OSPF 协议基础

OSPF (Open Shortest Path First) 开放式最短路径优先

OSPF 知识点

OSPF 基本配置，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（虚链路），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 在 FR 中的应用，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。

========

邻居及邻接的区别：

邻居 neighbor---必须有直连的链路

邻接 adjacency--- 1. 必须是邻居, 2. 链路两边同一区域的数据库必须同步(状态为:FULL).

邻接建不起来的原因

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

SPF 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 值不同时会停留在 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 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）

========

OSPF 协议　link-type

OSPF 的有以下几种 LSA:
Type-1 lsa (router lsa)
Type-2 lsa (network lsa)
Type-3 lsa (network summary lsa)
Type-4 lsa (asbr summary lsa)
Type-5 lsa (as external lsa)
Type-7 lsa (nssa external lsa)

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

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 Loopback 接口）的链接，属于路由信息。

LINK 包括:
1.Link-ID
2.Link-type
3.Link-data

第一类 LINK Point to Point link
1.Link -id 邻居的 Router-id
2.Link-type Point to Point
3.Link-data 本路由器在本链路的接口 IP

第二类 LINK Stub network link
1.Link-id 网络地址
2.Link-type Stub network
3.Link-data 子网掩码

第三类 LINK Transmit network link
1.Link-id DR 在本网段接口 IP
2.Link-type Transmit network
3.Link-data 本链路的接口 IP

第四类 LINK Virtual-link link
1.Link-id 邻居 Router-id
2.Link-type Virtual-link
3.Link-data 虚链路所使用的物理口的接口 IP

==========

OSPF 选路规则 ：

区域内的 > 区域间的 > TYPE 1 > TYPE2

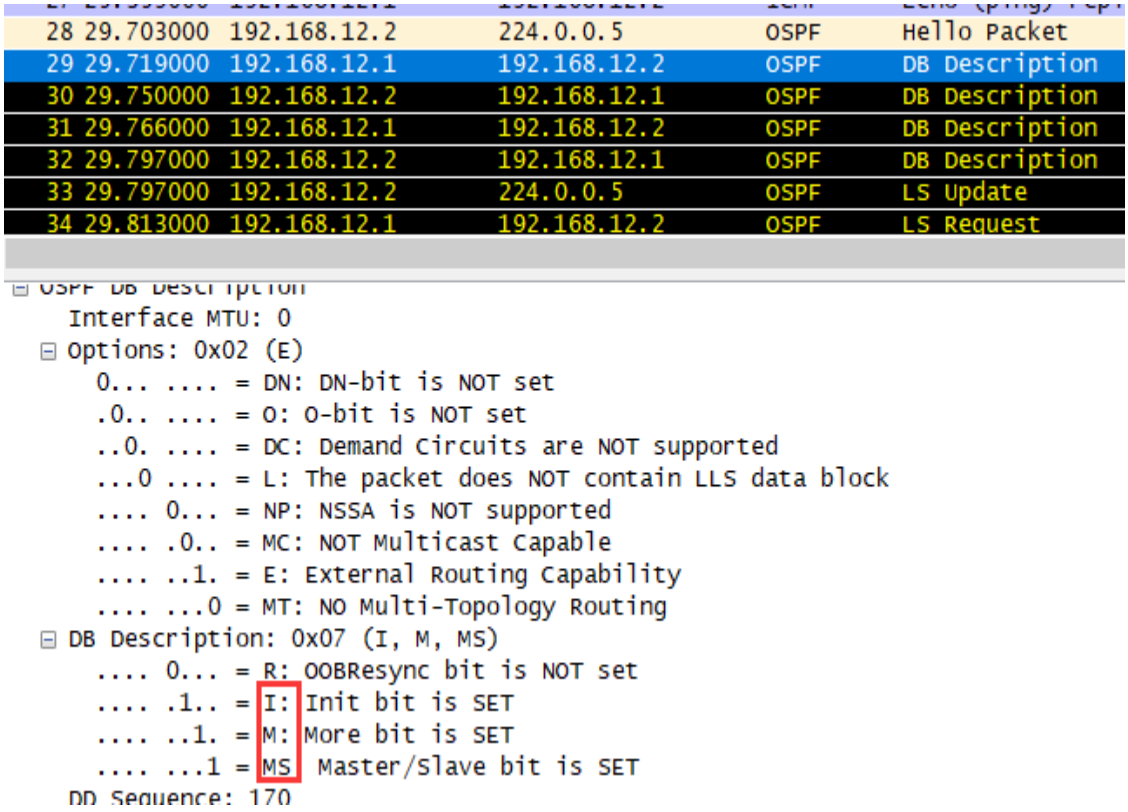
避免域间路由环路

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。

两台设备 R1　192.168.1.1 , R2 192.168.1.2
R1 发送的 DD 报文， I ,M ,MS

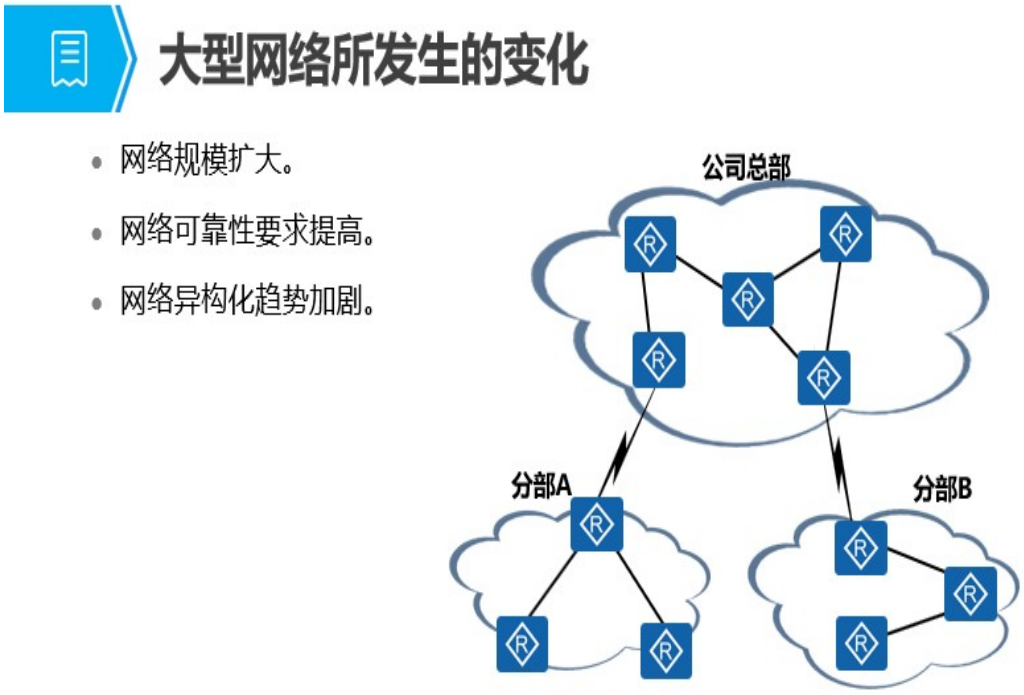
I=1 这个一个初始报文
M=1 后面还有其他的报文
MS=1 我是主张自己是 master



前言

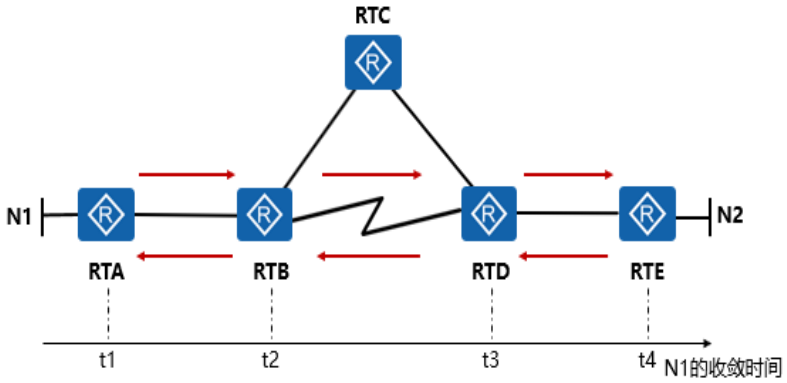
- RIP是基于距离矢量算法的路由协议，应用在大型网络中存在收敛速度慢、度量值不科学、可扩展性差等问题。
- IETF提出了基于SPF算法的链路状态路由协议OSPF（Open Shortest Path First）。通过在大型网络中部署OSPF协议，弥补了RIP协议的诸多不足。那么OSPF协议是如何实现的呢？面对网络扩展的需求，又该如何应对呢？

互联网工程任务组：The Internet Engineering Task Force(IETF)。



- 网络规模扩大：
- 企业新业务层出不穷，且业务呈现大集中趋势，使得网络规模不断扩大。
- 网络可靠性要求提高：
- 各种应用程序对网络可靠性要求越来越高，网络发生故障后，需要在更短的时间内恢复正常。
- 网络异构化，多厂商设备互联需求：
- 在日常的运营维护中，硬件设备不断升级或更新，不同设备之间性能差异较大，设备间互连链路带宽也存在一定的差异。
- 需要一种各厂商均支持的开放路由协议。
- 面对越来越高的要求与挑战，如果通过 RIP 来部署，会遇到什么问题？

RIP在大型网络中部署所面临的问题



RIP特性	带来的问题
逐跳收敛	收敛慢，故障恢复时间长
传闻路由更新机制	缺少对全局网络拓扑的了解
最多有效跳数为15	环形组网中，使远端路由不可达
以“跳数”为度量	存在选择次优路径的风险

- 逐跳收敛：
- 如图所示，N1 网络发生变化，RTA 向 RTB 发出更新，RTB 收到更新之后进行本地计算，完成计算后再向 RTC 发送路由变化通知，如此循环。逐跳收敛的方式，造成了网络收缓慢的问题。
- 传闻路由更新机制：
- RIP 在计算路由完全依赖于从邻居路由器收到的路由信息，RTE 仅依靠从 RTD 获取的信息计算路由，对 RTA、RTB 和 RTC 之间的网络情况并不了解。RIP 在计算路由时，缺少对全局网络拓扑的了解。
- 以“跳数”为度量：
- 因为 RIP 基于跳数的度量方式，所以 N1 与 N2 网络互访时会选择 RTA->RTB->RTD->RTE 作为最优路径。显然 RTB->RTC->RTD 之间的以太链路要比 RTB->RTD 的串行链路带宽要高的多。
- 针对 RIP 遇到的问题，可以通过什么方式优化或者解决？

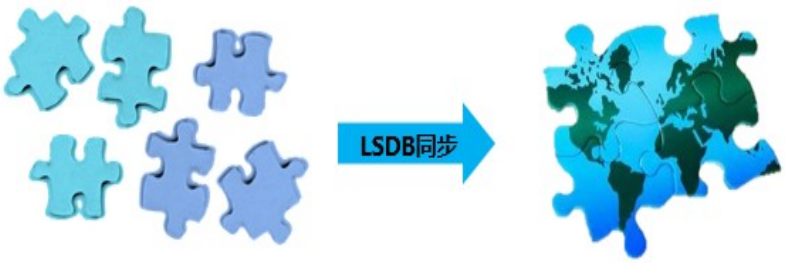
如何解决RIP的问题?

RIP的问题	优化或解决的方式
收敛慢，故障恢复时间长	“收到更新->计算路由->发送更新” 改为“收到更新->发送更新->计算路由”
缺少对全局网络拓扑的了解	路由器基于拓扑信息，独立计算路由
最多有效跳数为15	不限定跳数
存在选择次优路径的风险	将链路带宽作为选路参考值

- 在“收到更新”、“计算路由”、“发送更新”的路由收敛过程中，RIP 的局限性在于路由器需要在完成路由计算之后才可以向邻居发送路由变化通知。如果将这个过程调整为：“收到更新”、“发送更新”、“计算路由”，即路由器从邻居收到路由更新后立刻向其他邻居路由器转发，然后再本地计算新的路由。这样的收敛方式可以大大降低全网路由收敛的时间。
- 因为 RIP 路由器仅从邻居路由器获取路由信息，所以对于非最优或者错误路由信息，RIP 路由器并不能识别或屏蔽。解决此问题的关键最佳方式是路由器收集全网的信息，并基于这些信息独立计算路由。
- 基于跳数的度量方式并没有考虑数据包的链路转发延迟，如果采用以累积带宽为选路参考依据，可以更好的规避选择次优路径的风险。
- 与 RIP 这种距离矢量路由协议不同的链路状态路由协议是以怎样的方式来解决上述问题的呢？

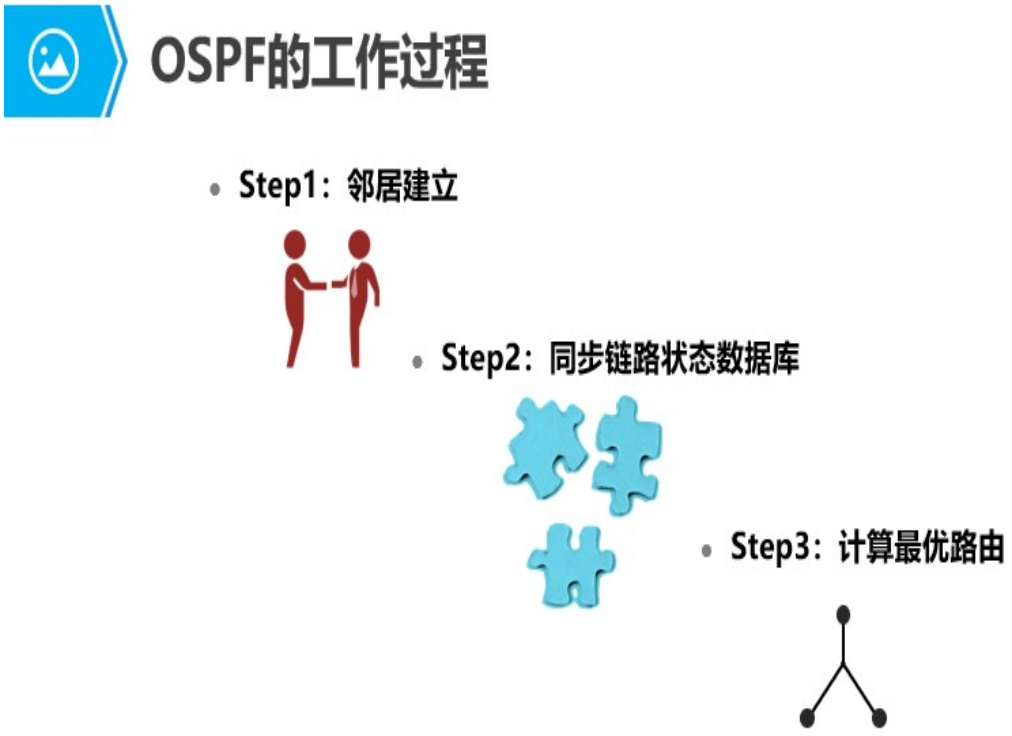
链路状态路由协议OSPF

- 路由信息传递与路由计算分离。
- 基于SPF算法。
- 以“累计链路开销”作为选路参考值。



- 所谓 Link State（链路状态）指的就是路由器的接口状态。在 OSPF 中路由器的某一接口的链路状态包含了如下信息：
- 该接口的 IP 地址及掩码。
- 该接口的带宽。
- 该接口所连接的邻居。
- ……
- OSPF 作为链路状态路由协议，不直接传递各路由器的路由表，而传递链路状态信息，各路由器基于链路状态信息独立计算路由。
- 所有路由器各自维护一个链路状态数据库。邻居路由器间先同步链路状态数据库，再各自基于 SPF（Shortest Path First）算法计算最优路由，从而提高收敛速度。
- 在度量方式上，OSPF 将链路带宽作为选路时的参考依据。“累计带宽”是一种要比“累积跳数”更科学的计算方式。
- RIP 在大型网络中部署所面临的问题，OSPF 都有相对

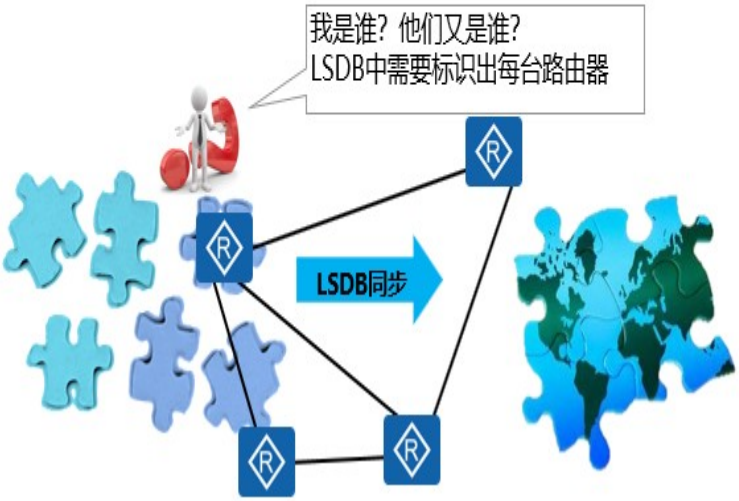
应的解决办法，接下来详细地介绍下 OSPF 的实现过程。



- 企业网络是由众多的路由器、交换机等网络设备之间互相连接组成的，类似一张地图。由于众多不同型号的路由器、不同类型的链路及其连接关系，造成了路由计算的复杂性。
- OSPF 的路由计算过程可以简化描述为：
- 路由器之间发现并建立邻居关系。
- 每台路由器产生并向邻居泛洪链路状态信息，同时收集来自其他路由器链路状态信息，完成 LSDB（Link State Data base）的同步。
- 每台路由器基于 LSDB 通过 SPF 算法，计算得到一棵以自己为根的 SPT（Shortest Path Tree），再以 SPT 为基础计算去往各目的网络的最优路由，并形成路由表。
- 下面我们依照这三个步骤为主线，来学习和掌握 OSPF 的原理与实现。

Router ID

- 用于在自治系统中唯一标识一台运行OSPF的路由器，每台运行OSPF的路由器都有一个Router ID。



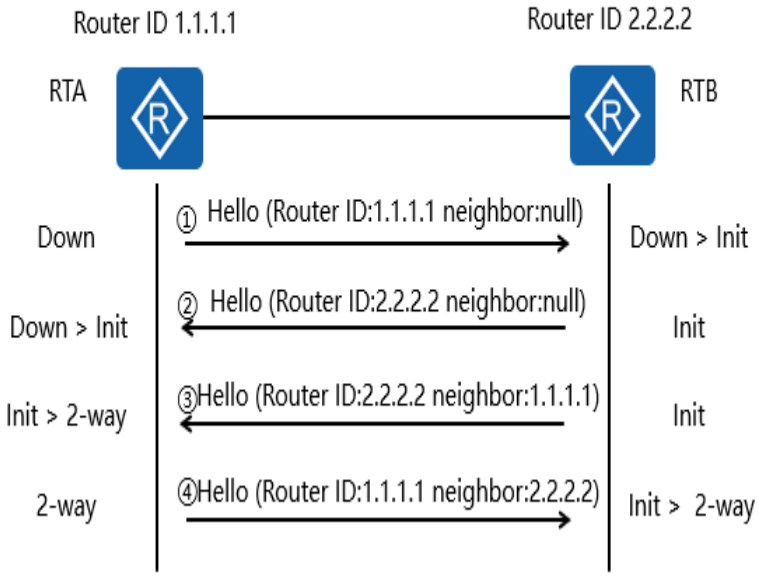
- 企业网中的设备少则几台多则几十台甚至几百台，每台路由器都需要有一个唯一的 ID 用于标识自己。
- Router ID 是一个 32 位的无符号整数，其格式和 IP 地址的格式是一样的，Router ID 选举规则如下：
- 手动配置 OSPF 路由器的 Router ID（通常建议手动配置）；
- 如果没有手动配置 Router ID，则路由器使用 Loopback 接口中最大的 IP 地址作为 Router ID；
- 如果没有配置 Loopback 接口，则路由器使用物理接口中最大的 IP 地址作为 Router ID。
- OSPF 的路由器 Router ID 重新配置后，可以通过重置 OSPF 进程来更新 Router ID。

发现并建立邻居 - Hello报文

- Hello报文的作用：
  - 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Keepalive机制，检测邻居运行状态。

- OSPF 路由器之间在交换链路状态信息之前，首先需要彼此建立邻居关系，通过 Hello 报文实现。
- OSPF 协议通过 Hello 报文可以让互联的路由器间自动发现并建立邻居关系，为后续可达性信息的同步作准备。
- 在形成邻居关系过程中，路由器通过 Hello 报文完成一些参数的协商。
- 邻居关系建立后，周期性的 Hello 报文发送还可以实现邻居保持的功能，在一定时间内没有收到邻居的 Hello 报文，则会中断路由器间的 OSPF 邻居关系。

OSPF邻居建立过程

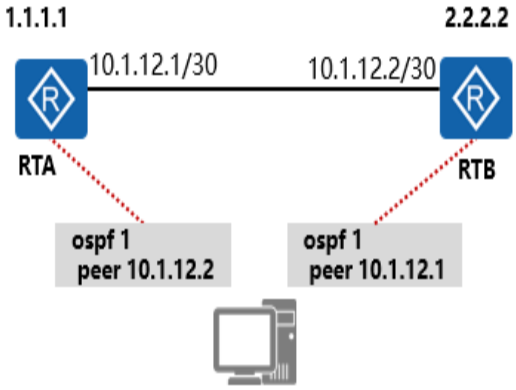


- 状态含义：
- Down：这是邻居的初始状态，表示没有从邻居收到任何信息。
- Init：在此状态下，路由器已经从邻居收到了 Hello 报文，但是自己的 Router ID 不在所收到的 Hello 报文的邻居列表中，表示尚未与邻居建立双向通信关系。
- 2-Way：在此状态下，路由器发现自己的 Router ID 存在于收到的 Hello 报文的邻居列表中，已确认可以双向通信。
- 邻居建立过程如下：
- RTA 和 RTB 的 Router ID 分别为 1.1.1.1 和 2.2.2.2。当 RTA 启动 OSPF 后，RTA 会发送第一个 Hello 报文。此报文中邻居列表为空，此时状态为 Down，RTB 收到 RTA 的这个 Hello 报文，状态置为 Init。
- RTB 发送 Hello 报文，此报文中邻居列表为空，RTA 收到 RTB 的 Hello 报文，状态置为 Init。

- RTB 向 RTA 发送邻居列表为 1.1.1.1 的 Hello 报文，RTA 在收到的 Hello 报文邻居列表中发现自己的 Router ID，状态置为 2-way。
- RTA 向 RTB 发送邻居列表为 2.2.2.2 的 Hello 报文，RTB 在收到的 Hello 报文邻居列表中发现自己的 Router ID，状态置为 2-way。
- 因为邻居都是未知的，所以 Hello 报文的目的 IP 地址不是某个特定的单播地址。邻居从无到有，OSPF 采用组播的形式发送 Hello 报文（目的地址 224.0.0.5）。对于不支持组播的网络，OSPF 路由器如何发现邻居呢？

发现并建立邻居 - 手动建立

- OSPF支持通过单播方式建立邻居关系。
- 对于不支持组播的网络可以通过手动配置实现邻居的发现与维护。



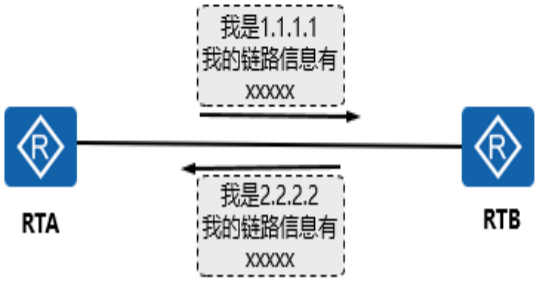
- 对于不支持组播的网络可以通过手动配置实现邻居的发现与维护。
- 当网络规模越来越大或者设备频繁更新，相关联的 OSPF 路由器都需要更改静态配置，手动更改配置的工作量变大且

容易出错。除了特殊场景，一般情况下不适用手动配置的方式。

- OSPF 路由器之间建立邻居关系是为了同步链路状态信息，接下来学习 OSPF 如何实现链路状态数据库的同步。

链路状态信息

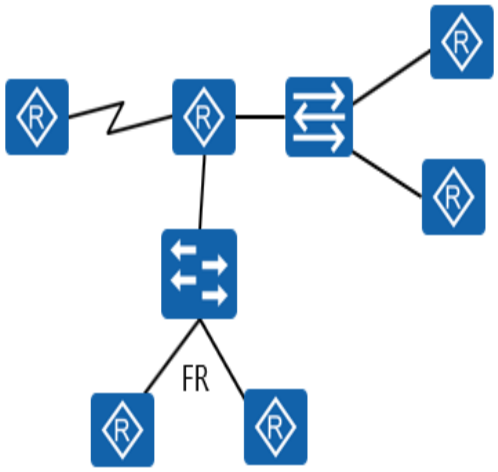
- 链路信息主要包括:
  - 链路的类型;
  - 接口IP地址及掩码;
  - 链路上所连接的邻居路由器;
  - 链路的带宽 (开销) 。



- 区别于 RIP 路由器之间交互的路由信息，OSPF 路由器同步的是最原始的链路状态信息，而且对于邻居路由器发来的链路状态信息，仅作转发。最终所有路由器都将拥有一份相同且完整的原始链路状态信息。
- 每台运行 OSPF 协议的路由器所描述的信息中都应该包括链路的类型、接口 IP 地址及掩码、链路上的邻居、链路的开销等信息。
- 路由器只需要知道目的网络号/掩码、下一跳、开销（接口 IP 地址及掩码、链路上的邻居、链路的开销）即可，为什么要有链路的类型呢？

丰富的数据链路层支持能力

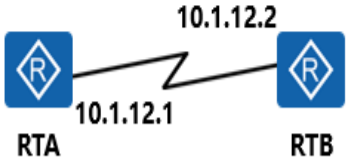
- 数据链路层协议类型多种多样，工作机制也各不相同。
- 为适配多种数据链路层协议，必须考虑各类链路层协议在组网时的应用场景。



- 网络技术的发展包含了设备、链路以及通信协议的发展。设备性能日趋提高，互联链路也从串行链路、ATM、帧中继发展到当前的以太网、xPON、SDH、MSTP、OTN 等。技术升级不是一蹴而就的，而是一个循序渐进的过程。各种不同的物理链路各具特点，也正因为如此，一个成熟的路由协议必须能够根据不同物理链路特性进行适配。
- 下面将介绍 OSPF 是如何定义多种网络的。

网络类型 - P2P网络

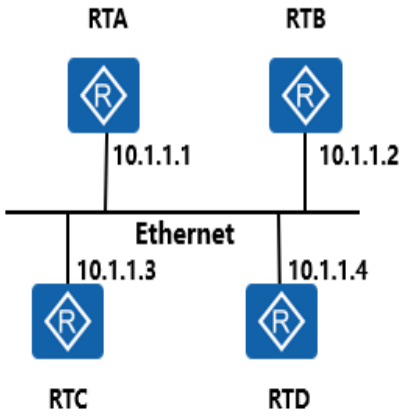
- 仅两台路由互连。
- 支持广播、组播。



- OSPF 划分了四种网络类型并以此来组成拓扑信息的一部分。
- P2P 网络连接了一对路由器，广播、组播数据包都可以转发。
- P2P 网络的例子：两台通过 PPP（Point-to-Point Protocol）链路相连的路由器网络。

网络类型 - 广播型网络

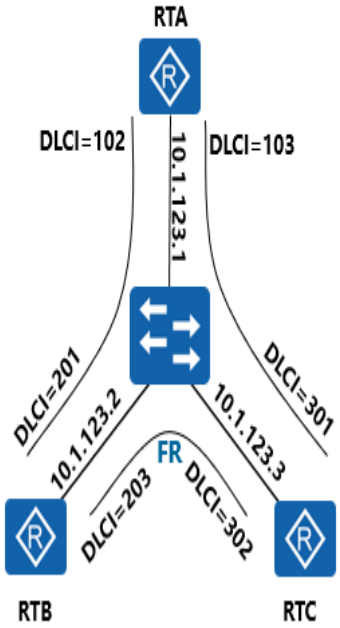
- 两台或两台以上的路由器通过共享介质互连。
- 支持广播、组播。



- 广播型网络支持两台及两台以上的设备接入同一共享链路且可以支持广播、组播报文的转发，是 OSPF 最常见的网络类型。
- 广播型网络的例子：通过以太网链路相连的路由器网络。
- 同时因为一个广播型网络中存在多台设备，邻居关系建立以及链路信息同步方面，OSPF 都有对应的特性来减少同一网络多台设备带来的不利影响。
- 以上两种网络类型是最常见的，此外，还有两种少见的网络类型。

网络类型 - NBMA网络

- 两台或两台以上路由器通过VC互连。
- 不支持广播、组播。

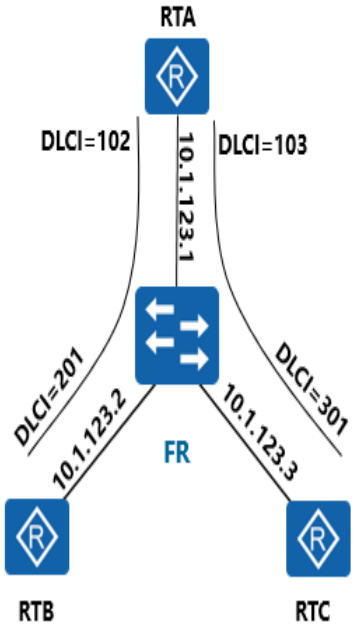


全连接的帧中继网络

- 与广播型网络不同的是 NBMA 网络默认不支持广播与组播报文的转发。在 NBMA 网络上，OSPF 模拟在广播型网络上的操作，但是每个路由器的邻居需要手动配置。
- NBMA（non-broadcast multiple access）型网络的例子：通过全互连的帧中继链路相连的路由器网络。
- 在现在的网络部署中，NBMA 网络已经很少了。

网络类型 - P2MP网络

- 多个点到点网络的集合。
- 支持广播、组播。



非完全连接的帧中继网络

- 将一个非广播网络看成是一组 P2P 网络，这样的非广播网络便成为了一个点到多点（P2MP）网络。在 P2MP 网络上，每个路由器的 OSPF 邻居可以使用反向地址解析协议（Inverse ARP）来发现。P2MP 可以看作是多个 P2P 的集合，P2MP 可以支持广播、组播的转发。
- 没有一种链路层协议默认属于 P2MP 类型网络，也就是说必须是由其他的网络类型强制更改为 P2MP。常见的做法是将非完全连接的帧中继或 ATM 改为 P2MP 的网络。
- 此外 OSPF 的链路状态信息中的开销值是如何度量的呢？

OSPF的度量方式

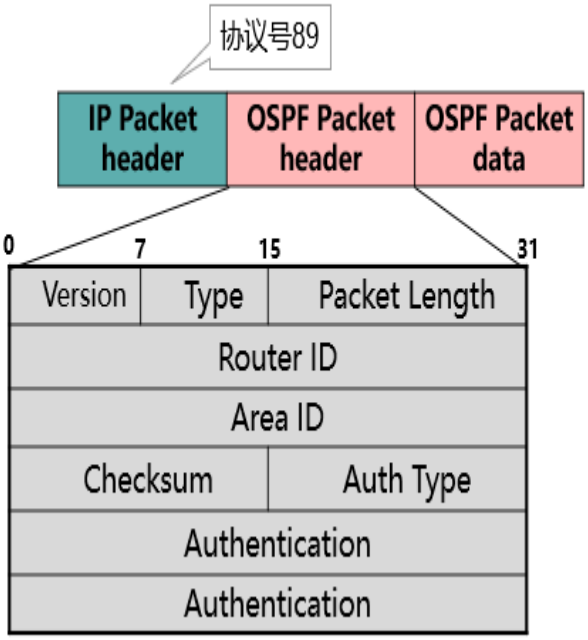
- 某接口cost=参考带宽/实际带宽。
- 更改cost的两种方式:
  - 直接在接口下配置;
  - 修改参考带宽 (所有路由器都需要修改，确保选路一致性) 。

- RTA到达192.168.3.0/24的“累计cost”=G1’s cost+G3’s cost

- OSPF 在计算接口的 cost 时，cost=参考带宽/实际带宽，默认参考带宽为 100M。当计算结果有小数位时，只取整数位；结果小于 1 时，cost 取 1。
- 若需要调整接口 cost 值有两种方式：
- 直接在接口下配置，需要注意的是，配置的 cost 是此接口最终的 cost 值，作用范围仅限于本接口。
- 修改 OSPF 的默认参考带宽值，作用范围是本路由器使能 OSPF 的接口。建议参考整个网络的带宽情况建立参考基线，所有路由器修改相同的参考带宽值，从而确保选路的一致性。
- 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，以 RTA 访问 RTC Loopback 1 接口 192.168.3.3 为例，其 cost=G1’s cost+ G3’s cost。

- 相比于 RIP，OSPF 的度量方式不仅考虑“跳数”，而且还考虑了“带宽”，比 RIP 更可靠的选择最优的转发路径。
- 那么 OSPF 路由器怎么表达链路状态信息并完成同步呢？

OSPF协议报文头部



- RIP 路由器之间是基于 UDP 520 的报文进行通信，OSPF 也有其规定的通信标准。OSPF 使用 IP 承载其报文，协议号为 89。
- 在 OSPF Packet 部分，所有的 OSPF 报文均使用相同的 OSPF 报文头部：
- Version ：对于当前所使用的 OSPFv2，该字段的值为 2。
- Type：OSPF 报文类型。
- Packet length：表示整个 OSPF 报文的长度，单位是字节。
- 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。
- 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 AuType 的值不同而不同。
- OSPF 的报文头部定义了 OSPF 路由器之间的通信的标准与规则，基于这个标准 OSPF 报文需要实现什么功能呢？

OSPF报文类型

Type	报文名称	报文功能
1	Hello	发现和维护邻居关系
2	Database Description	交互链路状态数据库摘要
3	Link State Request	请求特定的链路状态信息
4	Link State Update	发送详细的链路状态信息
5	Link State Ack	发送确认报文

- 思考：DD、LSR、LSU、LSAck报文都包含哪些信息？这么设计有什么好处？

- Type=1 为 Hello 报文，用来建立和维护邻居关系，邻居关系建立之前，路由器之间需要进行参数协商。
- Type=2 为数据库描述报文（DD），用来向邻居路由器描述本地链路状态数据库，使得邻居路由器识别出数据库中的 LSA 是否完整。
- Type=3 为链路状态请求报文（LSR），路由器根据邻居

的 DD 报文，判断本地数据库是否完整，如不完整，路由器把这些 LSA 记录进链路状态请求列表中，然后发送一个 LSR 给邻居路由器。

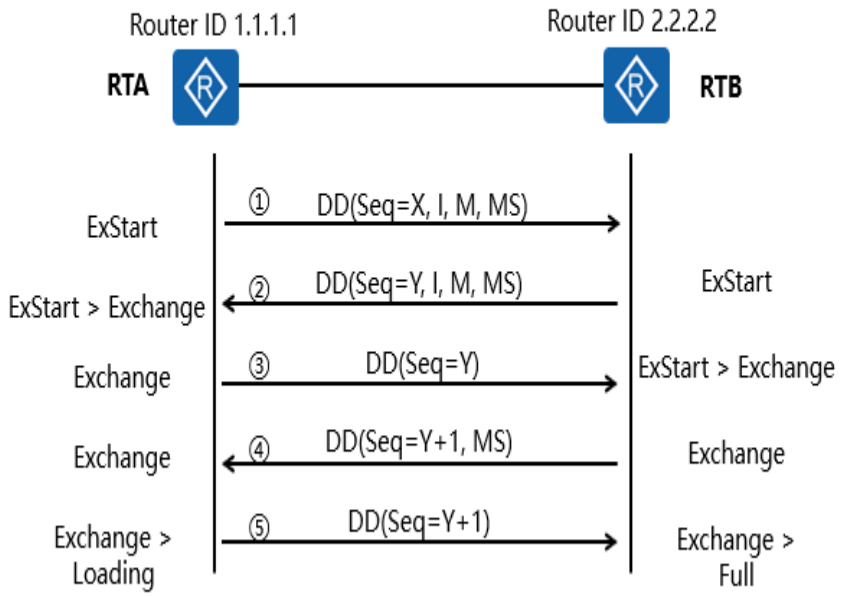
- Type=4 为链路状态更新报文（LSU），用于响应邻居路由器发来的 LSR，根据 LSR 中的请求列表，发送对应 LSA 给邻居路由器，真正实现 LSA 的泛洪与同步。
- Type=5 为链路状态确认报文（LSAck），用来对收到的 LSA 进行确认，保证同步过程的可靠性。
- DD、LSR、LSU、LSAck 与 LSA 的关系：
- DD 报文中包含 LSA 头部信息，包括 LS Type、LS ID、Advertising Router 、LS Sequence Number、LS Checksum。
- LSR 中包含 LS Type 、LS ID 和 Advertising Router 。
- LSU 中包含完整的 LSA 信息。
- LSAck 中包含 LSA 头部信息，包括 LS Type、LS ID、Advertising Router、LS Sequence Number、LS Checksum。
- 五种报文可以高效地完成 LSA 的同步，那么实际的报文交互过程是什么呢？

OSPF报文的功能需求

功能	实现分析
发现邻居与保持	Hello机制即可实现
LSA同步	双方互相发送LSA，完成同步； 同时同步速度更快，占用资源更少
可靠性	确保LSA同步过程的可靠性

- Hello 机制动态发现并维护邻居前文已介绍，不再赘述。
- RIP 设置了 Request 和 Response 两种报文来完成路由信息的同步，OSPF 路由器之间为了完成 LSA 的同步，可以直接把本地所有 LSA 发给邻居路由器，但是邻居路由器直接同步 LSA 并不是最好的方式。
- 更快速、更高效的方式是先在邻居路由器之间传送关键信息，路由器基于这些关键信息识别出哪些 LSA 是没有的、哪些是需要更新的，然后向邻居路由器请求详细的 LSA 内容。对于 OSPF 来说，需要有比 RIP 更高效、更可靠的方式来完成路由器之间的信息同步。

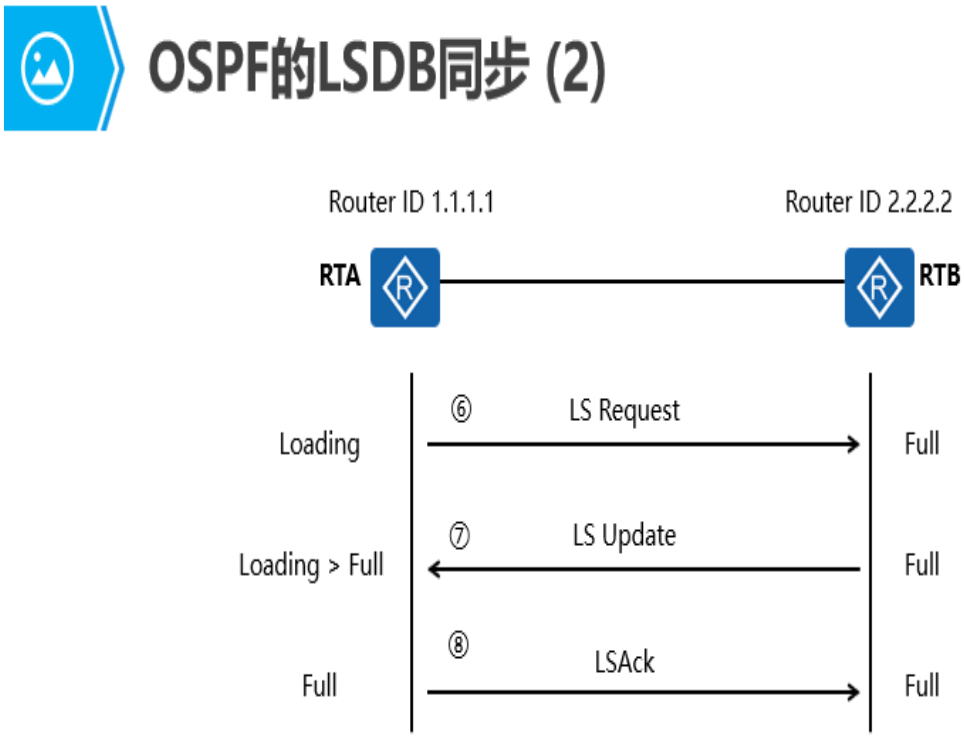
OSPF的LSDB同步 (1)



- 状态含义：
- 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。Master/Slave 关系是在此状态下形成的，初始 DD 序列号也是在此状态下确定的。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。
- Full：LSDB 同步过程完成，路由器与邻居之间形成了完全的邻接关系。
- LSDB 同步过程如下：
- RTA 和 RTB 的 Router ID 分别为 1.1.1.1 和 2.2.2.2 并且二者已建立了邻居关系。当 RTA 的邻居状态变为 ExStart 后，

RTA 会发送第一个 DD 报文。此报文中，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，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 RTA 宣告自己为 Master。

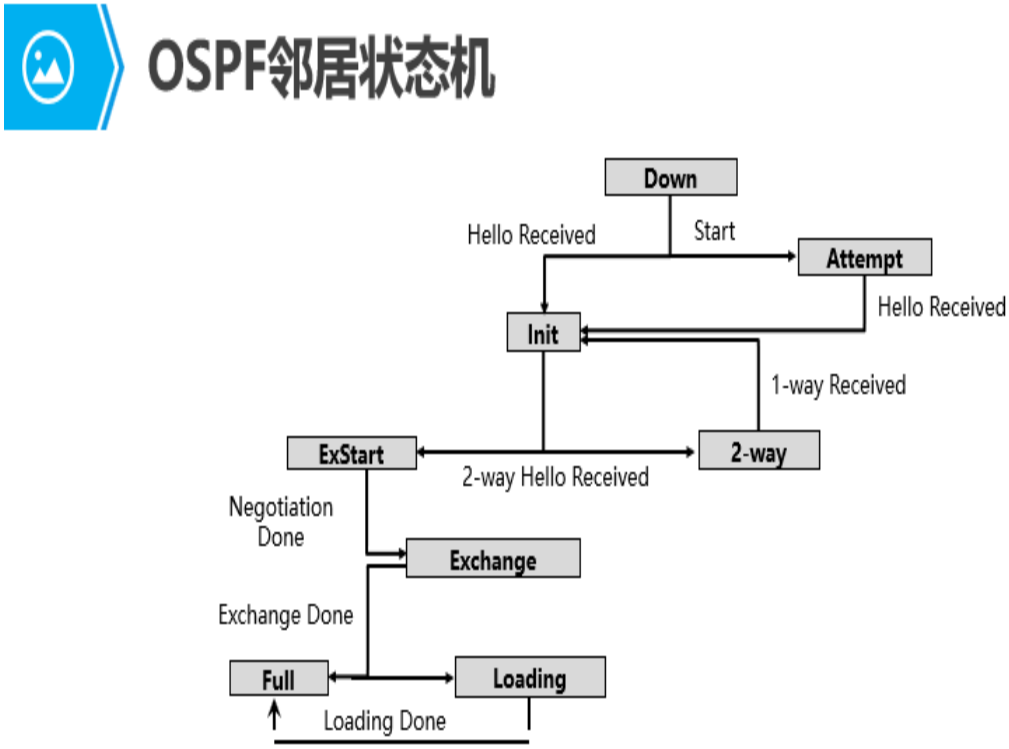
- 当 RTB 的邻居状态变为 ExStart 后，RTB 会发送第一个 DD 报文。此报文中，DD 序列号被随机设置为 Y（I-bit=1，M-bit=1，MS-bit=1，含义同上）。由于 RTB 的 Router ID 较大，所以 RTB 将成为真正的 Master。收到此报文后，RTA 会产生一个 Negotiation-Done 事件，并将邻居状态从 ExStart 变为 Exchange。



- RTA 开始向 RTB 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。
- RTB 向 RTA 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。RTA 在完成 LSU 报文的接收之

后，会将邻居状态从 Loading 变为 Full。

- RTA 向 RTB 发送 LSAck 报文，作为对 LSU 报文的确认。RTB 收到 LSAck 报文后，双方便建立起了完全的邻接关系。
- 从建立邻居关系到同步 LSDB 的过程较为复杂，错误的配置或设备链路故障都会导致无法完成 LSDB 同步。为了快速排障，最关键的是要理解不同状态之间切换的触发原因。

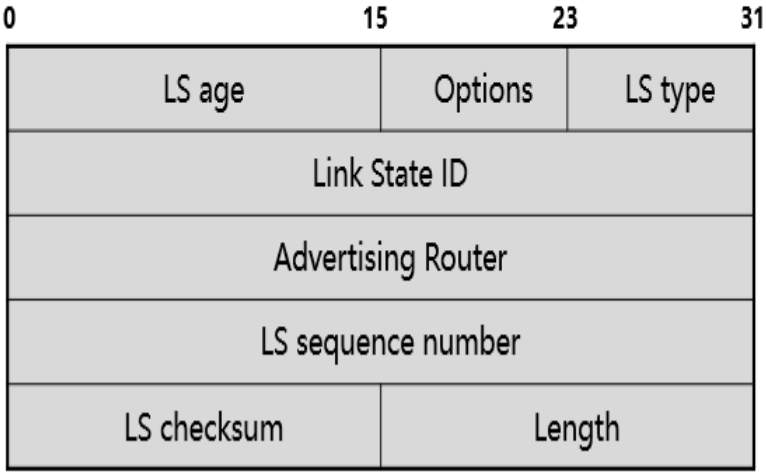


- 这是形成邻居关系的过程和相关邻居状态的变换过程。
- Down：这是邻居的初始状态，表示没有从邻居收到任何信息。在 NBMA 网络上，此状态下仍然可以向静态配置的邻居发送 Hello 报文，发送间隔为 PollInterval，通常和 Router DeadInterval 间隔相同。
- Attempt：此状态只在 NBMA 网络上存在，表示没有收到邻居的任何信息，但是已经周期性的向邻居发送报文，发送间隔为 HelloInterval。如果 Router DeadInterval 间隔内未收到邻居的 Hello 报文，则转为 Down 状态。

- Init：在此状态下，路由器已经从邻居收到了 Hello 报文，但是自己不在所收到的 Hello 报文的邻居列表中，表示尚未与邻居建立双向通信关系。在此状态下的邻居要被包含在自己所发送的 Hello 报文的邻居列表中。
- 2-Way Received：此事件表示路由器发现与邻居的双向通信已经开始（发现自己在邻居发送的 Hello 报文的邻居列表中）。Init 状态下产生此事件之后，如果需要和邻居建立邻接关系则进入 ExStart 状态，开始数据库同步过程，如果不能与邻居建立邻接关系则进入 2-Way。
- 2-Way：在此状态下，双向通信已经建立，但是没有与邻居建立邻接关系。这是建立邻接关系以前的最高级状态。
- 1-Way Received：此事件表示路由器发现自己没有在邻居发送 Hello 报文的邻居列表中，通常是由于对端邻居重启造成的。
- ExStart：这是形成邻接关系的第一个步骤，邻居状态变成此状态以后，路由器开始向邻居发送 DD 报文。主从关系是在此状态下形成的；初始 DD 序列号是在此状态下决定的。在此状态下发送的 DD 报文不包含链路状态描述。

LSA头部

- LSA是OSPF链路状态信息的载体。



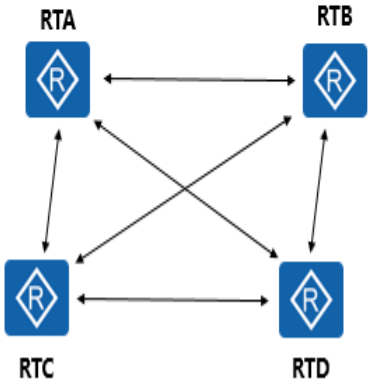
- LSA（Link State Advertisement）是路由器之间链路状态信息的载体。LSA 是 LSDB 的最小组成单位，也就是说 LSDB 由一条条 LSA 构成的。
- 所有的 LSA 都拥有相同的头部，关键字段的含义如下：
- LS age：此字段表示 LSA 已经生存的时间，单位是秒。
- LS type：此字段标识了 LSA 的格式和功能。常用的 LSA 类型有五种。
- Link State ID：此字段是该 LSA 所描述的那部分链路的标识，例如 Router ID 等。
- Advertising Router：此字段是产生此 LSA 的路由器的 Router ID。
- LS sequence number：此字段用于检测旧的和重复的 LSA。
- LS type，Link State ID 和 Advertising Router 的组合共

同标识一条 LSA。

- LSDB 中除了自己生成的 LSA，另一部分是从邻居路由器接收的。邻居路由器之间相互更新 LSA 必然需要一个“通道”。

MA网络中的问题

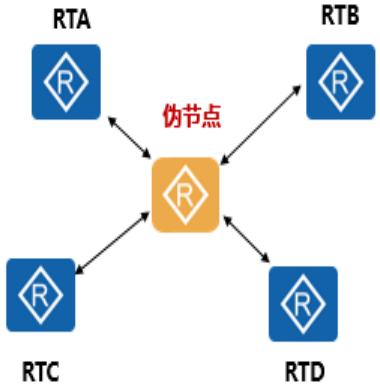
- n×(n-1)/2个邻接关系，管理复杂。
- 重复的LSA泛洪，造成资源浪费。



- 问题引出，在运行 OSPF 的 MA 网络包括广播型和 NBMA 网络，会存在两个问题：
- 在一个有 n 个路由器的网络，会形成(n×(n−1))/2 个邻接关系。
- 邻居间 LSA 的泛洪扩散混乱，相同的 LSA 会被复制多份，如 RTA 向其邻居 RTB、RTC、RTD 分别发送一份自己的 LSA，RTB 与 RTC、RTC 与 RTD、RTB 与 RTD 之间也会形成邻居关系，也会发送 RTA 的 LSA。
- 这样的工作效率显然是很低的，消耗资源的。作为高级的路由协议，OSPF 是怎样解决这些问题的呢？

DR与BDR作用

- 减少邻接关系。
- 降低OSPF协议流量。

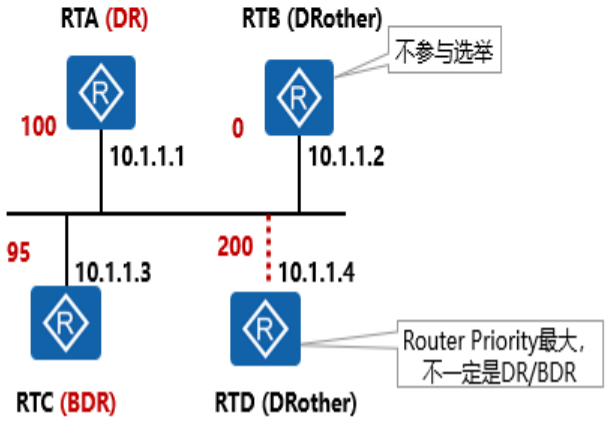


- 思考：DR的单点故障怎么解决?

- DR（Designated Router）即指定路由器，其负责在 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。这样就大大减少了 MA 网络中的邻接关系数量及交换链路状态信息消耗的资源。
- DR 一旦出现故障，其与其他路由器之间的邻接关系将全部失效，链路状态数据库也无法同步。此时就需要重新选举 DR，再与非 DR 路由器建立邻接关系，完成 LSA 的同步。为了规避单点故障风险，通过选举备份指定路由器 BDR，在 DR 失效时快速接管 DR 的工作。
- 伪节点是一个虚拟设备节点，其功能需要某台路由器来承载，下面将介绍 DR/BDR 的选举规则。

DR与BDR选举

- 选举规则：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。



邻居与邻接关系

网络类型	是否和邻居建立邻接关系
P2P	是
Broadcast	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	
P2MP	是

- 邻居（Neighbor）关系与邻接（Adjacency）关系是两个不同的概念。OSPF 路由器之间建立邻居关系后，进行 LSDB 同步，最终形成邻接关系。
- 在 P2P 网络及 P2MP 网络上，具有邻居关系的路由器之间会进一步建立邻接关系。
- 在广播型网络及 NBMA 网络上，非 DR/BDR 路由器之间只能建立邻居关系，不能建立邻接关系，非 DR/BDR 路由器与 DR/BDR 路由器之间会建立邻接关系，DR 与 BDR 之间也会建立邻接关系。
- 邻接关系建立完成，意味着 LSDB 已经完成同步，接下来 OSPF 路由器将基于 LSDB 使用 SPF 算法计算路由。

思考题

1. 下列哪些选项属于OSPF报文类型?
   A. Hello
   B. Database Description
   C. Link State Request
   D. Link State DD
   E. Link State Advertisement
2. OSPF的网络类型包括（　）?

答案：ABC。
答案：P2P 网络、P2MP 网络、广播型网络、NBMA 网络。