就是个久经沙场的武将。

“大隐隐于市，交换机，你可莫要小瞧了人家。举凡智计无双的奇人异士，纵有通天彻地之能，大都锋芒内敛，不显山露水。群英会上战RIP，本公是亲眼所见，此子风采，当真了得。”

被唤作主公的那人头戴侯冠，身披银[甲](https://www.baidu.com/s?wd=%E5%8F%B6%E7%94%B2&hl_tag=textlink&tn=SE_hldp01350_v6v6zkg6)，内罩白袍，威严凛凛。

“嘿嘿，那也未必。主公，当日此子确是舌灿莲花，但空有理论，说不定难于践行。”说话的这人天生一双鼠目，仍谁看了都觉得一脸奸猾相，“我们三人跟在他后面行了这么久，说不定他也只是故弄玄虚，区区一绣花枕头罢了。”

“呵呵，若能求得一贤才，多行一段路又何妨？”那主公见OSPF渐行渐远，立时催马扬鞭。并辔二人左右对视一眼，亦随即而去。

# 访诸葛三顾茅庐 论运行机制谋治世妙法

“三位跟随我多日，可悟出了什么？”行至一荫凉处，OSPF从牛背上跳了下来，对后面跟着的三人说道。

“…….”要是悟出来了，还能来寻你么？那主公心中如此想，但是脸上却没表现出来。他翻身下马，对着OSPF双手毕恭毕敬道：“先生说笑了，我三人皆是草莽出身，哪能悟得什么？此番前来，也是希望先生能助本公一臂之力，于互联网上开疆拓土，成不世伟业。”

“恩，悟得开了，悟得开了……”OSPF却似是自说自话一般，装出一副高人模样，心下却笑开了花。

“哼，黄口小儿，我家主公与你这般客气，你好生无礼！”交换机早就憋了一股气，终于忍不住喝道。

“恩，榆木脑袋，悟不开，悟不开……”

“你！”“嗳，”那主公拦下了交换机作势欲挥出去的臂腕，接着对OSPF道：“群英会上，本公听过先生所说的SPF与LSDB，很是钦佩。若真如先生所描绘那样，定当互联网之福，众网元之福。望先生入我麾下，本公必奉先生为上宾。”

“哈哈哈哈，主公谬赞，”OSPF再也不似先前散漫模样，既而毕恭毕敬道：“主公身旁两员猛将当是交换机、路由器，那主公必是携众网元开拓繁多网点的网管了。今幸得主公如此礼遇，再推脱就是OSPF的不是了。”

“嘿嘿嘿嘿，你这人倒是懂得顺坡骑驴，”路由器一脸奸笑，“SPF与LSDB莫不是一纸空谈吧？”

“呵呵，无妨，主公请看，”说话间OSPF顺手打开随身带的几幅图，“下山前，师父曾赐与我五行八阵图，嘱我多加揣摩，如今终得其精髓。此图在手，定能助主公争霸天下！”

“哈哈！先生快快说予我听！”网管早已按捺不住，忙对OSPF道。

OSPF见此情景，当下再不废话，徐徐与他们道来。

此组图说的，正是OSPF协议的基本运行机制，恰似百尺高楼之基石。其实概括起来，也比较容易理解，包括下面5步：

**（1）Hello协议交互-形成邻居关系**

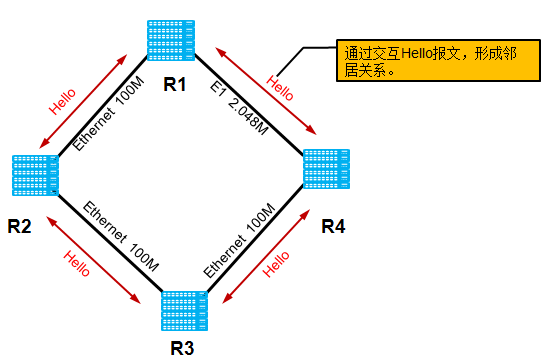


图1 Hello协议交互-形成邻居关系

如图1所示，设备运行OSPF协议后，会从所有启动OSPF协议的接口上发出Hello数据包。如果两台设备共享一条公共数据链路，并且能够互相成功协商他们各自Hello数据包中所指定的某些参数，那么他们就成了邻居关系。

**（2）LSAs的泛洪-通告链路状态信息**

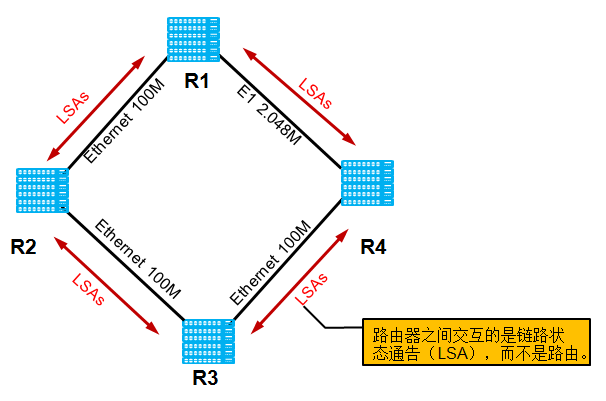


图2 LSAs的泛洪-通告链路状态信息

形成邻接关系的设备之间就可以交互LSA了，如图2所示。通过问鼎OSPF（1）我们知道，LSA全称是链路状态通告，它描述了设备所有的链路、接口、邻居及链路状态等信息。设备间正是不断的泛洪交互这些链路信息，来了解整个网络的拓扑信息。由于链路的多样性，OSPF协议定义了许多LSA类型，问鼎OSPF（3）会详细介绍到。

同时我们可以了解到OSPF比RIP的高明之处在于，设备之间交互的是LSA信息，而RIP协议交互的是路由。也就是说，OSPF协议中，设备的选路是一种“自主行为”，LSA只不过是一种选路的参考信息；而RIP协议中，设备的选路过度的依赖于邻居路由器的路由信息，即使邻居路由器传达的信息是错误的。

**（3）LSDB的组建-形成带权有向图**

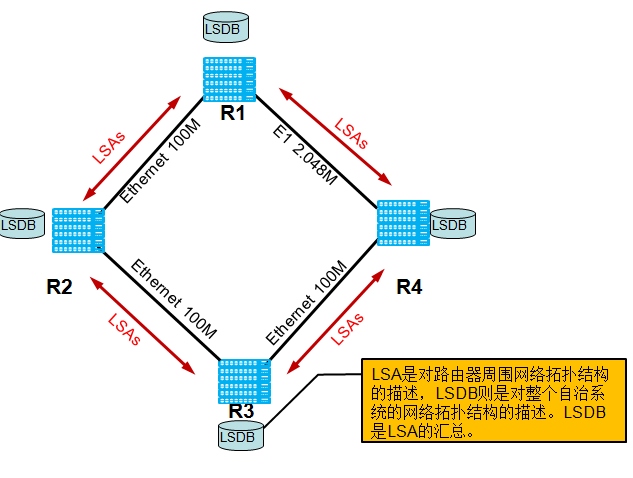


图3 LSDB的组建-形成带权有向图

通过泛洪LSA，设备接下来会把收到的LSA汇总记录在LSDB（链路状态数据库）当中，最终，所有路由器都会形成同样的LSDB，如图3所示。LSA是对设备周围网络拓扑结构的描述，而LSDB则是对整个自治系统的网络拓扑结构的描述，LSDB是LSA的汇总。

**（4）**SPF的计算-形成路由

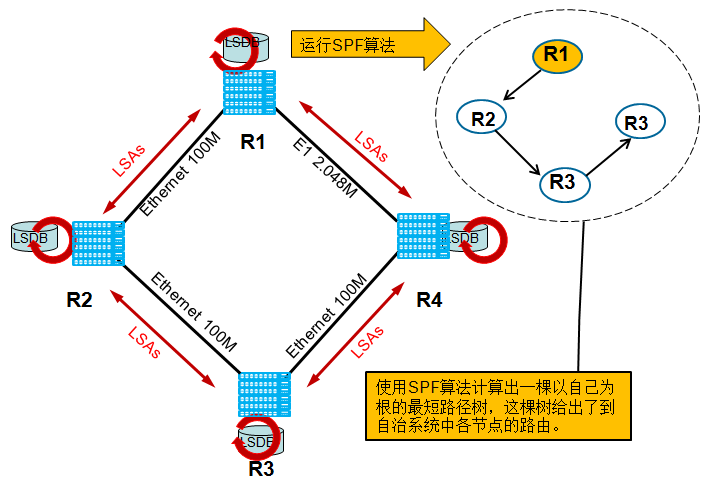


图4 SPF的计算-形成路由

如图4所示，当LSDB同步完之后，每一台设备都将以其自身为根，使用SPF算法来计算一个无环路的拓扑图，以描述它所知道的到达每一个目的地的最短路径（最小的路径代价）。这个拓扑图就是SPF算法树，有了这棵“树”，事实上路由器就已经知道了到达网络各个角落的最优路径。

**（5）路由表的维护更新**

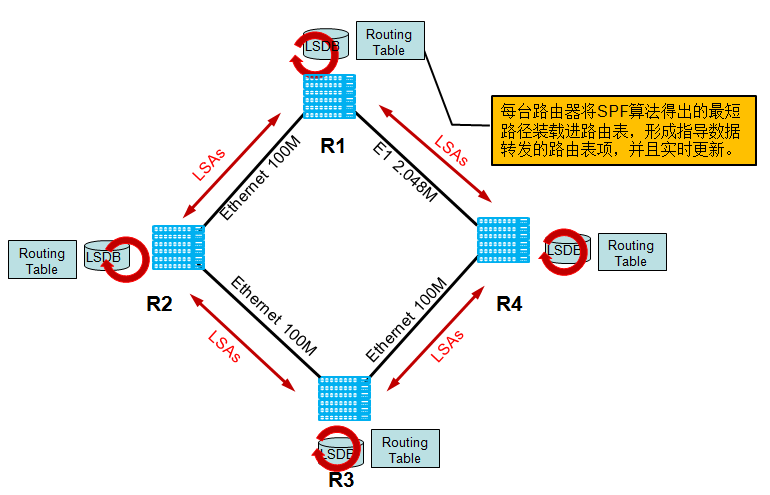


图5路由表的维护更新

最终，经过SPF算法得出了最短路径树，每台路由器将SPF算法得出的最短路径装载进路由表形成指导数据转发的路由表项，并且实时更新，如图5所示。同时，邻居之间交互Hello报文进行Keepalive，维持邻居/邻接关系，并且每30min重传一次LSA。如果网络拓扑稳定，那么网络中将不会有什么活动或行为发生。

# 领虎符整饬三军 设Router ID知己知彼

“先生这妙法如此运用，令本公茅塞顿开，我想再复杂的网点也必能攻的下来。先生若入我帐下，本公愿与你调兵虎符，麾下三军尽皆任你调遣！”到现在，网管是愈加的对OSPF佩服的五体投地。

“蒙主公抬爱，微臣自当鞠躬尽瘁，死而后已！”

自那日网管三顾茅庐之后，OSPF便随网管来到了某网点大营之中。借由帐中沙盘，OSPF知道此网点名唤X网点，内里错综复杂，地域辽阔。网点内不同网元鳞次栉比，而且网元间需要交错相连，任什么路由协议见了，肯定都会头疼不已。

“军师，俺叫路由器！”在大营中某一处，一个大头兵正憨憨的答着OSPF的话。OSPF并未急于攻城掠地，来了之后，便先是熟悉三军。

“那么你呢？”OSPF持扇轻摇，转身对另一士兵道。

“军师，俺也叫路由器！”

“俺也叫路由器！”

“……”

“你叫路由器，你也叫路由器，真是叫我有些晕头转向了。”OSPF摇头笑道。

“嘿嘿，军师，俺们本来就都是路由器啊。何况俺们这些大老粗，名讳也不值几个钱，哪能与那些有头有脸的协议比。”

“不然，知己知彼，方百战不殆。也莫要小瞧自己，哪个协议离了你们能玩的转。本军师要了解整网拓扑，亦要靠你们，”OSPF沉吟半晌，既而肃容道：“众将听令！卯时三刻，校场集合，军师来给你们发Router ID了！”

Router ID就是用于在自治系统中唯一标识一台运行OSPF的路由器的32位整数。每个运行OSPF的路由器都有一个Router ID。Router ID的格式和IP地址的格式是一样的，在实际网络部署中，考虑到协议的稳定，推荐使用路由器Loopback0的IP地址做为路由器的Router ID。

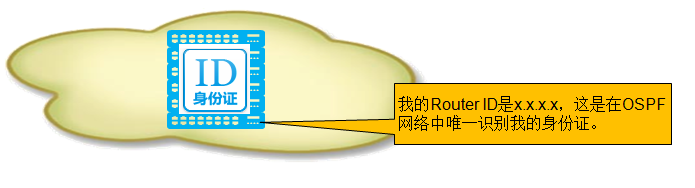


图6 OSPF自治系统中路由器的的Router ID

Router ID的选取有两种方式：

* 通过命令行手动配置。
* 设备自动设定。

如果没有手动配置Router ID，设备会从当前接口的IP地址中自动选取一个作为Router ID。其选择顺序是：

1、优先从Loopback地址中选择最大的IP地址作为Router ID。

2、如果没有配置Loopback接口，则在接口地址中选取最大的IP地址作为Router ID。

只有重新配置系统的Router ID或OSPF的Router ID，并且重新启动OSPF进程后，才会进行Router ID的重新选取。

# 因地域分四路掠城 携天威攻万变贼寇

“9.5.2.7！”

“到！”

“9.5.2.8！”

“到！”

数日过后，三军阵前校场点兵，上至虎虎将尉，下至大头小兵，莫不精神抖擞，焕然一新。OSPF心下了然，便开始排兵布阵，准备拿下X网点。而他根据不同的链路类型，将三军分作了四路，每路阵法不尽相同：

1. 广播（Broadcast）类型

当链路层协议是Ethernet、FDDI时，OSPF缺省认为网络类型是Broadcast。在该类型的网络中，通常以组播形式（224.0.0.5：含义是OSPF路由器的预留IP组播地址；224.0.0.6：含义是OSPF DR的预留IP组播地址）发送Hello报文、LSU报文和LSAck报文；以单播形式发送DD报文和LSR报文。

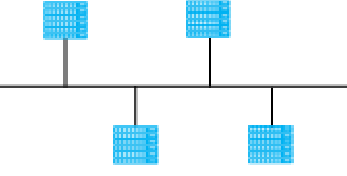


图7 广播（Broadcast）类型网络

1. NBMA（Non-Broadcast Multi-Access）类型

当链路层协议是帧中继、ATM或X.25时，OSPF缺省认为网络类型是NBMA。在该类型的网络中，以单播形式发送协议报文（Hello报文、DD报文、LSR报文、LSU报文、LSAck报文）。

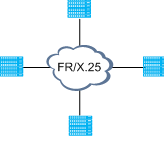


图8 NBMA（Non-Broadcast Multi-Access）类型网络

1. 点到多点P2MP（point-to-multipoint）类型。

没有一种链路层协议会被缺省的认为是Point-to-Multipoint类型。点到多点必须是由其他的网络类型强制更改的。常用做法是将非全连通的NBMA改为点到多点的网络。在该类型的网络中，以组播形式（224.0.0.5）发送协议报文（Hello报文、DD报文、LSR报文、LSU报文、LSAck报文）。

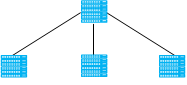


图9 点到多点P2MP（point-to-multipoint）类型类型网络

1. 点到点P2P（point-to-point）类型

当链路层协议是PPP、HDLC和LAPB时，OSPF缺省认为网络类型是P2P。在该类型的网络中，以组播形式（224.0.0.5）发送协议报文（Hello报文、DD报文、LSR报文、LSU报文、LSAck报文）。

图10 点到点P2P（point-to-point）类型网络.png

图10 点到点P2P（point-to-point）类型网络

# 拔敌寨遭遇险境 举能人解广播/NBMA邻间困局

“哈哈哈，先生真乃神人也！据探子来报，P2P和P2MP网络已被攻破。不出五日，X网点将被全部拿下！”网管坐在马上，抚掌大笑道。

“主公，此话言之尚早，剩下的两路可不好拿。”OSPF一反往日淡然自若，神色中竟带有一丝隐忧。

“报！”正当时，一探子快马加鞭至两人跟前。只见那探子衣衫尽皆血染，满目疲惫之色，“广播链路久攻不下，如不及时驰援，我军危矣！”

“什么！怎会如此！”网管身形一颤，险些跌下马来。

“初时我军攻敌势如破竹。但….”探子回话突地一顿，悲戚中难掩一丝愤恨，“但后来路将军不知与敌贼言语了什么，广播链路上需要连的设备竟越来越多。我军结两两相交之阵，一开始堪堪抵住汹涌来势，及至最后阵法渐露疲相，我军亦深陷泥潭！”

OSPF听罢静默无言，待心绪稍自平复，忙对来人道：“我问你，阵中那白袍小将安在？”

“这……身披白袍的将士有好些个，却不知军师说的是哪一位？”

“就是身边总有个容貌清秀的小兵围着他叽叽喳喳转个不停的那一位。”

“哦，原来是那一位！”探子恍然大悟，“我知道的！幸有此人一马当先，不然只怕我军早已败下阵来。”

“我赐你锦囊一枚，速交于他手！”

“是！”探子接过锦囊，一骑绝尘而去。

“先生，此子是何人物？能替本公拿下这一仗么？”网管言语间很是急切。开始如此顺利，现在若停歇下来，可真是功亏一篑了。

“早先我料到可能有此劫，攻打广播链路/NMBA链路时便定一计，但需营中素有威名之人相助。此子由众人推举而来，原先籍籍无名，现名DR！”

## DR/BDR选举的原因

在广播网和NBMA网络中，任意两台路由器之间都要传递路由信息。如图11所示，网络中有n台路由器，则需要建立n\*(n-1)/2个邻接关系。这使得任何一台路由器的路由变化都会导致多次传递，浪费了带宽资源。为解决这一问题，OSPF定义了指定路由器DR。通过选举产生DR（Designated 路由器）后，所有其他设备都只将信息发送给DR，由DR将网络链路状态LSA广播出去。为了防止DR发生故障时，重新选举DR时会产成业务中断，除了DR之外，还会选举一个备份指定路由器BDR。这样除DR和BDR之外的路由器（称为DR Other）之间将不再建立邻接关系，也不再交换任何路由信息，这样就减少了广播网和NBMA网络上各路由器之间邻接关系的数量。这就是OSPF给那白袍小将的锦囊妙计。

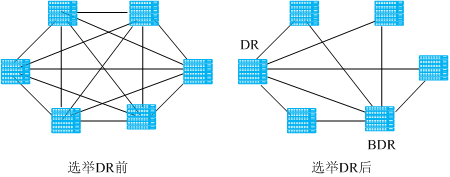


图11 DR和BDR的选举

## DR/BDR选举的原则

在广播网和NBMA网络中，为了稳定地进行DR和BDR的选举，OSPF规定了一系列的选举规则，遵循三个基本原则：**选举制、终身制、世袭制**。下面我们来逐个介绍一下。

**1、选举制**

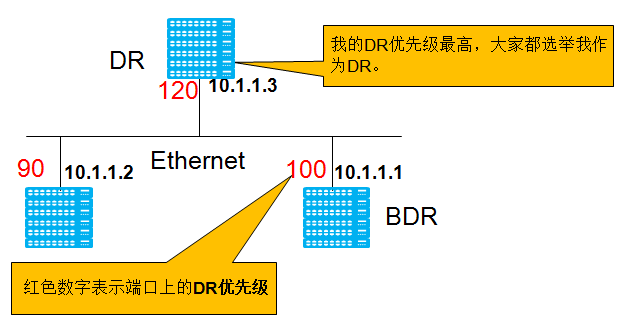


图12 DR和BDR的选举-选举制

如图12所示，所谓选举制即DR和BDR不是人为指定的，而是由本网段中所有的路由器共同选举出来的。路由器接口的DR优先级决定了该接口在选举DR、BDR时所具有的资格。本网段内DR优先级大于0的路由器都可作为“候选人”。选举中使用的“选票”就是Hello报文。每台路由器将自己选出的DR写入Hello报文中，发给网段上的其他路由器。当处于同一网段的两台路由器同时宣布自己是DR时，DR优先级高者胜出。如果优先级相等，则Router ID大者胜出。如果一台路由器的优先级为0，则它不会被选举为DR或BDR。

**2、终身制**

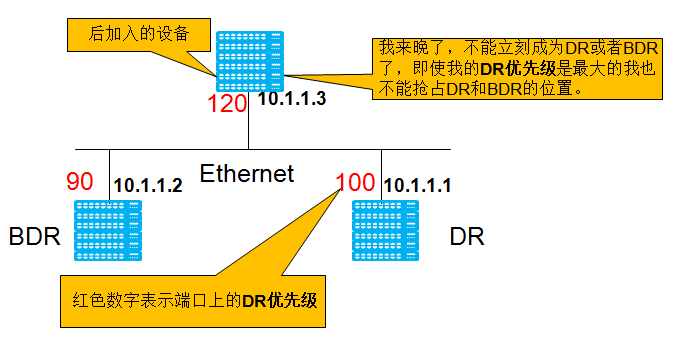


图13 DR和BDR的选举-终身制

所谓终身制也叫非抢占制。每一台新加入的路由器并不急于参加选举，而是先考察一下本网段中是否已有DR存在。如果目前网段中已经存在DR，即使本路由器的DR优先级比现有的DR还高，也不会再声称自己是DR了。而是承认现有的DR。因为网段中的每台路由器都只和DR/BDR建立邻接关系，如果DR频繁的更迭，则每次都要重新引起本网段内的所有路由器与新的DR/BDR建立邻接关系。 这样会导致在短时间内网段中有大量的OSPF协议报文在传输，降低网络的可用带宽。终身制有利于增加网络的稳定性、提高网络的可用带宽。实际上在一个多访问网络上，最先初始化启动的两台具有DR选举资格的路由器将成为DR和BDR路由器。

**3、世袭制**

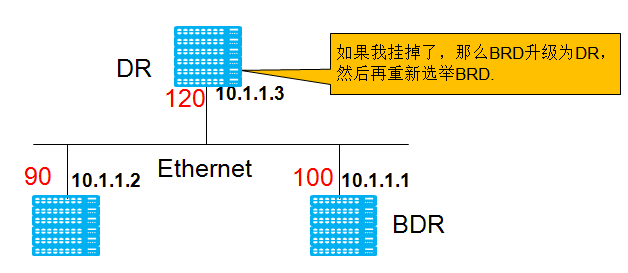


图14 DR和BDR的选举-世袭制

所谓世袭制就是如果DR故障了，那么下一个当选为DR的一定是BDR，其他的路由器只能去竞选BDR的位置。这个原则都是为了保证DR是比较稳定的，不会经常进行选举的，并且DR是有备份的（BDR），一旦DR失效，可以马上由BDR来承担DR的角色，由于DR和BDR的数据库是完全同步的，这样当DR故障后，BDR立即成为DR，履行DR的职责，而且邻接关系已经建立，所以从角色切换到承载业务的时间会很短。同时，在BDR成为新的DR之后，还会选举出一个新的BDR，虽然这个过程所需的时间比较长，但这个已经不会影响路由的计算了。

## DR/BDR选举的详细步骤

广播链路或者NMBA链路上DR/BDR详细的选举过程如下：

（1）接口UP后，发送Hello报文，同时进入到waiting状态。在waiting状态下会有一个waiting timer，该timer的长度与dead timer是一样的。默认值40s，用户不可自行调整。

（2）在waiting timer触发前，发送的hello报文是没有DR和BDR字段的。在waiting阶段，如果收到Hello报文中有DR和BDR，那么直接承认网络中的DR和BDR，而不会触发选举。直接离开waiting状态，开始邻居同步。

（3）假设网络中已经存在一个DR和一个BDR，这时新加入网络中的设备，不论它的Router ID或者DR优先级有多大，都会承认现网中已有的DR和BDR。

（4）当DR因为故障down掉之后，BDR会继承DR的位置，剩下的优先级大于0的设备会竞争成为新的BDR。

（5）只有当不同Router ID，或者配置不同DR优先级的设备同时起来，在同一时刻进行DR选举才会应用DR选举规则产生DR。该规则是：优先选择DR优先级最高的作为DR，次高的作为BDR。DR优先级为0的设备只能成为DRother；如果优先级相同，则优先选择Router ID较大的设备成为DR，次大的成为BDR，其余设备成为DRother。

**江湖小贴士：**

**看完这一节之后，小伙伴们是不是已经满腹疑问？什么是waiting状态？dead timer又是个什么鬼？不要急，等看了第三篇，再回头看这里，你就会一目了然。**

## DR/BDR选举的实验验证过程

前面给大家介绍了OSPF在广播和NBMA网络中DR/BDR的选举原则和选举步骤，下面我们通过实验来验证一下DR和BDR的选举过程中的一些问题。

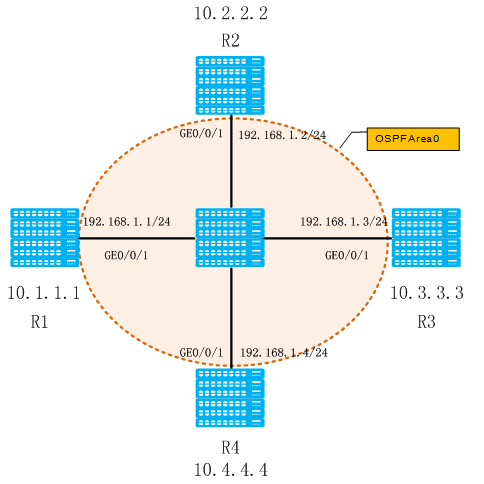


图15 实验验证DR和BDR的选举

实验环境如图15所示，五台路由器组成一个广播网络，中间的交换机作为纯二层设备，四周的四个路由器R1、R2、R3、R4作为路由设备。四个路由器都规划在OSPF的Area0区域内，各路由器的IP地址及Router ID如图15所示。

R1、R2、R3、R4四台路由设备的关键配置如下：

|  |  |
| --- | --- |
| R1的关键配置 | #  ospf 1 Router ID 10.1.1.1  area 0.0.0.0  network 192.168.1.0 0.0.0.255  # |
| R2的关键配置 | #  ospf 1 Router ID 10.2.2.2  area 0.0.0.0  network 192.168.1.0 0.0.0.255  # |
| R3的关键配置 | #  ospf 1 Router ID 10.3.3.3  area 0.0.0.0  network 192.168.1.0 0.0.0.255  # |
| R4的关键配置 | #  ospf 1 Router ID 10.4.4.4  area 0.0.0.0  network 192.168.1.0 0.0.0.255  # |

配置完成后，待网络稳定后我们来查看一下当前这个广播网络的DR和BDR的选举情况。

在R1上通过display ospf peer命令查看OSPF的邻居信息，显示如图16所示。

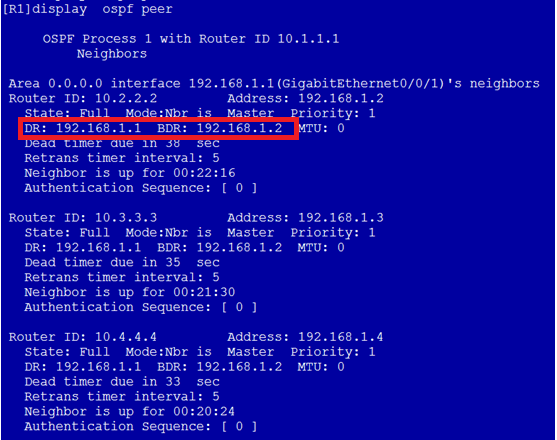


图16 R1上通过display ospf peer命令查看OSPF的邻居信息

从上述表项可以看出，目前这个广播网络已经完成了DR和BRD的选举，R1是DR，R2是BDR，R3和R4是DRother。这里R1是DR，R2是BDR跟设备的启动顺序也是直接相关的，本例中我们按照R1、R2、R3、R4的顺序启动设备，所以R1和R2首先完成了初始化，自然成为了DR和BDR。

下面分别在R1、R2、R3和R4上查看OSPF邻居关系的概要信息。

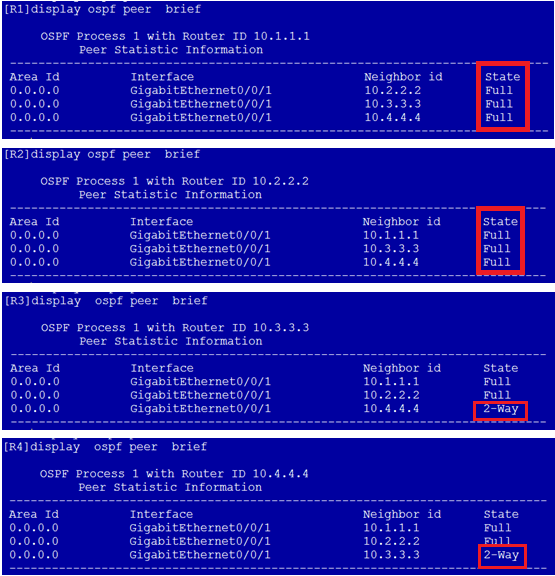


图17 完成DR和BDR选举后的OSPF邻居关系的概要信息

我们可以从图17的表项中看到：R1/R2和其他三个设备的邻居关系都是Full，

而R3和R4之间的OSPF邻居关系是2-Way状态。这表示DR/BDR与邻居间建立的是邻接关系，

而DROther之间建立的只是邻居关系。

**说到这儿，小伙伴们是不是又产生了很多新的疑问？为什么Full就表示邻接关系，2-Way就表示邻居关系？邻居关系和邻接关系又有什么区别？嘿嘿，咱们还是等问鼎OSPF（3）讲解邻居状态机的时候会详细展开介绍这些内容。**

**江湖小贴士：**

1. **如果广播网络里只存在DR，但并不存在BDR，这时候网络是什么状况？**

我们通过实验来回答读者的疑问，我们在R2、R3、R4的接口视图下配置**ospf DR优先级 0**命令将设备的DR优先级配置为0，那么这个时候这3台设备将失去DR和BDR的选举资格，只能作为DRother，网络中仅存在一台具备DR/BDR选举资格的设备，就是R1。

配置完成后我们查看OSPF邻居状态可以看到，此时DR是R1，BDR显示为None，即网络中不存在BDR，如图18-1所示。

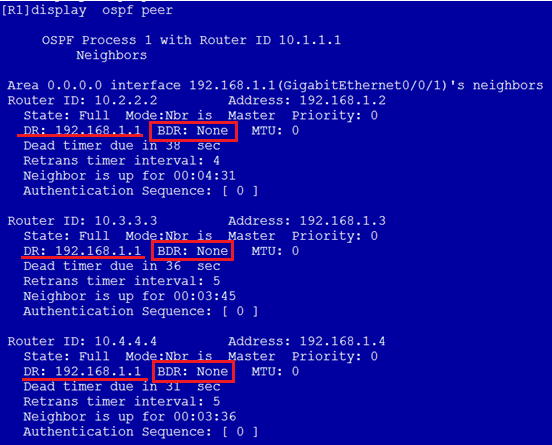


图18-1 网络中只存在DR不存在BDR的情形

此时我们查看四台设备的OSPF简要信息，如图图18-2所示，可以看到R2、R3、R4分别和R1建立了邻接关系（状态为FULL），而R2、R3、R4之间的邻居状态只停留在2-Way的状态。

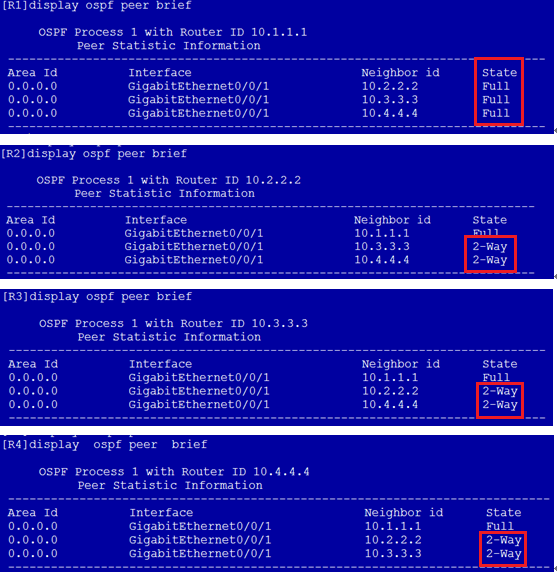


图18-2 网络中只存在DR不存在BDR的情形

实验过程已经基本看到了这个问题的答案，这里我们总结说明一下：如果在一个多访问网络上只有唯一的一台具有选取资格的路由器相连，那么这台路由器将成为DR路由器，而且在这个网络上没有BDR路由器。其他所有的路由器都将只和这台DR路由器建立邻接关系。聪明的读者此时可能会问另外一个问题，如果这个时候R1故障了，或者失去了DR的资格了呢？请看紧接着的下面第二个小贴士。

**江湖小贴士：**

1. **如果广播网络里不存在DR和BDR，这时候网络是什么状况？**

我们同样用实验来回答这个问题，我们在上面的基础上，继续在R1的接口视图下配置**ospf DR优先级 0**命令，让R1也失去DR/BDR的选举资格，此时这个广播网络中将没有任何设备具备DR/BDR的选举资格。如图19-1所示，DR和BDR字段都显示None，表示这个网络中不存下DR和BDR。

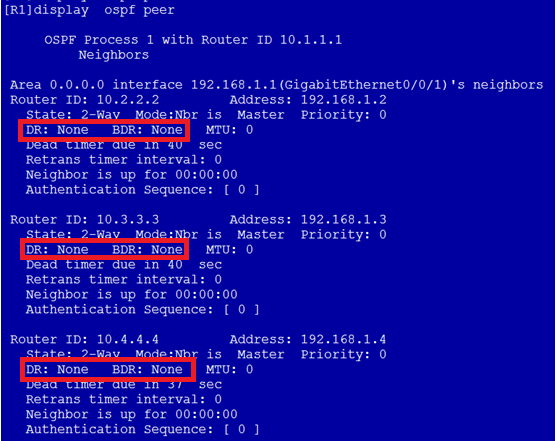


图19-1 网络中不存在DR和BDR的情形

此时我们查看四台设备的OSPF简要信息，如图图19-2所示，可以看到此时所有的邻居状态都只停留在2-Way的状态，网络不能建立邻接关系，各个设备之间不能完成路由信息的交互。

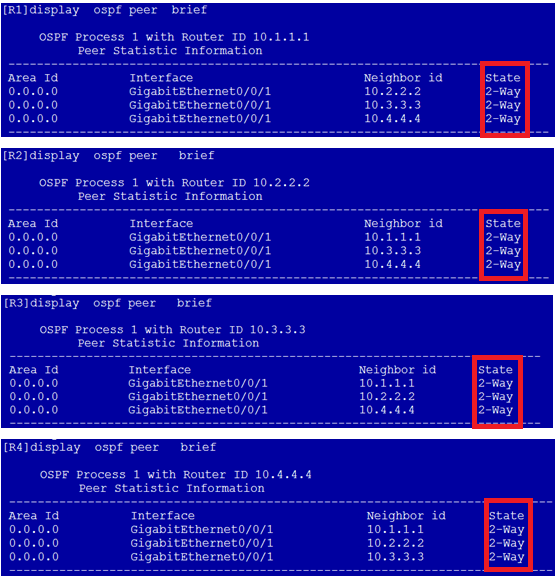


图19-2 网络中不存在DR和BDR的情形

根据上述实验，我们可以给出这样的结论：如果广播网络上不存在具备DR/BDR选举资格的设备，那么这个网络上将没有DR或者BDR路由器，而且也不会建立任何邻接关系。在这种情况下，网络上所有的路由器的邻居状态都将停留在“2-Way”状态。

# 尾声

残阳如血。

寂寥的苍穹中偶有几只归雁飞过，伴随着一声声悠扬的啼鸣在长空中回荡。

然而，眼前的一众人却无暇他顾，他们都在目不斜视的盯着远方，每个人脸上都像崩了一根弦，似是那里有着什么[择人而噬](http://www.baidu.com/link?url=wFxPdlLCfMZrsvSeTEUSZvicfj6n7iHtSPEaVRj68_VB22TR-hC6nRCHJfvLvWnyZSDZags3pmXOGxNIHtMRxK)的恐怖怪物在等着他们。

突然，远方的地方线上出现了一个小黑点，紧接着，两个，三个…….那些黑点越来越多，也越来越大，及至最后汇聚成一列铁骑，由远及近呼啸而来。

“主公，我们胜了。”

白袍小将，横刀立马。