

第 5 章 链路状态路由协议 OSPF

本章重点介绍 OSPF 的工作流程,根据不同的网络类型介绍单区域 OSPF 的配置,根据不同的区域类型介绍多区域 OSPF 的配置等。

5.1 OSPF 的基本概念

OSPF(Open Shortest Path First,开放最短路径优先)是一种典型的链路状态路由协议,启用 OSPF 协议的路由器彼此交换并保存整个网络的链路信息,从而掌握全网的拓扑结构,再通过 SPF(最短路径优先)算法计算出到达每一个网络的最佳路由。

OSPF 作为一种内部网关协议(Interior Gateway Protocol),其网关和路由器都在同一个自治系统内部,用于在同一个自治域(AS)中的路由器之间发布路由信息。运行 OSPF 的每一台路由器中都维护一个描述自治系统拓扑结构的统一的数据库(链路状态数据库),该数据库由每一个路由器的链路状态信息(该路由器可用的接口信息、邻居信息等)、路由器相连的网络状态信息(该网络所连接的路由器)、外部状态信息(该自治系统的外部路由信息)等组成。所有的路由器并行运行着同样的算法(SPF),根据该路由器的链路状态数据库(拓扑结构),构造出以它自己为根节点的最短路径树,该最短路径树的叶子节点是自治系统内部的其他路由器。当到达同一目标路由器存在多条相同代价的路由时,OSPF 能够在多条路由上分配流量,实现负载均衡。

OSPF 不同于距离矢量协议(如: RIP),有如下特性:

- 支持大型网络,路由收敛快,占用网络资源少。
- 无路由环路。
- 支持 VLSM 和 CIDR。
- 支持等价路由。
- 支持区域划分,构成结构化网络,提供路由分级管理。

1. 路由器 ID

(1) 通过 router-id 命令指定的路由器 ID 最优先:

```
Router(config-router)# router-id 1.1.1.1
```

(2) 选择具有最高 IP 地址的环回接口:

```
Router(config)# int loopback 0  
Router(config)# ip address 10.1.1.1 255.255.255.255
```

(3) 再选择具有最高 IP 地址的已激活的物理接口:

```
Router(config)# int f1/1  
Router(config)# ip address 170.10.1.1 255.255.255.255
```

2. 邻居

启用 OSPF 的第一步是建立毗邻关系。路由器 A 从自己的端口向外多播发送 Hello 报文,通告自己的路由器 ID 等,所有与路由器 A 物理上直连且同样运行 OSPF 协议的路由器称为邻居路由器。如果邻居路由器 B 收到这个 Hello 报文,就将这个报文内路由器 A 的 ID 信息加入自己的 Hello 报文内。当路由器 A 的某端口收到从邻居路由器 B 发送的含有自身 ID 信息的 Hello 报文后,A、B 两台路由器就处于 Two-way 状态,从而建立了邻居关系。

3. 邻接

两台路由器建立了邻居关系后,再根据该端口所在的网络类型来确定这两台路由器是否需要交换链路状态信息,此时两台路由器处于 Full 状态,需要交换链路状态信息时称建立了邻接(adjacency)关系。

4. 链路状态

与链路的工作状态(是正常工作还是发生故障)相关的信息称为链路状态(Link-State)。

OSPF 路由器收集其所在网络区域中各路由器的连接状态信息,即链路状态信息,生成链路状态数据库(Link-State Database,LSDB)。路由器掌握了该区域上所有路由器的链路状态信息,也就等于掌握了该区域的网络拓扑状况。

5. 链路状态公告和链路状态数据库

OSPF 路由器之间使用链路状态通告(Link-State Advertisement,LSA)来交换各自的链路状态信息,并把获得的信息存储在链路状态数据库中。

根据路由器的不同类型,定义了 7 种类型的 LSA。LSA 中包括的信息有路由器 ID、邻居路由器 ID、链路的带宽、路由条目、掩码等信息。

路由器 LSA(第 1 类 LSA)由区域内所有路由器产生,并且只能在本区域内泛洪。这些最基本的 LSA 列出了路由器所有的链路和接口、链路状态及代价。

6. 链路开销

OSPF 路由协议通过计算链路的带宽来计算最佳路径的选择。每条链路根据带宽不同具有不同的度量值,这个度量值在 OSPF 路由协议中称为链路开销。其计算公式是 $10^8/\text{带宽}$ (单位是 b/s)。通常,环回接口的链路开销是 1,10Mb/s 以太网的链路开销是 10,16Mb/s 令牌环网的链路开销是 6,FDDI 或快速以太网的链路开销是 1,2Mb/s 的串行链路的链路开销是 48。

两台路由器之间链路开销之和的最小值为最佳链路。

7. 邻居表、拓扑表和路由表

OSPF 路由协议维护 3 张表:邻居表、拓扑表和路由表。最基础的就是邻居表。

路由器通过发送 Hello 包将与其物理直连且同样运行 OSPF 路由协议的路由器作为邻居放在邻居表中。

当路由器建立了邻居表之后,运行 OSPF 路由协议的路由器会互相通告自己所了解的网络拓扑,从而建立拓扑表。在一个区域内,一旦收敛,所有的路由器就具有相同的拓扑表。

当完整的拓扑表建立起来后,路由器便会按照链路带宽的不同,使用 SPF 算法从拓扑表中找出最佳路由,放在路由表中。

8. 指定路由器

在接口所连接的各邻居路由器中将具有最高优先级的路由器作为指定路由器 (Designative Router, DR)。端口的优先权值为 0~255, 在优先级相同的情况下, 选 ID 值最高的路由器作为 DR。

9. 备份指定路由器

在各邻居路由器中选择具有次高优先级的路由器作为备份指定路由器 (Backup Designative Router, BDR)。优先级相同时比较路由器 ID。

10. OSPF 网络类型

根据路由器所连接的物理网络不同, OSPF 将网络划分为 4 种类型: 广播多路访问型、非广播多路访问型、点到点型、点到多点型。

- 广播多路访问型 (BMA) 网络, 如以太网 (Ethernet)、令牌环网 (Token Ring)、FDDI。它选举 DR 和 BDR。涉及 IP 地址和 MAC 地址, 用 ARP 实现二层和三层映射。
- 非广播多路访问型 (NBMA) 网络, 如帧中继 (Frame Relay)、X.25、SMDS。它选举 DR 和 BDR。网络中允许存在多台路由器, 在物理上共享链路, 通过二层虚链路建立逻辑上的连接。广播针对每一条虚链路发送, 而不是针对全网发送的广播或多播分组, 所以其他路由器收不到广播。
- 点到点型 (point-to-point) 网络, 一个网络里仅有两个接口, 使用 HDLC 或 PPP 封装, 不需寻址, 地址字段固定为 FF。
- 点到多点型 (point-to-multipoint) 网络, 又分为点到多点广播式网络和点到多点非广播式网络。

11. 区域

OSPF 引入了分层路由的概念, 将网络分割成一个主干连接的一组相互独立的部分, 这些相互独立的部分称为区域 (Area), 主干的部分称为主干区域。一个区域就如同一个独立的网络, 该区域的 OSPF 路由器只保存该区域的链路状态, 同一区域的链路状态数据库保持同步, 使得每个路由器的链路状态数据库都可以保持合理的大小, 路由计算的时间、报文数量都不会过大。

多区域的 OSPF 必须存在一个主干区域 (Area0), 主干区域负责收集非主干区域发出的汇总路由信息, 并将这些信息返回各区域。

OSPF 区域不能随意划分, 应该合理地选择区域边界, 使不同区域之间的通信量最小。在实际应用中, 区域的划分往往不是根据通信模式而是根据地理或政治因素来完成的。分区域的好处如下:

- (1) 减少路由更新。
- (2) 加速收敛。
- (3) 将不稳定限制在一个区域内。
- (4) 提高网络性能。

12. 路由器的类型

根据路由器在区域中的位置不同分为 4 种类型, 如图 5-1 所示。

(1) 内部路由器 (IR): 所有端口都在同一区域的路由器, 它们都维护着一个相同的链路状态数据库。

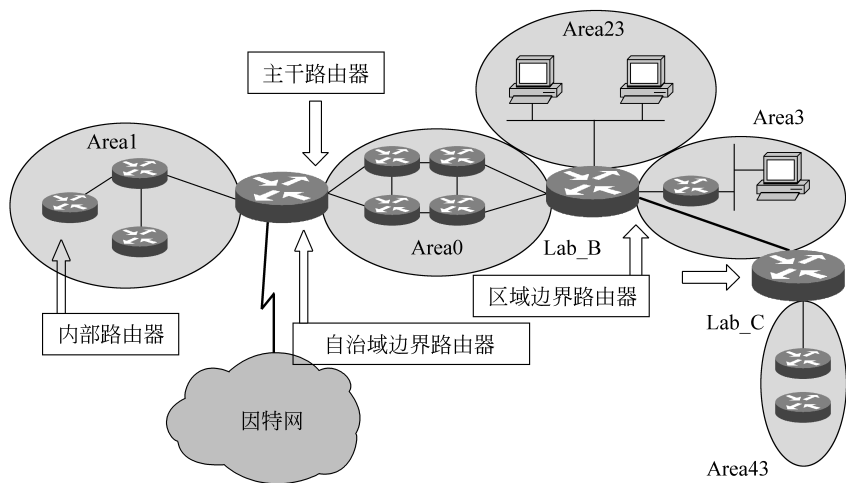


图 5-1 OSPF 路由器的类型

(2) 主干路由器：至少有一个连接主干区域端口的路由器。

(3) 区域边界路由器(ABR)：具有连接多区域端口的路由器，一般作为一个区域的出口。ABR 为其连接的每一个区域单独建立链路状态数据库，负责将其连接区域的路由摘要信息发送到主干区域，而主干区域上的 ABR 则负责将这些信息发送到其连接的所有其他区域。

(4) 自治域边界路由器(ASBR)：至少拥有一个连接外部自治域网络(如非 OSPF 的网络)端口的路由器，负责将非 OSPF 网络信息传入 OSPF 网络。

5.2 OSPF 的工作过程

图 5-2 给出了同一区域内 OSPF 的工作流程。

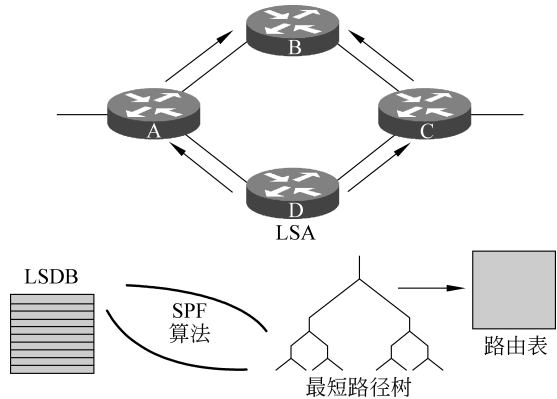


图 5-2 OSPF 的工作流程

运行 OSPF 协议的路由器通过发送 Hello 数据包建立邻居关系，并彼此交换链路状态信息，链路状态信息被加载在 LSA 中，以 LSU(链路状态更新包)的形式在网络中进行洪

泛。OSPF 把这些链路状态信息存放在本地链路状态数据库中。在掌握了区域内所有链路状态信息后,每一个 OSPF 路由器都以自己为根节点,用 Dijkstra 算法计算到其他路由器(其他叶节点)的最短路径树(SPF Tree),从而产生路由表。

具体步骤如下:

- (1) 建立路由器的邻居关系。
- (2) 进行必要的 DR/BDR 选举。
- (3) 保持链路状态数据库的同步。
- (4) 产生路由表。
- (5) 维护路由信息。

5.2.1 建立路由器的邻居关系

OSPF 协议通过 Hello 报文建立路由器的邻居关系。每个 Hello 数据包都包含以下信息:始发路由器 ID、始发路由器接口的区域地址、始发路由器的接口地址掩码、始发路由器的认证信息和类型、始发路由器的 Hello 时间间隔、始发路由器的无效路由的时间间隔、路由器的优先级、DR 和 BDR、标识(可选 5 个标记位)和始发路由器所有有效邻居路由器 ID。

邻居关系的建立要经过 3 个状态,如图 5-3 所示。

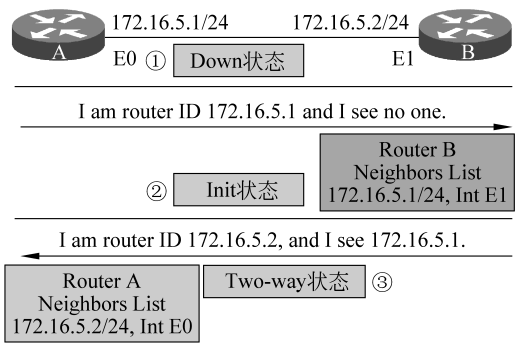


图 5-3 建立路由器的邻接关系

当路由器 A、B 启动时,它们处于 Down 状态。

路由器 A 从其各个接口通过 224.0.0.5 以固定的时间间隔(10s)向所有邻居(包括 B)发送 Hello 报文,通告自己的路由器 ID(172.16.5.1)。其他路由器收到这个 Hello 报文后,就会把它加入自己的邻居表中(路由器 B 把路由器 A 的 ID 加入 Hello 报文的邻居 ID 字段),从而进入 Init 状态。

路由器 B 向路由器 A 发送 Hello 报文,其包含自己和其他邻居路由器的信息(路由器 A 的 ID 在路由器 B 的邻居表中);当路由器 A 看到自己出现在另一邻居路由器的 Hello 报文中时,就把其中的邻居关系加入自己的数据库中,进入 Two-way 状态,路由器 A 和路由器 B 就建立了双向通信,从而建立了邻居关系。

进入 Two-way 状态后,路由器 A 将决定和谁建立邻接关系,这是根据各接口所连接的网络类型决定的。即使两台路由器是邻居,但它们不一定建立邻接关系。如果是点到点(PPP、HDLC)的网络,就与其直连的路由器建立邻接关系;如果是多路访问型,包括广播

(以太网、令牌环、FDDI)和非广播(帧中继、X.25)的网络,则进入第二步,进行必要的 DR/BDR 选举。每台路由器只与 DR/BDR 建立邻接关系,其他路由器之间不建立邻接关系。如果不需要进行 DR/BDR 选举,路由器就进入第三步,交换链路状态数据库,使拓扑结构保持一致。

5.2.2 选举 DR 和 BDR

在初始状态下,路由器将活动接口设置为 DR 和 BDR 为 0.0.0.0,这意味着没有 DR 和 BDR 被选举出来。同时路由器设置 Wait Timer,其值为 Router Dead Interval,其作用是如果在这段时间内还没有收到有关 DR 和 BDR 的宣告,那么它就宣告自己为 DR 或 BDR。然后在发送 Hello 包后进行 DR 和 BDR 的选举:

首先比较 Hello 包中的优先级,最高的为 DR,次高的为 BDR。默认优先级都为 1;当优先级相同时,再比较路由器 ID,最高的为 DR,次高的为 BDR;当优先级设置为 0 时,不参加 DR/BDR 的选举,只能成为 DRoother。

DR 和 BDR 选举不具有抢占性,选举完成后,将一直保持,直到 DR 和 BDR 失效(或强行关闭 DR 及 BDR 的路由器,或用 clear ip ospf process 手工配置重新开始运行 OSPF 路由协议),否则即使新加入更高优先级的路由器,也不会再进行选举。

在多路访问网络(广播或非广播)中,选举 DR 和 BDR。在点到点的网络中,不选举 DR 和 BDR。在点到多点的网络中,若将其分解配置为点到多点非广播多路访问网络(NBMA),全互联的邻居路由器属于同一个子网的,采用人工配置产生 DR 和 BDR;若为点到多点广播多路访问网络(BMA)属于同一个子网的,自动选举 DR 和 BDR。

DR/BDR 选举完成后,DRoother 只和 DR/BDR 形成邻接关系,各 DRoother 之间不建立邻接关系。224.0.0.5 是 DRoother 的多播地址,224.0.0.6 是 DR、BDR 的多播地址。因此 DRoother 向 224.0.0.6 这个多播地址发送自己的 LSU,DR/BDR 收到此 LSU 并汇总后,向 224.0.0.5 这个多播地址发送 LSU,从而泛洪到所有 DRoother 路由器上。

由 DR(或 BDR)与本区域内所有其他路由器之间交换链路状态信息,进入准启动(Exstart)状态。

在点到点的网络中,两台路由器之间建立主从关系,路由器 ID 高的作为主路由器,另一台作为从路由器,也进入 Exstart 状态。

5.2.3 链路状态数据库的同步

在 OSPF 中,必须保持同一区域范围内所有路由器的链路状态数据库同步。

通过建立并保持邻接关系,OSPF 首先使具有邻接关系的路由器的数据库同步,进而保证同一区域范围内所有路由器的数据库同步。

数据库同步过程从建立邻接关系(Exstart 状态)开始,在完全邻接关系(Full 状态)时结束。

在点到点的网络中,当路由器的端口状态为 Exstart 时,路由器通过发一个空的数据库描述包来协商主从关系以及数据库描述包的序号,路由器 ID 大的为主,反之为从。主路由器首先将自己的链路状态信息发给从路由器。主从相互交换链路状态数据库汇总后,进入 Exchange 状态,如图 5-4 所示。

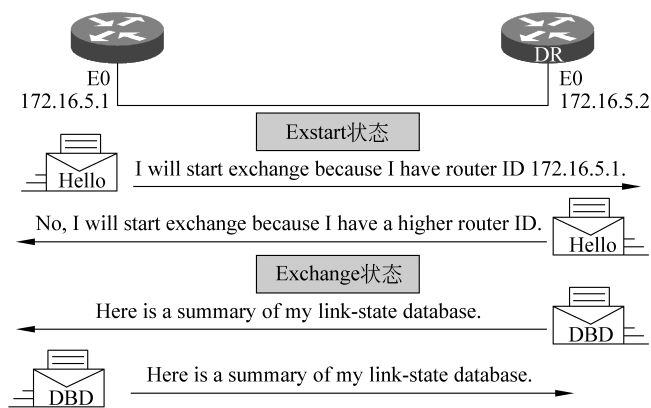


图 5-4 链路状态数据库同步过程一

在多路访问网络中,DR 和 BDR 选举好后,进入 Exstart 状态。DR 或 BDR 先将自己的链路状态汇总信息发给其他路由器,其他路由器再将各自的链路状态汇总信息发给 DR 或 BDR,而在其他路由器之间不相互交换链路状态信息。最后,当同一区域内链路状态数据库汇总达到一致后,进入 Exchange 状态。

在链路状态数据库同步过程中,有以下几种形式的数据包:

- 链路状态描述包(DBD),发送路由器的链路状态数据库汇总数据包。
- 链路状态请求包(LSR),请求链路状态数据库中某一条目的完整信息。
- 链路状态更新包(LSA),给出链路状态数据库中某一条目的完整信息。
- 链路状态确认包(LSAck),收到一个链路状态更新包后的确认。

以点到点的网络为例,主路由器发送链路状态描述包(数据库描述包),从路由器收到链路状态描述包后,向主路由器发送链路状态确认包,并检查自己的链路状态数据库,如果发现链路状态数据库里没有某些项,则添加它们,并将这些项加入到链路状态请求列表中,向主路由器发送链路状态请求包,如图 5-5 所示。当主路由器收到链路状态请求包时,返回链

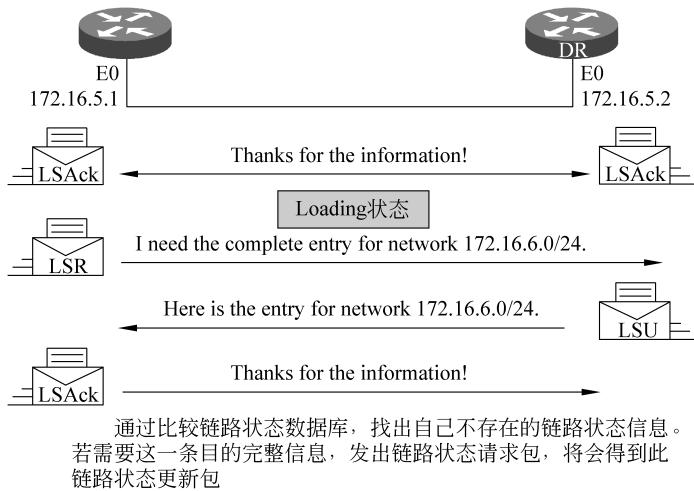


图 5-5 链路状态数据库同步过程二

路状态更新包,进行链路状态的更新。从路由器收到链路状态更新包后发出链路状态确认包,进行确认,表示收到该更新包,否则主路由器就在重发定时器的启动下进行重复发送。当所有的数据库请求包都已被主路由器处理后,主从路由器也就进入了 Full(完全邻接)状态。

同理,在多路访问网络中,当 DR 与整个区域内所有的路由器都完成链路状态数据库的更新,整个区域中所有路由器的链路状态数据库同步时,即进入 Full 状态。

5.2.4 路由表的产生

当链路状态数据库达到同步以后,同一区域内所有的路由器都具有了相同的链路状态数据库(即拓扑表),通过 SPF 算法计算并产生路由表。SPF 算法的核心是:将当前路由器到目标路由器之间的所有链路开销相加,并选出一个开销最低的路径作为最佳路径,从而得到以当前路由器为根节点,到达每一台路由器(叶节点)的一条最佳路径,形成一棵最小生成树。图 5-6 给出了 SPF 算法的基本过程。OSPF 最多允许 4 个等值的路由项以进行负载均衡。

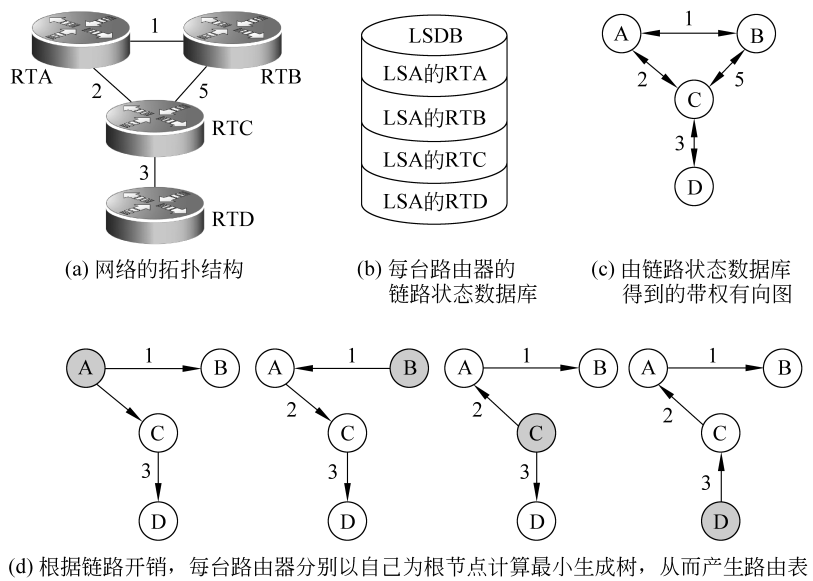


图 5-6 SPF 算法的基本过程

OSPF 协议中的 SPF 算法计算路由的过程如下(图 5-6):

- (1) 各路由器发送自己的 LSA,其描述了自己的链路状态信息。
- (2) 各路由器汇总收到的所有 LSA,生成 LSDB。
- (3) 各路由器以自己为根节点计算出最小生成树,依据是链路开销。
- (4) 各路由器按照自己的最小生成树得出路由条目并加入路由表。

5.2.5 维护路由信息

在 OSPF 路由协议中,所有路由器的拓扑结构数据库必须保持同步。当链路状态发生变化时,路由器通过扩散过程将这一变化通知给网络中的其他路由器。图 5-7 显示了链路

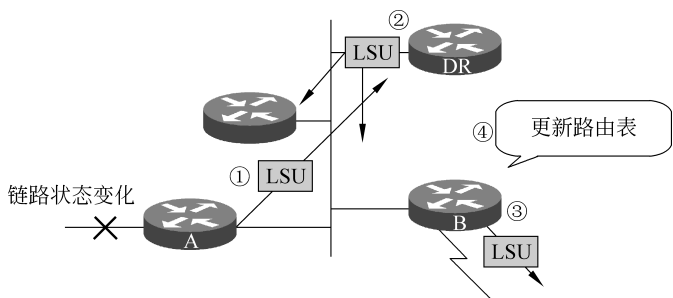


图 5-7 OSPF 中链路状态更新

状态的更新过程。

路由器对某一条链路的状态更新称为 LSA,对一组链路的状态更新称为 LSU,LSU 更新包里可包含多个 LSA。

当路由器 A 的链路出现故障时,发送链路状态更新 LSU 到 DR 和 BDR(其多播地址为 224.0.0.6)。

DR 和 BDR 利用多播地址 224.0.0.5,再把此 LSU 泛洪到除路由器 A 以外的所有路由器,以通知其他路由器。

路由器 B 收到 DR 或 BDR 发来的 LSU 后再扩散到它的邻居(其他区域),最终扩散到整个网络。

当整个网络的拓扑结构保持同步时,每台路由器开始利用 SPF 算法,重新计算路由,得到新的路由表。

5.2.6 OSPF 运行状态和协议包

OSPF 路由器在完全邻接之前,要经过以下几个运行状态,各运行状态之间的关系如图 5-8 所示。

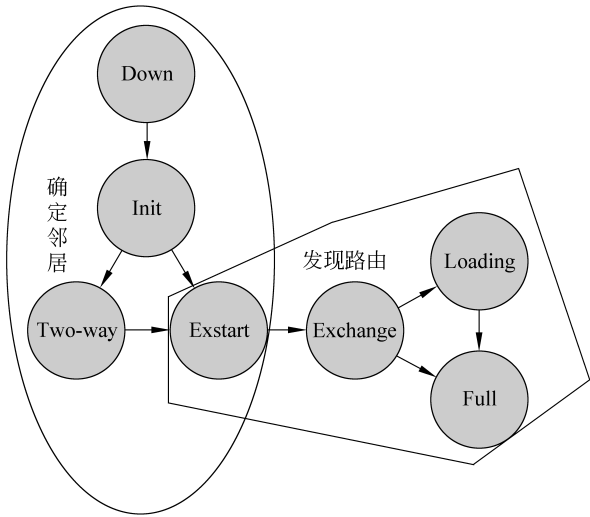


图 5-8 OSPF 中各运行状态之间的关系

- Down: 此状态还没有与其他路由器交换信息。首先从其 OSPF 接口使用多播地址 224.0.0.5 向外发送 Hello 报文,此时并不知道谁是 DR/BDR(若为广播网络)和任何其他路由器。
- Attempt: 只适于 NBMA 网络,在 NBMA 网络中邻居是手动指定的,在该状态下,路由器将使用 HelloInterval 取代 PollInterval 来发送 Hello 报文。
- Init: 在 DeadInterval 里收到了 Hello 报文,将邻居信息放在自己的邻居表中,并将其包含在 Hello 报文中,再从自己的所有接口使用多播地址 224.0.0.5 发送出去。
- Two-way: 当路由器 ID 彼此出现在对方的邻居列表中时,建立双向会话。
- Exstart: 信息交换初始状态,在这种状态下,选举了 DR/BDR,路由器和它的邻居将建立主从关系,并确定链路状态描述包的序列号。
- Exchange: 信息交换状态,路由器和它的邻居交换一个或多个链路状态描述包。DBD 报文中包含有关 LSDB 中 LSA 条目的摘要信息。
- Loading: 信息加载状态,收到 DBD 后,使用 LSAck 报文确认已收到 DBD。将收到的信息同 LSDB 中的信息进行比较。如果 DBD 中有更新的链路状态条目,则向对方发送一个 LSR,用于请求新的 LSA。
- Full: 完全邻接状态,当网络中所有路由器的 LSDB 同步时,即拓扑表保持一致,进入完全邻接状态。

OSPF 共使用 5 种路由协议包: Hello 包、链路状态描述包(DBD)、链路状态请求包(LSR)、链路状态更新包(LSA)、链路状态确认包(LSAck),用于 OSPF 运行过程中不同状态下各个路由器之间交换信息。每种协议包都包含 24B 的 OSPF 协议包的首部,如图 5-9 所示。

版本号	类型	包长度
路由器ID		
区域ID		
校验和	认证类型	
认证		

图 5-9 OSPF 协议包的首部字段

DBD 是类型为 2 的 OSPF 包,在形成邻接关系的过程中,路由器之间交换链路状态信息。根据接口数和网络数,可能需要不止一个 DBD 来传输整个网络链路状态数据库。在交换的过程中所涉及的路由器建立主从关系。主路由器发送包,而从路由器通过使用数据库描述(Database Description,DD)序列号认可接收到的包。图 5-10 给出了 DBD 的参数字段。其中 Interface MTU 指示通过该接口可发送的最大 IP 包长度,当通过虚链路发送包时,这个字段设置为 0。选项字段包含 3 位,用于显示路由器的能力。I 位是 Init 位,对数据库序列中的第一个包,设置为 1。M 位设置为 1,表示在序列中还有更多的数据库描述包。MS 位是主从位,在数据库描述包交换期间,1 表示路由器是主路由器,而 0 表示路由器是从路由器。数据库描述序列号是 DBD 的序列号。包的其余部分是一个或多个 LSA 首部。

链路状态请求包是类型为 3 的 OSPF 包,它包含 LS 类型、链路状态 ID、宣告路由器几个主要字段。当两个路由器完成交换数据库描述包时,路由器可检测链路状态数据库是否