

OSPF 基础

OSPF 知识点：

OSPF 基本配置，查看 OSPF 的 3 张表，OSPF 5 种报文，7 种邻居状态，4 种网络类型，4 种特殊区域，7 类 LSA，4 种 link type，标识一条 LSA 的 3 个要素，OSPF 邻居与邻接关系，单区域、多区域、OSPF 认证 (keychain) ,DR 的选举，OSPF 被动接口，OSPF 域间聚合，外部聚合，不规则区域的解决办法 (虚链接)，OSPF 开销值、协议优先级及计时器的修改，OSPF 外部路由引入，下放默认路由，OSPF 协议中 Forwarding Address 的理解

=====

先选举 BDR，后有 DR

因为 DR 和 BDR 的切换状态机是：当 DR 失效时，BDR 成为 DR

如果先选举 DR，再选举 BDR，那么当选举 BDR 的过程中 DR 失效，此时网络中既没有 DR 也没有 BDR，切换将无法进行，状态机也就没办法做了。

所以先有 BDR，后有 DR 是为了保证状态机能正常工作。

DR（村长）的选举过程

- ◆ 登记选民

本网段内的OSPF路由器；

本村内的18岁以上公民。

- ◆ 登记候选人

本网段内的priority>0的OSPF路由器；

本村内的30岁以上的公民，且在本村居住3年以上。

- ◆ 竞选演说

所有的priority>0的OSPF路由器都认为自己是DR；

所有的候选人都自认为应该当村长。

- ◆ 投票

选priority值最大的，若priority值相等，选Router ID最大的；

选年纪最大的，若年龄相等，按姓氏笔划排序。

OSPF 邻居建不起来的原因

- 1.hello 间隔和 dead 间隔不同；
- 2.接口下 OSPF 网络类型不匹配。
- 3.区域号码不一致；
- 4.特殊区域（如 stub，nssa 等）区域类型不匹配；
- 5.认证类型或密码不一致；
- 6.路由器 router-id 相同；
- 7.链路上的 MTU 不匹配；
- 8.在 broadcast 链路上的子网掩码不匹配
- 9.在 MA 网络中，没有 DR
- 10.接口设置为 silent-interface

```
int g0/0/0
mtu 1400
ospf mtu-enable
```

ospf mtu-enable 两端都要设置

ospf mtu-enable 命令用来使能接口在发送 DD 报文时填 MTU 值。缺省情况下，接口发送 DD 报文时 MTU 值为 0，即不填接口的实际 MTU 值。

大家直到 OSPF 在达到 FULL 关系过程中，可能会由于 MTU 不匹配导致 DD 报文交互出现问题，无法达到 FULL 状态。因此为了避免不同厂商设备对接时出现该问题，默认情况下发送 DD 报文时不填写接口的实际 MTU 值，即为 0。这样避免对端由于 DD 报文中 MTU 值大于接口 MTU，导致邻居关系无法建立。

强调下，该命令只是控制发 DD 报文的情况。HW VRP 对于收的 DD 报文，就是不检查 MTU 的。所以 MTU 的匹配要通过专门确认。当然不检查 MTU 也会存在一些问题，就不能通过 OSPF 发现两端接口 MTU 不一致的情况。如果没有及时发现 MTU 不一致，上业务后会出现有丢包情况。

OSPF 几个需要注意的地方：

- 1) 当 hello 时间不同时是永远起不来邻居的
- 2) 当 hello 时间不同时会停留在 INIT 状态
- 3) 如果路由的优先级都改成了 0，会停留在 TWO-WAY 状态
- 4) 当 MTU 值不同时停留在 EXSTART 或 EXCHANGE 状态

OSPF 使用两个多播地址

224.0.0.5---All OSPF Routers.

224.0.0.6---All DR Routers (DR+BDR)

224.0.0.5 指在任意网络中所有运行 OSPF 进程的接口都属于

该组，于是接收所有 224.0.0.5 的组播数据包
224.0.0.6 指一个多路访问网络中 DR 和 BDR 的组播接收地址

MA 网络 路由器发向 DR.BDR 的为 224.0.0.6
DR.BDR 回应的是 224.0.0.5

OSPF 进程号

OSPF ProcessID 只是本地有效，用来在本地区分不同的 OSPF 进程

虽然 OSPF ProcessID 本地有效，但是我们仍然建议，除非有特殊需求，否则全网使用一致的 OSPF 进程 ID，这将方便网络管理和维护

什么情况下需要用到多 OSPF 进程呢？这个问题可就大了，一般而言，一台路由器创建一个 OSPF 进程，可以实现网络的需求，但是在一些特殊的、复杂的环境当中，单一的 OSPF 进程可能无法满足要求，因此在一台路由器上创建多个 OSPF 进程，这样一来的直接好处是，这台路由器从原先的内部路由器--OSPF interArea Router，华丽的变成了一台 ASBR，它可以在进程之间引入、可以过滤路由或 LSA、可以控制路由、可以做基于 route-policy 的策略、可以做路由聚合等等。



R2: 不同进程的引入

```

ospf 1 router-id 2.2.2.2
area 0
net 192.168.12.2 0.0.0.0
ospf 2 router-id 2.2.2.2
area 0
net 192.168.23.2 0.0.0.0

```

```

ospf 1
import-route ospf 2
ospf 2
import-route ospf 1

```

R1: 查看路由表，
学习到的3.3.3.3外部路由

```

3.3.3.3/32  O_ASE  150  1

```

R1: dis ospf abr-asbr

RtType	Destination	Type
Intra-area	2.2.2.2	ASBR

OSPF 链路状态序列号

最大寿命定时器、刷新定时器和链路状态序列号一起确保数据库只包含最新的链路状态记录。

1) .LSDB 中每一个 LSA 都有一个序列号，序列号越大，LSA 越新。

2) .序列号范围从 0x80000001-0x7FFFFFFF

3) .OSPF 每 30 分钟 flood 一次 LSA 来维持 LSDB 同步，每次 flood 序列号加 1

4) .当一个路由器遇到同一个 LSA 的两个实例时，它必须能够确定哪一个是最接近的 LSA。（根据序列号来识别）

当一条 LSA 的序列号到达最大序列号时，始发路由器会发送一个生存时间为最大值的 LSA，让其它的路由器从 LSDB 中清除这条 LSA，当其它路由器确认后，再发送一个初始序列号的 LSA。注意：只有始发路由器才可以提前使这条 LSA 老化，LSA 条目的老化时间默认是一小时（0-3600S）

判断 LSA 的新旧：

1. 序列号越大，则 LSA 越新
2. 若序列号相同，则校验和较大被认为更新
3. 若序列号与校验和的比较都不能确定出最新的 LSA 时，则会比较老化时间。若 LSA 的老化时间被设置为 max age（默认 3600），则为最新
4. 若没有 LSA 的老化时间被设置为 max age，则看老化时间的相差值。在 LSDB 中，如果老化时间相差在大于 15min 以上，则 Ls age 的值越小，说明 LSA 越新；如果相差在 15min 内，则认为两条 LSA 是一样的。

当收到某条相同的 LSA 的多个实例时，路由器将通过下面的算法来确认哪个是最新的 LSA 实例：

- 1、比较 LSA 实例的序列号，拥有最大序列号的 LSA 就是最新的 LSA；
- 2、如果 LSA 实例的序列号相同，那么将会比较它们的校验和。拥有最大的无符号校验和的 LSA 就是最新的 LSA；
- 3、如果 LSA 实例的校验和也相同，那么就进一步比较老化时间。如果只有一条 LSA 拥有大小为最大生存时间的老化时间，那么就认为这条 LSA 是最新的 LSA；
- 4、如果这些 LSA 的老化时间之间的差别多余 15min，那么拥有较小老化时间的 LSA 将是最新的 LSA；
- 5、如果上述的条件都无法区别最新的 LSA，那么这两个 LSA 就被认为是相同的。

以上内容摘自《TCP/IP 路由技术第一卷》

OSPF 协议 link-type

表 3-3

各种链路类型及其描述

链路类型	描述	链路 ID	链路数据
1	点对点连接到另一台路由器	邻居的 Router-ID	产生该 LSA 的路由器的接口 IP 地址
2	连接到一个传输网络	DR 的接口 IP 地址	产生该 LSA 的路由器的接口 IP 地址
3	连接到一个末梢网络	网络 IP 地址	网络掩码
4	虚链路	邻居的 Router-ID	产生该 LSA 的路由器的接口 IP 地址

> OSPF Header

✓ LS Update Packet

Number of LSAs: 1

✓ LSA-type 1 (Router-LSA), len 48

.000 0000 0000 0001 = LS Age (seconds): 1

0... = Do Not Age Flag: 0

> Options: 0x02, (E) External Routing

LS Type: Router-LSA (1)

Link State ID: 2.2.2.2

Advertising Router: 2.2.2.2

Sequence Number: 0x80000010

Checksum: 0xe16d

Length: 48

✓ Flags: 0x00

.... .0.. = (V) Virtual link endpoint: No

.... ..0. = (E) AS boundary router: No

.... ...0 = (B) Area border router: No

Number of Links: 2

Router Lsa 中的 link type 又分为 4 类:

Point to Point link

描述链路是 P to P

Stub network link

描述网段信息

Trans network link

描述 DR.BDR

Virtual-link link

描述虚链接

- ▼ Type: **Stub** ID: 2.2.2.2 Data: 255.255.255.255 Metric: 0
 Link ID: 2.2.2.2 - IP network/subnet number
 Link Data: 255.255.255.255
 Link Type: 3 - Connection to a stub network
 Number of Metrics: 0 - TOS
 0 Metric: 0
- ▼ Type: **Transit** ID: 192.168.123.1 Data: 192.168.123.2 Metric: 1
 Link ID: 192.168.123.1 - IP address of Designated Router
 Link Data: 192.168.123.2
 Link Type: 2 - Connection to a transit network
 Number of Metrics: 0 - TOS
 0 Metric: 1

- * Link ID: 2.2.2.2
 Data : 255.255.255.255
 Link Type: **StubNet**
 Metric : 0
 Priority : Medium
- * Link ID: 192.168.12.2
 Data : 192.168.12.2
 Link Type: **TransNet**
 Metric : 1
- * Link ID: 1.1.1.1
 Data : 192.168.23.2
 Link Type: **Virtual**

链接类型（并非 OSPF 定义的四种网络类型），Router LSA 描述的链接类型主要有：

Point-to-Point：描述一个从本路由器到邻居路由器之间的点到点链接，属于拓扑信息。

TransNet：描述一个从本路由器到一个 Trans 网段（例如 MA 网段或者 NBMA 网段）的链接，属于拓扑信息。

StubNet：描述一个从本路由器到一个 Stub 网段（例如 Loop back 接口）的链接，属于路由信息。

Virtual : 虚链路，拓扑信息

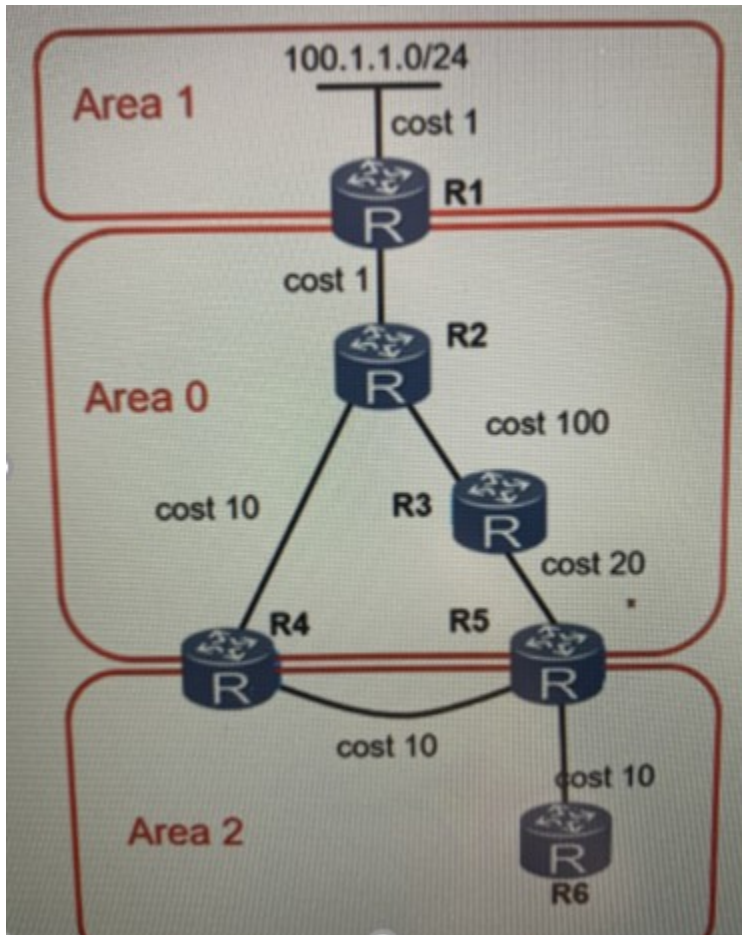
OSPF 选路规则 :

区域内的 > 区域间的 > TYPE 1 > TYPE2

避免域间路由环路

为防止区域间的环路 OSPF 定义了骨干区域和非骨干区域和三类 LSA 的传递规则。

1. OSPF 划分了骨干区域和非骨干区域，所有非骨干区域均直接和骨干区域相连且骨干区域只有一个，非骨干区域之间的通信都要通过骨干区域中转，骨干区域 ID 固定为 0。
2. OSPF 规定从骨干区域传来的三类 LSA 不再传回骨干区域。



R5 学习到的 100.1.1.0 路由，是从 R4 学，还是从 R3 学？ 答：
R3

R6 学习到的 100.1.1.0 路由的 cost 值是多少？ 答：132

如果 R3 R2 之间的连接断开了，R5 会从 R4 学 100.1.1.0 路由吗？ 答：不会

两台设备 R1 192.168.1.1, R2 192.168.1.2

R1 发送的 DD 报文，I, M, MS

一开始发的 DD 报文是空的，没有载荷，用来协商主从 mtu 用的

mtu 默认不携带在 DD 报文，

I=1 这个一个初始报文

M=1 后面还有其他的报文

MS=1 我是主张自己是 master

28	29.703000	192.168.12.2	224.0.0.5	OSPF	Hello Packet
29	29.719000	192.168.12.1	192.168.12.2	OSPF	DB Description
30	29.750000	192.168.12.2	192.168.12.1	OSPF	DB Description
31	29.766000	192.168.12.1	192.168.12.2	OSPF	DB Description
32	29.797000	192.168.12.2	192.168.12.1	OSPF	DB Description
33	29.797000	192.168.12.2	224.0.0.5	OSPF	LS Update
34	29.813000	192.168.12.1	192.168.12.2	OSPF	LS Request

```
OSPF DB Description
Interface MTU: 0
Options: 0x02 (E)
  0... .. = DN: DN-bit is NOT set
  .0.. .. = O: O-bit is NOT set
  ..0. .... = DC: Demand Circuits are NOT supported
  ...0 .... = L: The packet does NOT contain LLS data block
  .... 0... = NP: NSSA is NOT supported
  .... .0.. = MC: NOT Multicast Capable
  .... ..1. = E: External Routing Capability
  .... ...0 = MT: NO Multi-Topology Routing
DB Description: 0x07 (I, M, MS)
  .... 0... = R: OOBResync bit is NOT set
  .... .1.. = I: Init bit is SET
  .... ..1. = M: More bit is SET
  .... ...1 = MS: Master/slave bit is SET
DD Sequence: 170
```

计时器

int g0/0/0

ospf timer poll 130

缺省情况下，时间间隔为 120 秒。

用来配置 NBMA 网络上发送轮询 Hello 报文的时间间隔。当邻居失效后，将按 ospf timer poll 设置的轮询时间间隔定期地发送 Hello 报文。轮询时间间隔值至少应为 Hello 报文时间间隔的 4 倍。

```
int g0/0/0
```

```
ospf timer retransmit 6
```

缺省情况下，时间间隔为 5 秒。

用来设置接口重传 LSA 的时间间隔。

当一台设备向它的邻居发送一条 " 链路状态广播 " (LSA) 后，需要等到对方的确认报文。若在该重传 LSA 的时间间隔内未收到对方的确认报文，就会重传这条 LSA。相邻设备重传 LSA 时间间隔的值不要设置得太小，否则将会引起不必要的重传。

```
int g0/0/0
```

```
ospf trans-delay 2
```

缺省情况下，传输延迟时间为 1 秒。

用来设置接口上发送 LSA 过程中的传输延迟时间。

LSA 在本设备的链路状态数据库 (LSDB) 中会随时间老化 (每秒钟加 1)，但在网络的传输过程中却不会，所以有必要在发送之前在 LSA 的老化时间上增加本命令所设置的一段时间。此配置对低速率的网络尤其重要。

区域标识

Option：其中重点参考 Option 中的 N 和 E bit 位，
N/E 必须置为相同 (N/E 主要表示两端路由器在什么样的区域)

Option 字段是在 OSPF hello 报文中，如果 N / E 不同，是不能建立邻居关系的

```

Open Shortest Path First
> OSPF Header
  v OSPF Hello Packet
    Network Mask: 255.255.255.0
    Hello Interval [sec]: 10
    v Options: 0x02, (E) External Routing
      0... .... = DN: Not set
      .0.. .... = O: Not set
      ..0. .... = (DC) Demand Circuits: Not supported
      ...0 .... = (L) LLS Data block: Not Present
      .... 0... = (N) NSSA: Not supported
      .... .0.. = (MC) Multicast: Not capable
      .... ..1. = (E) External Routing: Capable
      .... ...0 = (MT) Multi-Topology Routing: No
    Router Priority: 1

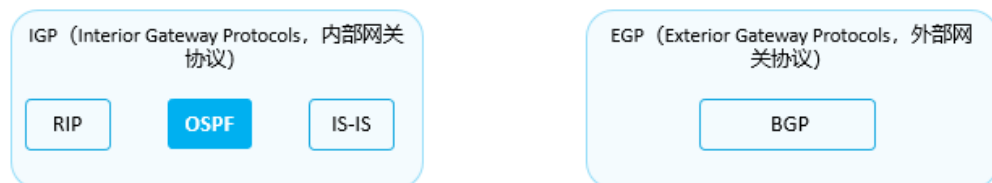
```

N=0 E=1 : 普通区域
 N=0 E=0 : STUB 区域
 N=1 E=0 : NSSA

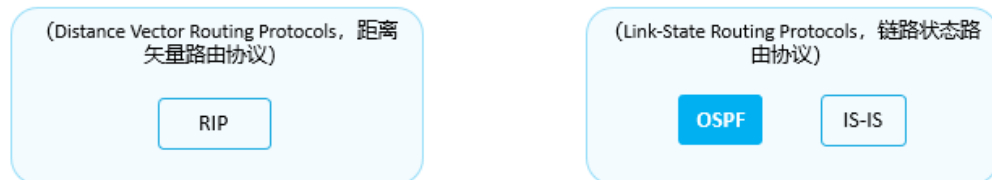
- 路由器根据路由表转发数据包，路由表项可通过手动配置和动态路由协议生成。
- 静态路由比动态路由使用更少的带宽，并且不占用 CPU 资源来计算和分析路由更新。当网络结构比较简单时，只需配置静态路由就可以使网络正常工作。但是当网络发生故障或者拓扑发生变化后，静态路由不会自动更新，必须手动重新配置。
- 相比较于静态路由，动态路由协议具有更强的可扩展性，具备更强的应变能力。
- OSPF (Open Shortest Path First，开放式最短路径优先) 具有扩展性强，收敛速度快等特点，作为优秀的内部网关协议被广泛使用。
- 本课程主要介绍 OSPF 的基本概念、OSPF 邻接关系的建立、以及 OSPF 的基本配置。

动态路由协议的分类

按工作区域分类



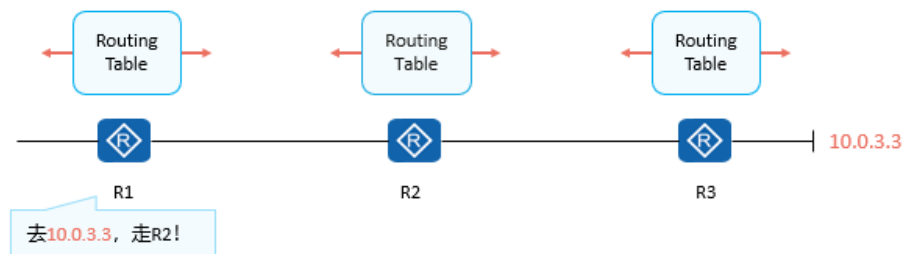
按工作机制及算法分类



- BGP 使用一种基于距离矢量算法修改后的算法，该算法被称为路径矢量 (Path Vector) 算法。因此在某些场合下，BGP 也被称为路径矢量路由协议。
- IGP (内部网关协议) 在同一个自治系统内交换路由信息，IGP 的主要目的是发现和计算自治域内的路由信息。
- EGP (外部网关协议) 主要用于 AS (自治系统) 之间的互联。

距离矢量路由协议

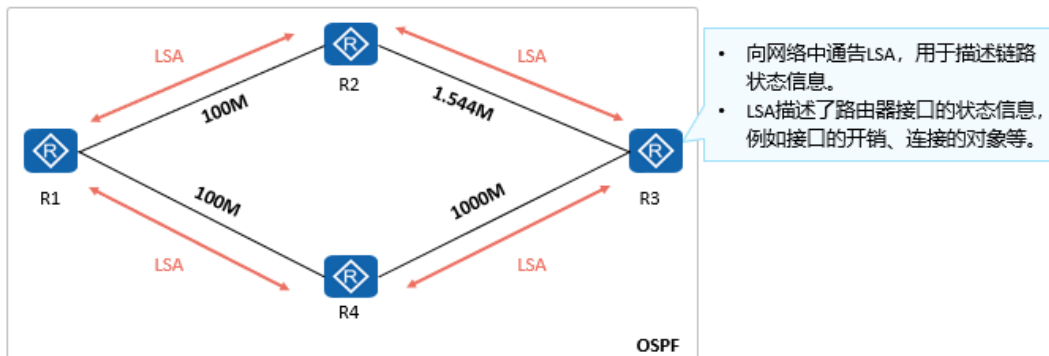
- 运行距离矢量路由协议的路由器周期性地泛洪自己的路由表。通过路由的交互，每台路由器都从相邻的路由器学习到路由，并且加载进自己的路由表中，然后再通告给其他相邻路由器。
- 对于网络中的所有路由器而言，路由器并不清楚网络的拓扑，只是简单的知道要去往某个目的网段方向在哪里，开销有多大。





链路状态路由协议 - LSA泛洪

- 链路状态路由协议通告的是链路状态而不是路由信息。
- 运行链路状态路由协议的路由器之间首先会建立邻居关系，然后彼此之间开始交互LSA（Link State Advertisement，链路状态通告）。

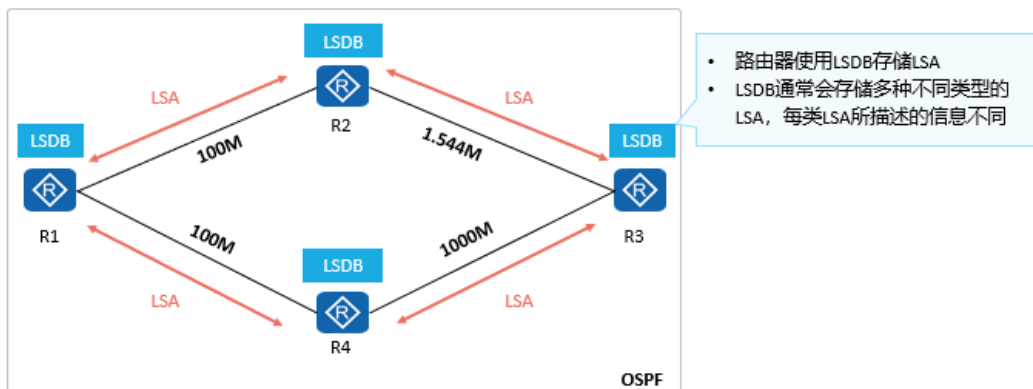


- 链路状态通告，可以简单的理解为每台路由器都产生一个描述自己直连接口状态（包括接口的开销、与邻居路由器之间的关系等）的通告。



链路状态路由协议 - LSDB维护

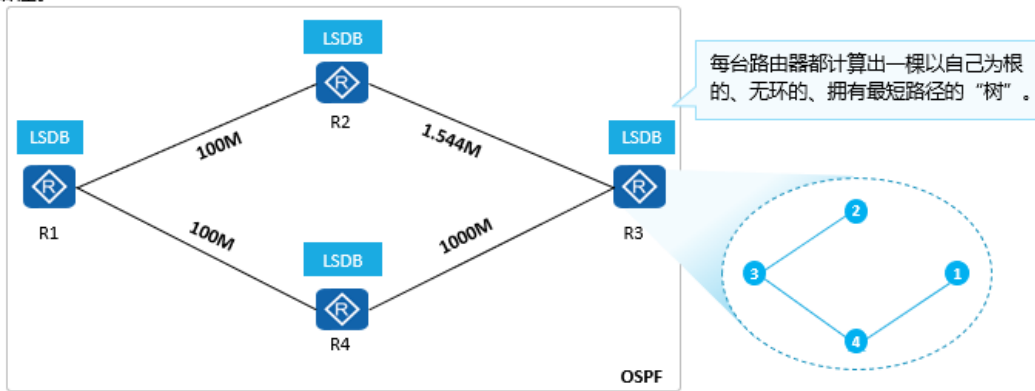
每台路由器都会产生LSA，路由器将接收到的LSA放入自己的LSDB（Link State DataBase，链路状态数据库）。路由器通过对LSDB中所存储的LSA进行解析，进而了解全网拓扑。





链路状态路由协议 - SPF计算

每台路由器基于LSDB，使用SPF（Shortest Path First，最短路径优先）算法进行计算。每台路由器都计算出一棵以自己为根的、无环的、拥有最短路径的“树”。有了这棵“树”，路由器就已经知道了到达所有网段的优选路径。

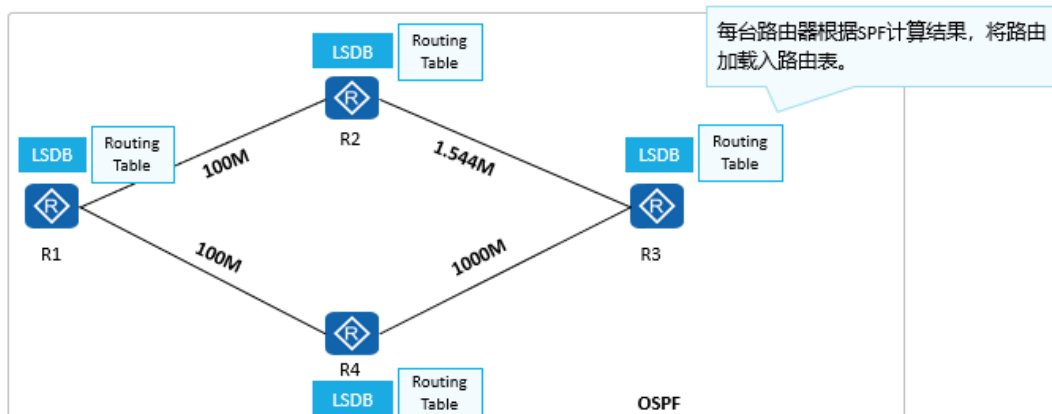


- SPF 是 OSPF 路由协议的一个核心算法，用来在一个复杂的网络中做出路由优选的决策。



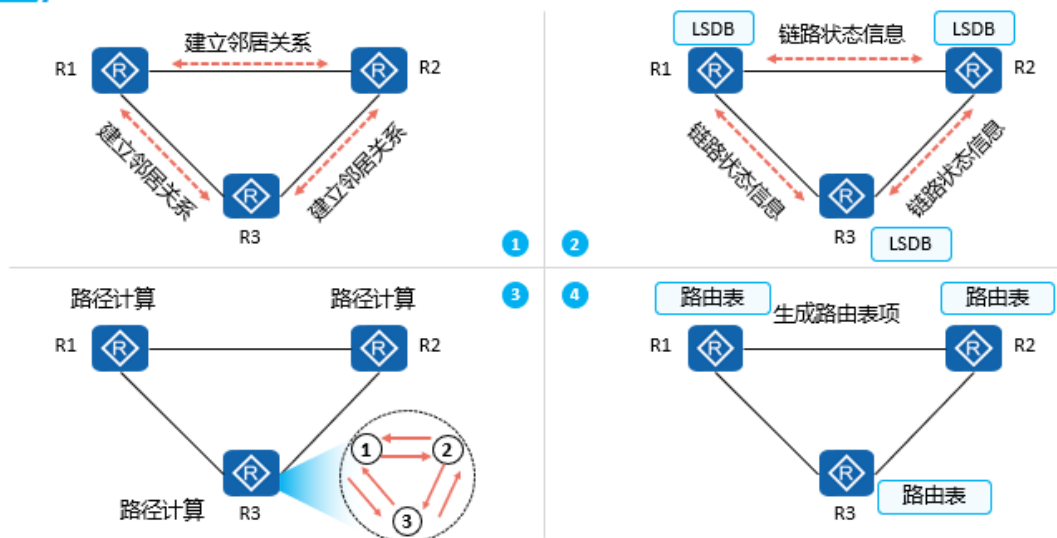
链路状态路由协议 - 路由表生成

路由器将计算出来的优选路径，加载进自己的路由表（Routing Table）。





链路状态路由协议总结



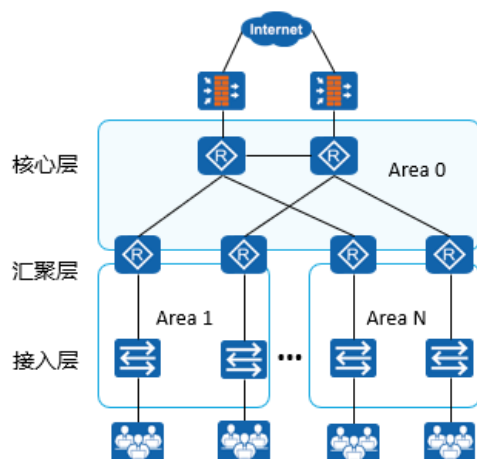
- 链路状态路由协议有四个步骤：
- 第一步是建立相邻路由器之间的邻居关系。
- 第二步是邻居之间交互链路状态信息和同步 LSDB。
- 第三步是进行优选路径计算。
- 第四步是根据最短路径树生成路由表项加载到路由表。



OSPF概述

- OSPF是IETF定义的一种基于链路状态的内部网关路由协议。目前针对IPv4协议使用的是OSPF Version 2 (RFC2328)；针对IPv6协议使用OSPF Version 3 (RFC2740)。
- OSPF有以下优点：
 - 基于SPF算法，以“累计链路开销”作为选路参考值
 - 采用组播形式收发部分协议报文
 - 支持区域划分
 - 支持对等价路由进行负载分担
 - 支持报文认证

OSPF应用场景



大型企业网络中通常部署OSPF实现各个楼宇的网络之间的路由可达。

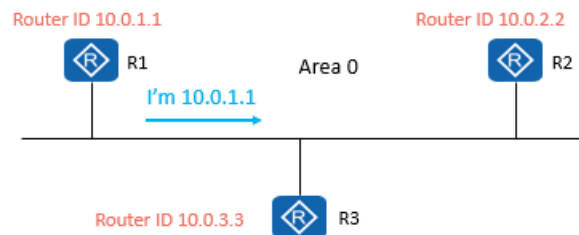
- 核心和汇聚层部署在OSPF骨干区域。
- 接入和汇聚层部署在OSPF非骨干区域。

- **接入层**：接入层利用光纤、双绞线、同轴电缆、无线接入技术等传输介质，实现与用户连接，并进行业务和带宽的分配。接入层目的是允许终端用户连接到网络，因此接入层交换机具有低成本和高端口密度特性。
- **汇聚层**：汇聚层为接入层提供基于策略的连接，如地址合并，协议过滤,路由服务，认证管理等。通过网段划分实现与网络隔离，可以防止网络故障蔓延和影响到核心层。汇聚层同时也可以提供接入层虚拟网之间的互连，控制和限制接入层对核心层的访问，保证核心层的安全和稳定。
- **核心层**：核心层的功能主要是实现骨干网络之间的优化传输，核心层任务的重点通常是冗余能力、可靠性和高速的传输。



OSPF基础术语 - Router ID

- Router ID用于在自治系统中唯一标识一台运行OSPF的路由器，它是一个32位的无符号整数。
- Router ID选举规则如下：
 - 手动配置OSPF路由器的Router ID（建议手动配置）
 - 如果没有手动配置Router ID，则路由器使用Loopback接口中最大的IP地址作为Router ID
 - 如果没有配置Loopback接口，则路由器使用物理接口中最大的IP地址作为Router ID

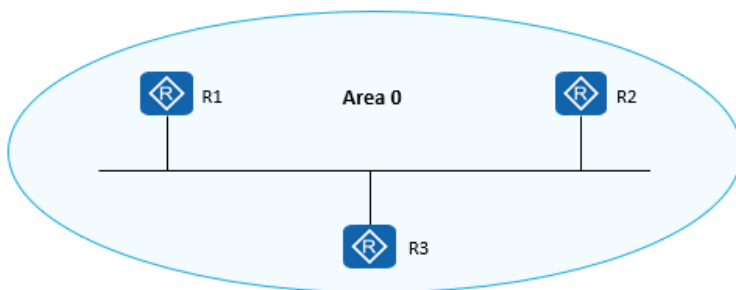


- Router ID 一旦选定，之后如果要更改的话就需要重启 OSPF 进程。
- 在实际工程中，推荐手工指定 OSPF 路由设备的 Router ID。首先规划出一个私有网段用于 OSPF 的 Router ID 选择，例如：192.168.1.0/24。在启用 OSPF 进程前在每个 OSPF 路由器上建立一个 Loopback 接口，使用一个 32 位掩码的私有地址作为其 IP 地址，这个 32 位的私有地址即作为该路由设备的 Router ID。如果没有特殊要求，这个 Loopback 接口地址可以不发布在 OSPF 网络中。



OSPF基础术语 - 区域

- OSPF Area用于标识一个OSPF的区域。
- 区域是从逻辑上将设备划分为不同的组，每个组用区域号（Area ID）来标识。
- OSPF的区域ID是一个32bit的非负整数，按点分十进制的形式（与IPv4地址的格式一样）呈现，例如Area0.0.0.1。为了简便起见，我们也会采用十进制的形式来表示。



- 几个例子：Area0.0.0.1 等同于 Area1，Area0.0.0.255 等同于 Area255，Area0.0.1.0 等同于 Area256。许多网络厂商的设备同时支持这两种区域 ID 配置及表示方式。



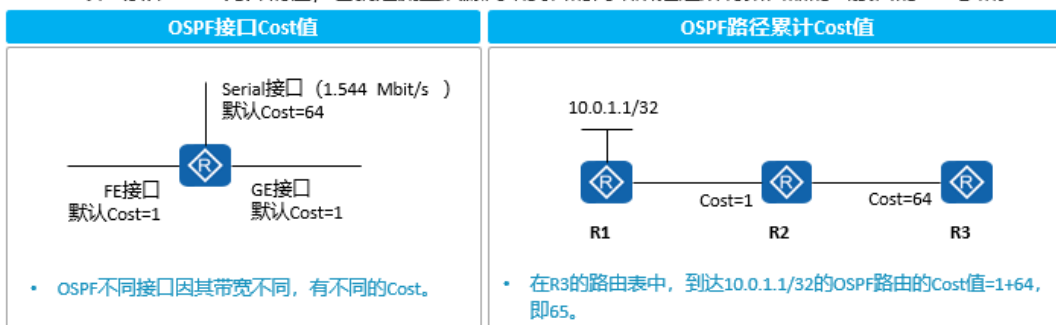
OSPF基础术语 - 度量值

- OSPF使用Cost（开销）作为路由的度量值。每一个激活了OSPF的接口都会维护一个接口Cost值，缺省的接

$$\text{Cost} = \frac{100 \text{ Mbit/s}}{\text{接口带宽}}$$

其中100 Mbit/s为OSPF指定的缺省参考值，该值是可配置的。

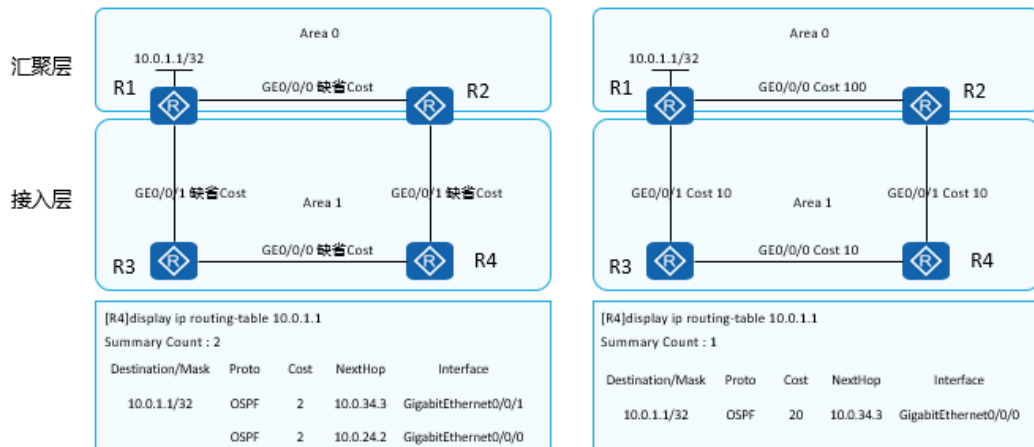
- OSPF以“累计cost”为开销值，也就是流量从源网络到目的网络所经过所有路由器的出接口的cost总和。



- 在实际应用中，推荐根据接口带宽大小手动配置 Cost 值，而不是修改 OSPF 参考带宽。



OSPF基础术语 - 度量值修改举例



缺省时，R4访问10.0.1.1/32存在两条路径，数据转发路径不可控。

如图所示修改设备接口Cost值，可确保接入层路由器访问R1时，流量无需绕行R2。

- 在规划流量路径时，推荐将汇聚层直连链路 Cost 值 > 接入环所有链路 Cost 值之和，这样可以保证从接入层访问 R1 或 R2 时，流量可通过接入层路由器直达 R1 或 R2。
- 以上图为例。R1 和 R2 位于企业网中的汇聚层，R1 和 R2 的直连链路属于区域 0，R1 有一个直连网段 10.0.1.1/32 通过区域 0。
- 在缺省情况下，R4 到达 10.0.1.1/32 的路由存在两个下一跳。
- 修改 Cost 之后，R4 到达 10.0.1.1 的路由仅存在一个下一跳。



OSPF三大表项 - 邻居表

OSPF有三张重要的表项，OSPF邻居表、LSDB和OSPF路由表。对于OSPF的邻居表，需要了解：

- OSPF在传递链路状态信息之前，需先建立OSPF邻居关系。
- OSPF的邻居关系通过交互Hello报文建立。
- OSPF邻居表显示了OSPF路由器之间的邻居状态，使用display ospf peer查看。

[R1]display ospf peer

Router ID:10.0.1.1



R1

GE1/0/0

10.0.12.1/30

Router ID:10.0.2.2



R2

GE1/0/0

10.0.12.2/30

```
<R1> display ospf peer
OSPF Process 1 with Router ID 10.0.1.1
Neighbors
Area 0.0.0.0 interface 10.0.12.1(GigabitEthernet1/0/0)'s neighbors
Router ID: 10.0.2.2 Address: 10.0.12.2 GR State: Normal
State: Full Mode:Nbr is Master Priority: 1
DR: 10.0.12.1 BDR: 10.0.12.2 MTU: 0
Dead timer due in 35 sec
Retrans timer interval: 5
Neighbor is up for 00:00:05
Authentication Sequence: [ 0 ]
```



OSPF三大表项 - LSDB

对于OSPF的LSDB，需要了解：

- LSDB会保存自己产生的及从邻居收到的LSA信息，本例中R1的LSDB包含了三条LSA。
- Type标识LSA的类型，AdvRouter标识发送LSA的路由器。
- 使用命令行display ospf lsdb查看LSDB表。

[R1]display ospf lsdb

Router ID:10.0.1.1



R1

GE1/0/0

10.0.12.1/30

Router ID:10.0.2.2



R2

GE1/0/0

10.0.12.2/30

```
<R1> display ospf lsdb
OSPF Process 1 with Router ID 10.0.1.1
Link State Database
Area: 0.0.0.0
```

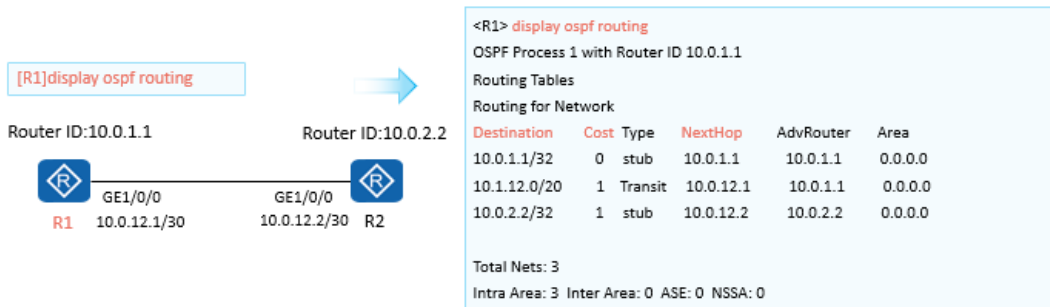
Type	LinkStateID	AdvRouter	Age	Len	Sequence	Metric
Router	10.0.2.2	10.0.2.2	98	36	8000000B	1
Router	10.0.1.1	10.0.1.1	92	36	80000005	1
Network	10.0.12.2	10.0.2.2	98	32	80000004	0



OSPF三大表项 - OSPF路由表

对于OSPF的路由表，需要了解：

- OSPF路由表和路由器路由表是两张不同的表。本例中OSPF路由表有三条路由。
- OSPF路由表包含Destination、Cost和NextHop等指导转发的信息。
- 使用命令display ospf routing查看OSPF路由表。

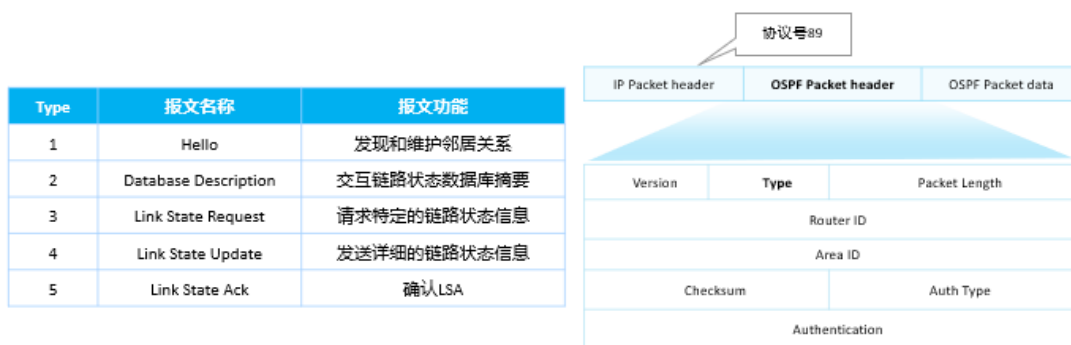


- 路由器路由表通常称为全局路由表，并非所有 OSPF 路由都可以放到路由器路由表。



OSPF报文格式和类型

- OSPF一共定义了5种类型的报文，不同类型的OSPF报文有相同的头部格式。
- OSPF报文直接采用IP封装，在报文的IP头部中，协议号为89。

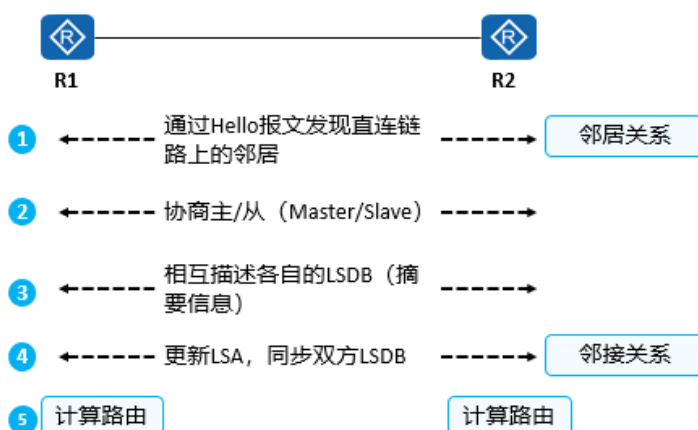


- 重要字段解释
- Version：对于当前所使用的 OSPFv2，该字段的值为 2。
- Router ID：表示生成此报文的路由器的 Router ID。
- Area ID：表示此报文需要被通告到的区域。
- Type：类型字段。
- Packet length：表示整个 OSPF 报文的长度，单位是字

节。

- Checksum：校验字段，其校验的范围是整个 OSPF 报文，包括 OSPF 报文头部。
- Auth Type：为 0 时表示不认证；为 1 时表示简单的明文密码认证；为 2 时表示加密（MD5）认证。
- Authentication：认证所需的信息。该字段的内容随 Auth type 的值不同而不同。

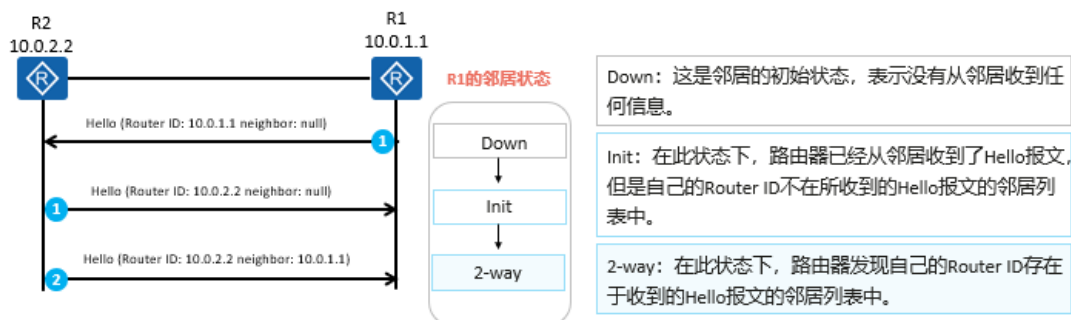
OSPF工作过程概览



步骤1-4由双方交互完成，步骤5独立完成。

建立邻居关系

- OSPF使用Hello报文发现和建立邻居关系。
- 在以太网链路上，缺省时，OSPF采用组播的形式发送Hello报文（目的地址224.0.0.5）。
- OSPF Hello报文中包含了路由器的Router ID、邻居列表等信息。



- R1 和 R2 路由器相互发送 Hello 报文，第一个 Hello 报文包含的邻居列表为空。
- R2 收到 R1 发送的 Hello 报文后，如果各项参数匹配，再次发送 Hello 报文时，将 R1 加入自己的邻居列表。
- 在以太网链路上，通常以组播形式发送 Hello 报文：
- 224.0.0.5 的组播地址为 OSPF 设备的预留 IP 组播地址。
- 224.0.0.6 的组播地址为 OSPF DR/BDR 的预留 IP 组播地址。
- 对于不支持组播的链路，OSPF 支持采用单播的方式发送 Hello 报文。



Hello报文

- Hello报文的主要作用：

- 邻居发现：自动发现邻居路由器。
- 邻居建立：完成Hello报文中的参数协商，建立邻居关系。
- 邻居保持：通过周期性发送和接收，检测邻居运行状态。

Network Mask		
Hello Interval	Options	Router Priority
RouterDeadInterval		
Designated Router		
Backup Designated Router		
Neighbor		
—		

- 重要字段解释

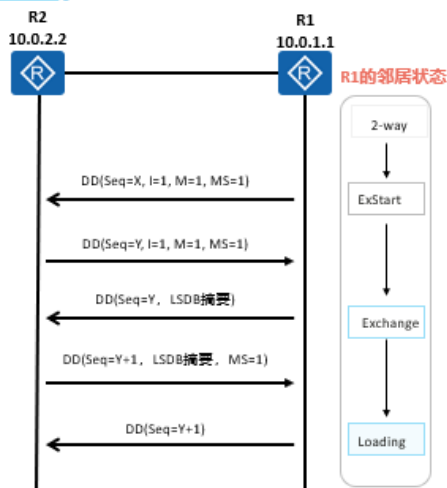
- Network Mask：发送Hello报文的接口的网络掩码。
- HelloInterval：发送Hello报文的时间间隔。通常为10s。
- RouterDeadInterval：失效时间。如果在此时间内未收到邻居发来的Hello报文，则认为邻居失效。通常为40s。
- Neighbor：邻居，以Router ID标识。

- 其它字段解释

- Options：
 - E：是否支持外部路由
 - MC：是否支持转发组播数据包
 - N/P：是否为NSSA区域
- Router Priority：DR优先级。默认为1。如果设置为0，则路由器不能参与DR或BDR的选举。
- Designated Router：DR的接口地址。
- Backup Designated Router：BDR的接口地址。



邻接关系建立 (1)



- DD报文部分字段解释

- I：当发送连续多个DD报文时，如果这是第一个DD报文，则置为1，否则置为0。
- M (More)：当发送连续多个DD报文时，如果这是最后一个DD报文，则置为0。否则置为1，表示后面还有其他的DD报文。
- MS (Master/Slave)：当两台OSPF路由器交换DD报文时，首先需要确定双方的主从关系，Router ID大的一方会成为Master，当值为1时表示发送方为Master。
- DD sequence number：DD报文序列号。主从双方利用序列号来保证DD报文传输的可靠性和完整性。

ExStart：邻居状态变成此状态以后，路由器开始向邻居发送DD报文。在此状态下发送的DD报文不包含链路状态描述。

Exchange：在此状态下，路由器与邻居之间相互发送包含链路状态信息摘要的DD报文。

Loading：在此状态下，路由器与邻居之间相互发送LSR报文、LSU报文、LSAck报文。

- R1 和 R2 的 Router ID 分别为 10.0.1.1 和 10.0.2.2 并且二者已建立了邻居关系。当 R1 的邻居状态变为 ExStart 后，R1 会发送第一个 DD 报文。此报文中，M-bit 设置为 1，表示后续还有 DD 报文要发送，MS-bit 设置为 1，表示 R1 宣告自己为 Master。DD 序列号被随机设置为 X，I-bit 设置为 1，表示这是第一个 DD 报文。
- 同样当 R2 的邻居状态变为 ExStart 后，R2 也会发送第一个 DD 报文。此报文中，DD 序列号被随机设置为 Y (I-bit=

1, M-bit=1, MS-bit=1, 含义同上)。由于 R2 的 Router ID 较大, 所以 R2 将成为真正的 Master。收到此报文后, R1 会产生一个 Negotiation-Done 事件, 并将邻居状态从 ExStart 变为 Exchange。

- 当 R1 的邻居状态变为 Exchange 后, R1 会发送一个新的 DD 报文, 此报文中包含了 LSDB 的摘要信息, 序列号设置为 R2 在步骤 2 中使用的序列号 Y, I-bit=0, 表示这不是第一个 DD 报文, M-bit=0, 表示这是最后一个包含 LSDB 摘要信息的 DD 报文, MS-bit=0, 表示 R1 宣告自己为 Slave。收到此报文后, R2 将邻居状态从 ExStart 变为 Exchange。

- 当 R2 的邻居状态变为 Exchange 后, R2 会发送一个新的 DD 报文, 此报文包含了 LSDB 的摘要信息。DD 序列号设置为 Y+1, MS-bit=1, 表示 R2 宣告自己为 Master。

- 虽然 R1 不需要发送新的包含 LSDB 摘要信息的 DD 报文, 但是作为 Slave, R1 需要对 Master 发送的每一个 DD 报文进行确认。所以, R1 向 R2 发送一个新的 DD 报文, 序列号为 Y+1, 该报文内容为空。发送完此报文后, R1 产生一个 Exchange-Done 事件, 将邻居状态变为 Loading。R2 收到此报文后, 会将邻居状态变为 Full (假设 R2 的 LSDB 是最新最全的, 不需要向 R1 请求更新)。

DD报文

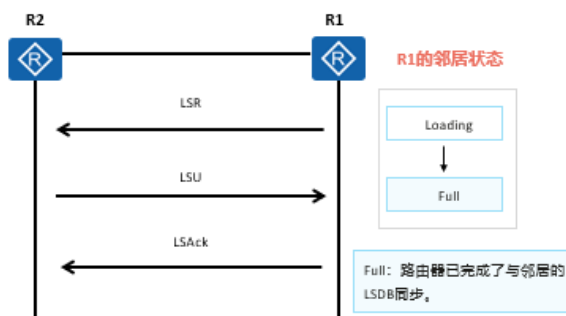
DD报文包含LSA头部信息，包括LS Type、LS ID、Advertising Router、LS Sequence Number、LS Checksum。

Interface MTU	Options	0 0 0 0 1 M M S
DD sequence number		
LSA Header		

• 其他字段解释

- Interface MTU：指示在不分片的情况下，此接口最大可发出的IP报文长度。在两个邻居发送DD报文中包含MTU参数，如果收到的DD报文中MTU和本端的MTU不相等，则丢弃该DD报文。缺省情况下，华为设备未开启MTU检查。
- Options：字段同Hello报文。

邻接关系建立 (2)



- R1开始向R2发送LSR报文，请求那些在Exchange状态下通过DD报文发现的、并且在本地LSDB中没有的链路状态信息。
- R2向R1发送LSU报文，LSU报文中包含了那些被请求的链路状态的详细信息。R1在完成LSU报文的接收之后，且没有其他待请求的LSA后，会将邻居状态从Loading变为Full。
- R1向R2发送LSAck报文，作为对LSU报文的确认。

思考：如果多台路由器处于同一个广播网络，按照如上方式建立邻接关系存在哪些问题？

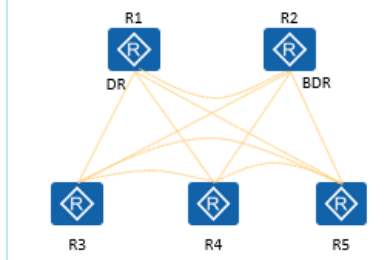


DR与BDR的作用

MA网络中的问题

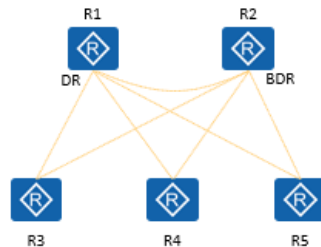
- $n \times (n-1)/2$ 个邻接关系，管理复杂。
- 重复的LSA泛洪，造成资源浪费。

解决方法



在MA网络中选举DR:

- DR (Designated Router, 指定路由器) 负责在MA网络建立和维护邻接关系并负责LSA的同步。
- DR与其他所有路由器形成邻接关系并交换链路状态信息，其他路由器之间不直接交换链路状态信息。
- 为了规避单点故障风险，通过选举BDR (Backup Designated Router, 备份指定路由器)，在DR失效时快速接管DR的工作。

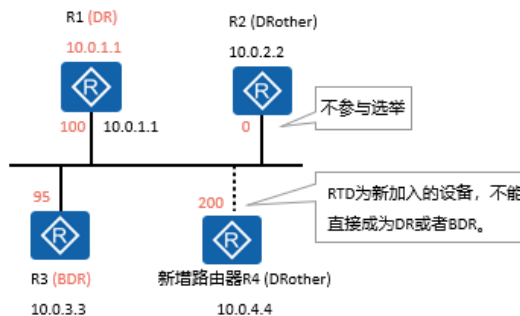


- MA (Multiple Access , 多路访问) 分为 BMA (Broadcast Multi-Access , 广播多路访问) 和 NBMA (Non-Broadcast Multiple Access , 非广播多路访问)。以太网链路组成的网络是典型的 BMA 网络。帧中继链路通过逻辑上的划分组成典型的 NBMA 网络 (注：帧中继相关知识不再介绍)。
- DRoother：既不是 DR 也不是 BDR 的路由器就是 DRoother 路由器。



DR与BDR的选举规则

- DR/BDR的选举是非抢占式的。
- DR/BDR的选举是基于接口的。
 - 接口的DR优先级越大越优先。
 - 接口的DR优先级相等时，Router ID越大越优先。



- 思考：
 - 如果将上图中4台路由器的优先级全部设置为0，OSPF是否可以正常工作？
 - 缺省情况下，哪些链路类型组成的网络是MA网络呢？

- 广播链路或者 NBMA 链路上 DR 和 BDR 的选举过程如下：
- 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 WaitingTimer，该计时器的长度与 DeadTimer 是一样的。默认值为 40 秒，用户不可自行调整。

- 在 WaitingTimer 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。
- 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路由器，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。
- 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 的路由器会竞争成为新的 BDR。
- 只有当不同 Router ID，或者配置不同 DR 优先级的路由器同时起来，在同一时刻进行 DR 选举才会应用 DR 选举规则产生 DR。

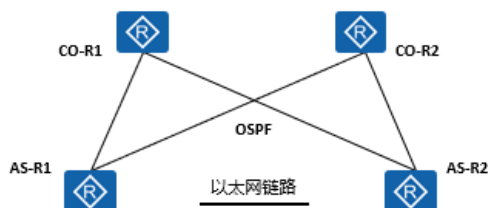
不同网络类型中DR与BDR的选举操作

OSPF网络类型	常见链路层协议	是否选举DR	是否和邻居建立邻接关系
Point-to-point	PPP链路；HDLC链路	否	是
Broadcast	以太网链路	是	DR与BDR、DRother建立邻接关系 BDR与DR、DRother建立邻接关系 DRother之间只建立邻居关系
NBMA	帧中继链路		
P2MP	需手工指定	否	是

- P2MP：point-to-multipoint，点到多点。



可按需调整设备接口的OSPF网络类型



- OSPF的网络类型是根据接口的数据链路层封装自动设置的。
- 图中的路由器采用以太网接口互联，因此这些接口的网络类型缺省均为Broadcast。
- 每段链路实际上都是点对点链路，因此在链路上选举DR与BDR是没有必要的。
- 为了提高OSPF的工作效率，加快邻接关系的建立过程，可以把这些互联接口的网络类型都修改为P2P。

在接口配置视图使用ospf network { p2p | p2mp | broadcast | nbma }即可修改该接口的网络类型

- OSPF 的网络类型是协议根据接口的数据链路层封装自动设置的，在图中，两台处于网络接入层的路由器 AS-R1 及 AS-R2 都通过双链路上联到核心层的路由器 CO-R1 及 CO-R2。四台路由器都在各自接口上激活 OSPF。由于这些路由器都是采用以太网接口互联，因此这些接口的网络类型缺省均为 Broadcast，于是在邻居关系的建立过程中，OSPF 会在每段以太网链路上选举 DR 及 BDR。
- 然而，这实际上是没有必要而且浪费时间的（DR 及 BDR 的选举过程涉及一个等待计时器，这增加了直连路由器形成邻接关系的时间），因为这些链路其实从逻辑的角度看都是点对点的连接，选举 DR 或 BDR 实在是画蛇添足。因此为了提高 OSPF 的工作效率，加快邻接关系的建立过程，可以把这些互联接口的网络类型都修改为 P2P。



配置命令介绍 (1)

1. 启动OSPF进程，进入OSPF视图。

```
[Huawei] ospf [ process-id | Router ID Router ID ]
```

路由器支持OSPF多进程，进程号是本地概念，两台使用不同OSPF进程号设备之间也能够建立邻接关系。

2. 创建并进入OSPF区域视图

```
[Huawei-ospf-1] area area-id
```

3. 在OSPF区域中使能OSPF

```
[Huawei-ospf-1-area-0.0.0.0] network network-address wildcard-mask
```

执行该命令配置区域所包含的网段。设备的接口IP地址掩码长度 \geq network命令指定的掩码长度，且接口的主IP地址必须在network命令指定的网段范围内，此时该接口才会在相应的区域内激活OSPF。

4. 接口视图下使能OSPF：

```
[Huawei-GigabitEthernet1/0/0] ospf enable process-id area area-id
```

ospf enable命令用来在接口上使能OSPF，优先级高于network命令。

- 缺省情况下，OSPF以32位主机路由的方式对外发布Loopback接口的IP地址，与Loopback接口上配置的掩码长度无关。如果要发布Loopback接口的实际网段，需要在接口下配置网络类型为NBMA或广播型。



配置命令介绍 (2)

5. 接口视图下：设置选举DR时的优先级

```
[Huawei-GigabitEthernet1/0/0] ospf dr-priority priority
```

缺省情况下，优先级为1。

6. 接口视图下：设置Hello报文发送的时间间隔

```
[Huawei-GigabitEthernet1/0/0] ospf timer Hello interval
```

缺省情况下，P2P、Broadcast类型接口发送Hello报文的时间间隔的值为10秒，且同一接口上邻居失效时间是Hello间隔时间的4倍

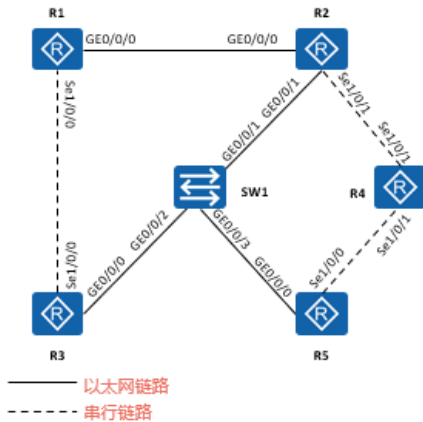
7. 接口视图下：设置网络类型

```
[Huawei-GigabitEthernet1/0/0] ospf network-type { broadcast | nbma | p2mp | p2p }
```

缺省情况下，接口的网络类型根据物理接口而定。以太网接口的网络类型为广播，串口和POS口（封装PPP协议或HDLC协议时）网络类型为P2P。



OSPF配置举例



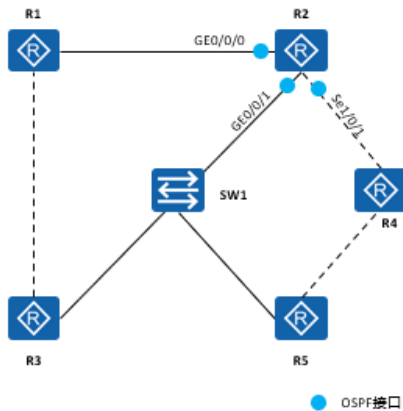
- 基础数据说明：每台设备的Router ID为10.0.x.x，其中x为路由器的编号，例如R5的Router ID为10.0.5.5。设备之间的互联IP地址为10.0.xy(z).x(y/z)/24其中xyz为路由器编号，编号小的放到前面，例如R2 GE0/0/1接口的IP地址为10.0.235.2/24。
- 拓扑说明：五台路由器均工作在区域0。

以R2为例：

```
[R2]ospf 1 router-id 10.0.2.2
[R2-ospf-1]area 0.0.0.0
[R2-ospf-1-area-0.0.0.0] network 10.0.12.0 0.0.0.255
[R2-ospf-1-area-0.0.0.0] network 10.0.24.2 0.0.0.0
[R2-ospf-1-area-0.0.0.0] network 10.0.35.2 0.0.0.0
```



OSPF配置验证 (1)



display ospf interface all可查看当前设备所有激活OSPF的接口信息：

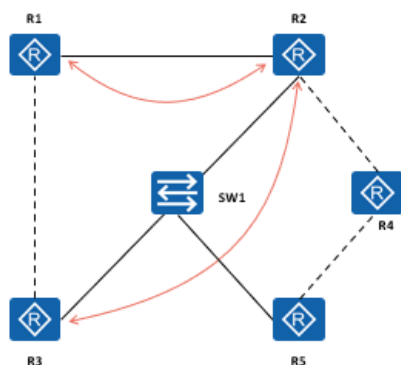
- 时间参数，例如Hello报文发送间隔，死亡时间。
- 接口的链路类型、接口的MTU。
- 对于以太网链路，可查看DR的接口地址，DR的优先级。

```
[R2]display ospf interface all
      OSPF Process 1 with Router ID 10.0.2.2
Area: 0.0.0.0
Interface: 10.0.12.2 (GigabitEthernet0/0/0)
Cost: 1 State: DR Type: Broadcast MTU: 1500 Priority: 1
Designated Router: 10.0.12.2
Backup Designated Router: 10.0.12.1
Timers: HELLO 10, Dead 40, Poll 120, Retransmit 5, Transmit Delay 1

Interface: 10.0.235.2 (GigabitEthernet0/0/1)
Cost: 1 State: DROther Type: Broadcast MTU: 1500 Priority: 1
Designated Router: 10.0.235.5
Backup Designated Router: 10.0.235.3
Timers: HELLO 10, Dead 40, Poll 120, Retransmit 5, Transmit Delay 1

Interface: 10.0.24.2 (Serial1/0/1) -> 10.0.24.4
Cost: 48 State: P-2-P Type: P2P MTU: 1500
Timers: HELLO 10, Dead 40, Poll 120, Retransmit 5, Transmit Delay 1
```

OSPF配置验证 (2)



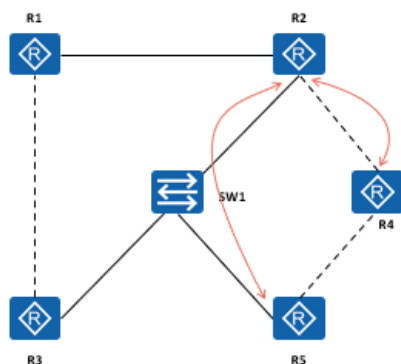
display ospf peer可查看当前设备的邻居状态:

- 邻居路由器的Router ID。
- 邻居状态，例如FULL, TWO-WAY, DOWN等。

```
<R2>display ospf peer
      OSPF Process 1 with Router ID 10.0.2.2
Area 0.0.0.0 interface 10.0.12.2(GigabitEthernet0/0/0)'s neighbors
Router ID: 10.0.1.1    Address: 10.0.12.1
State: Full Mode:Nbr is Slave Priority: 1
DR: 10.0.12.2 BDR: 10.0.12.1 MTU: 0
Dead timer due in 28 sec
Retrans timer interval: 5
Neighbor is up for 00:01:31
Authentication Sequence: [ 0 ]

Area 0.0.0.0 interface 10.0.235.2(GigabitEthernet0/0/1)'s neighbors
Router ID: 10.0.3.3    Address: 10.0.235.3
State: Full Mode:Nbr is Master Priority: 1
DR: 10.0.235.5 BDR: 10.0.235.3 MTU: 0
Dead timer due in 30 sec
Retrans timer interval: 5
Neighbor is up for 00:01:31
Authentication Sequence: [ 0 ]
```

OSPF配置验证 (3)

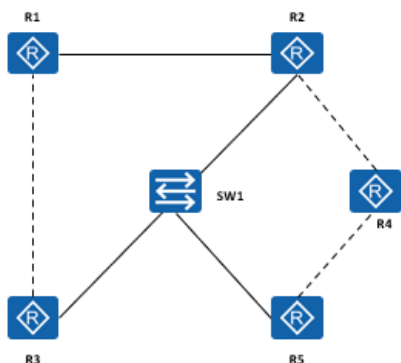


在P2P网络中不需要选举DR/BDR。因此在本例中，查看R2的OSPF邻居表时，会发现其Serial1/0/1接口的数据结构中DR/BDR字段为None。

```
<R2>display ospf peer
      OSPF Process 1 with Router ID 10.0.2.2
Area 0.0.0.0 interface 10.0.235.2(GigabitEthernet0/0/1)'s neighbors
Router ID: 10.0.5.5    Address: 10.0.235.5
State: Full Mode:Nbr is Master Priority: 1
DR: 10.0.235.5 BDR: 10.0.235.3 MTU: 0
Dead timer due in 40 sec
Retrans timer interval: 0
Neighbor is up for 00:01:27
Authentication Sequence: [ 0 ]

Area 0.0.0.0 interface 10.0.24.2(Serial1/0/1)'s neighbors
Router ID: 10.0.4.4    Address: 10.0.24.4
State: Full Mode:Nbr is Master Priority: 1
DR: None BDR: None MTU: 0
Dead timer due in 35 sec
Retrans timer interval: 5
Neighbor is up for 00:01:56
Authentication Sequence: [ 0 ]
```

OSPF配置验证 (4)



display ospf lsdb可查看当前设备的LSDB：

- LSDB由多种类型的LSA构成，所有的LSA都有相同的报文头部格式，其中关键字段如Type、LinkState ID、AdvRouter等。下节课程将重点介绍LSA的详细信息。

```
<R2>display ospf lsdb
```

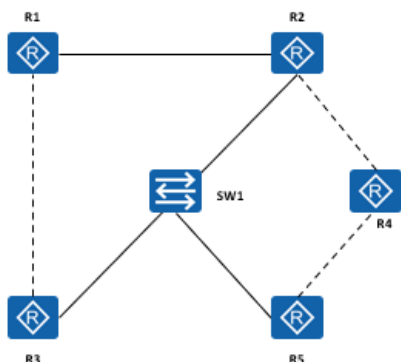
OSPF Process 1 with Router ID 10.0.2.2
Link State Database

Area: 0.0.0.0

Type	LinkState ID	AdvRouter	Age	Len	Sequence	Metric
Router	10.0.4.4	10.0.4.4	662	72	80000006	48
Router	10.0.2.2	10.0.2.2	625	72	8000000C	1
Router	10.0.1.1	10.0.1.1	638	60	80000007	1
Router	10.0.5.5	10.0.5.5	634	60	8000000B	1
Router	10.0.3.3	10.0.3.3	639	60	80000009	1
Network	10.0.235.5	10.0.5.5	634	36	80000005	0
Network	10.0.12.2	10.0.2.2	629	32	80000003	0

在其它设备上查看LSDB，是否完全一致呢？

OSPF配置验证 (5)



display ospf routing可查看当前设备的OSPF路由表：

- 从R2的OSPF路由表可看出，它已经通过OSPF获知到达全网的路由。

```
<R2>display ospf routing
```

OSPF Process 1 with Router ID 10.0.2.2
Routing Tables

Destination	Cost	Type	NextHop	AdvRouter	Area
10.0.12.0/24	1	Transit	10.0.12.2	10.0.2.2	0.0.0.0
10.0.24.0/24	48	Stub	10.0.24.2	10.0.2.2	0.0.0.0
10.0.235.0/24	1	Transit	10.0.235.2	10.0.2.2	0.0.0.0
10.0.13.0/24	49	Stub	10.0.12.1	10.0.1.1	0.0.0.0
10.0.13.0/24	49	Stub	10.0.235.3	10.0.3.3	0.0.0.0
10.0.45.0/24	49	Stub	10.0.235.5	10.0.5.5	0.0.0.0

本章总结：

- 本章简单介绍了 OSPF 的基本概念，内容包括 Router ID、区域、Cost。运行 OSPF 的路由器通过相互发送链路状态信息完成拓扑和路由计算。
- 本章详细介绍了 OSPF 邻居和邻接关系建立的过程。在 MA 网络中，需要选举 DR 和 BDR。OSPF 有 5 种类型的报文，所有报文有相同的报文头部格式。运行 OSPF 的路由器通过周期性的发送 Hello 报文发现和维持邻居关系，通过 DD、LS

R、LSU、LSAck 报文配合完成链路状态数据库的同步。最后介绍了 OSPF 单区域的简单配置。

思考题：

- (单选题) OSPF 使用以下哪种报文维护邻居关系？ ()
- Hello
- Database Description
- LSR
- LSU
- (多选题) OSPF 支持以下哪些网络类型？ ()
- P2P 网络
- P2MP 网络
- 广播网络
- NBMA 网络

答案：

- A
- ABCD
-