

OSPF

By ZhuLinfeng



- 1. 基础概念介绍
- 2. OSPF概念介绍
- 3. OSPF过程解析
- 4. OSPF包格式

PPT文档说明

本PPT主要以RFC 2328为参考，从原理上介绍OSPF协议，以全面、准

确、系统、简单为标准，节约大家学习OSPF的时间开销。

内容范围是RFC 2328前四个章节，这对理解OSPF运行过程绰绰有余。

在对OSPF有了整体把握之后，至于RFC 2328余下的章节，可以简单的

按需查阅，获取更细节的内容。

- 1. 图形含义约定
- 2. “网络”的语义
- 3. 最短路径优先算法
- 4. 自治系统
- 5. 路由
- 6. 路由协议分类

图形含义约定



主机：主机名称以H开头，如主机 H1 , H2



路由器：路由器名称以TR开头，如路由器 RT1

路由器接口以 I 开头，如接口 Ia

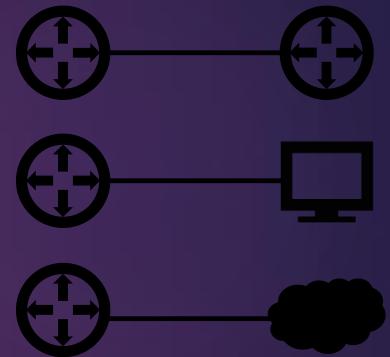


网络：网络名称以N开头，如网络 N1

图形含义约定

直线：直线会表示两种含义，但在一个图中只会选取一种

- □ 在交换链路信息时表示路由器间有邻接关系
- □ 在其他场合，表示两设备间有直接的物理连接



虚线：表示两设备间有虚连接。有时设备并不直接相连，中间会经过其他系列设备，但可通过配置虚链接，逻辑上认为这两台设备有邻接关系



1.1

图形含义约定

距离：距离用正整数表示

- 距离指设备接口的链路开销，带宽越大，开销越小，距离值也就越小
- 距离是单向的，即 $RTa \rightarrow RTb$ 的距离与 $RTb \rightarrow RTa$ 的距离是不同的



$RTa \rightarrow RTb$ 的距离是 3

$RTb \rightarrow RTa$ 的距离是 2



$RTa \rightarrow N1$ 的距离是 3

$N1 \rightarrow RTa$ 的距离是 0

网络到设备的距离一律为0，故在图中不予标明，但设备到网络的距离不一定为0。
(品味“距离指设备接口的链路开销”这句话可知原因)

1.1

图形含义约定

AS：自治系统AS以矩形虚线框表示

□ 处于AS边界的路由器为ASBR

AS



区域：AS可以划分为多个区域分治管理，

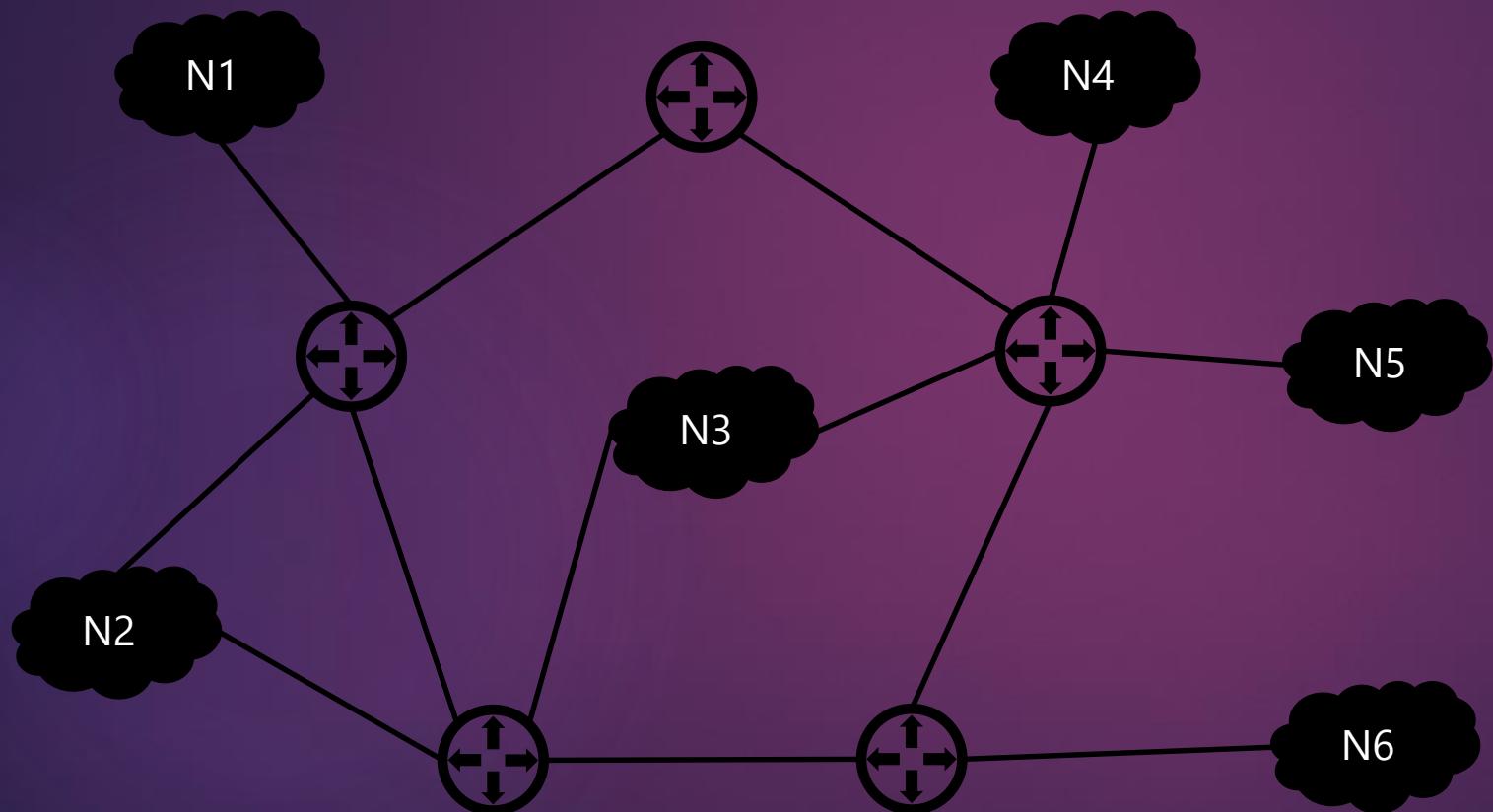
区域以半透明圆角矩形表示

□ 处于区域边界的路由器为ABR

①



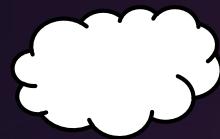
OSPF 协议中“网络”的语义



在OSPF的RFC文档中，“网络”是个使用频繁的词语。然而根据的场景不同，“网络”一词描述的对象会有所差别，这点RFC文并未明确区分。

在讨论自治系统（AS）时，左图整体可以看做是个自治系统网络。

OSPF 协议中“网络”的语义



用此图标标明的网络，其中可能包含着大量主机，你的PC正处于其中。它是 IP 报文的源地址或目的地址，路由器需要把IP报文最终转发到这样的网络。

该网络由子网号进行划分
子网号 = IP地址 + 子网掩码
子网号限定了此网络 IP 地址的范围

存根网络 (Stub Networks) :

是一种通过单一路由访问的网络，与外界只有一个输出连接。此网络中的设备向外界发报文时不存在选路的问题，因此可以配置默认路由，即配置所谓默认网关。

图中N1, N4, N5, N6都是存根网络

传输网络 (Transit Networks) :

源地址和目的地址都不属于此网络的 IP 报文也可在此网络内传输。

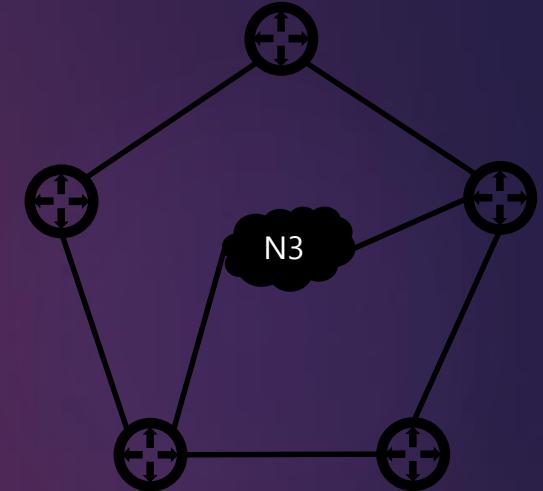
图中N2, N3都是传输网络



192.168.0.0/16

OSPF 协议中“网络”的语义

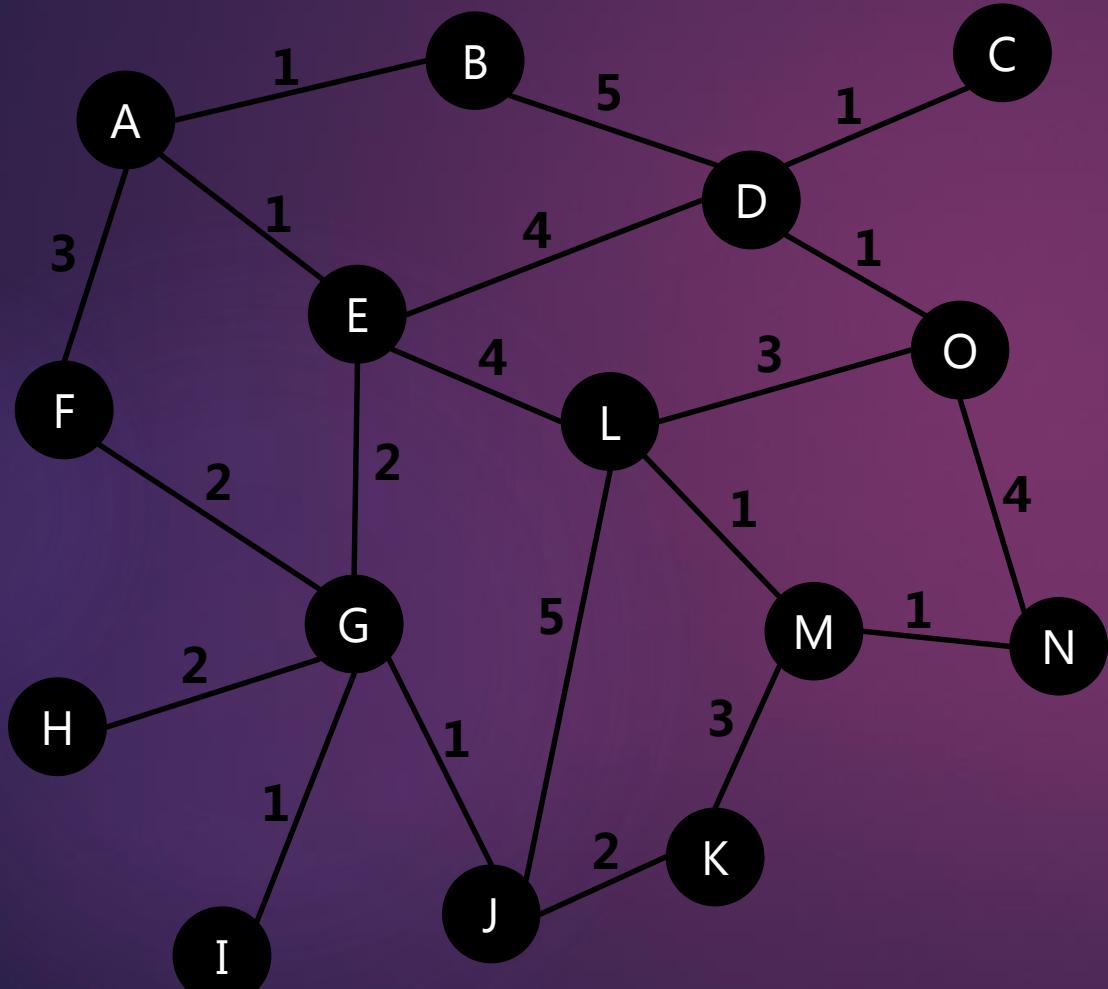
在 OSPF 协议运行的过程中，路由器需要相互通信。为处理方便，它们之间的连接关系也需要抽象为一个或多个网络。网络考虑的对象仅仅是路由器。



OSPF 的四种网络类型

- 点对点网络 (Point-to-point networks)
- 广播网络 (Broadcast Networks)
- 非广播网络 (Non Broadcast Networks)
- 点对多点网络 (Point-to-MultiPoint Networks)

最短路径优先算法 (Dijkstra 算法)



Dijkstra 算法是个著名的图论算法，具体可以参见《算法导论》有兴趣者可以自己编程实践，算法步骤不在此讨论。这里仅说明算法的目的和最终结果。

在一个有权拓扑图中，节点之间的连线有距离，距离越小越好，任意两个节点间可能有多条路径，不同路径的距离不同。算法的目的是算出每个节点到其他任意节点的距离最短的路径。

左边拓扑图中，为简单考虑，节点间连线是双向的，即 $A \rightarrow B$ 和 $B \rightarrow A$ 的距离一样。

$A \rightarrow L$ 存在多条路径，路径及其距离如下：

- $A \rightarrow B \rightarrow D \rightarrow E \rightarrow L$: COST = 14
- $A \rightarrow E \rightarrow L$: COST = 5
- $A \rightarrow E \rightarrow G \rightarrow J \rightarrow L$: COST = 9

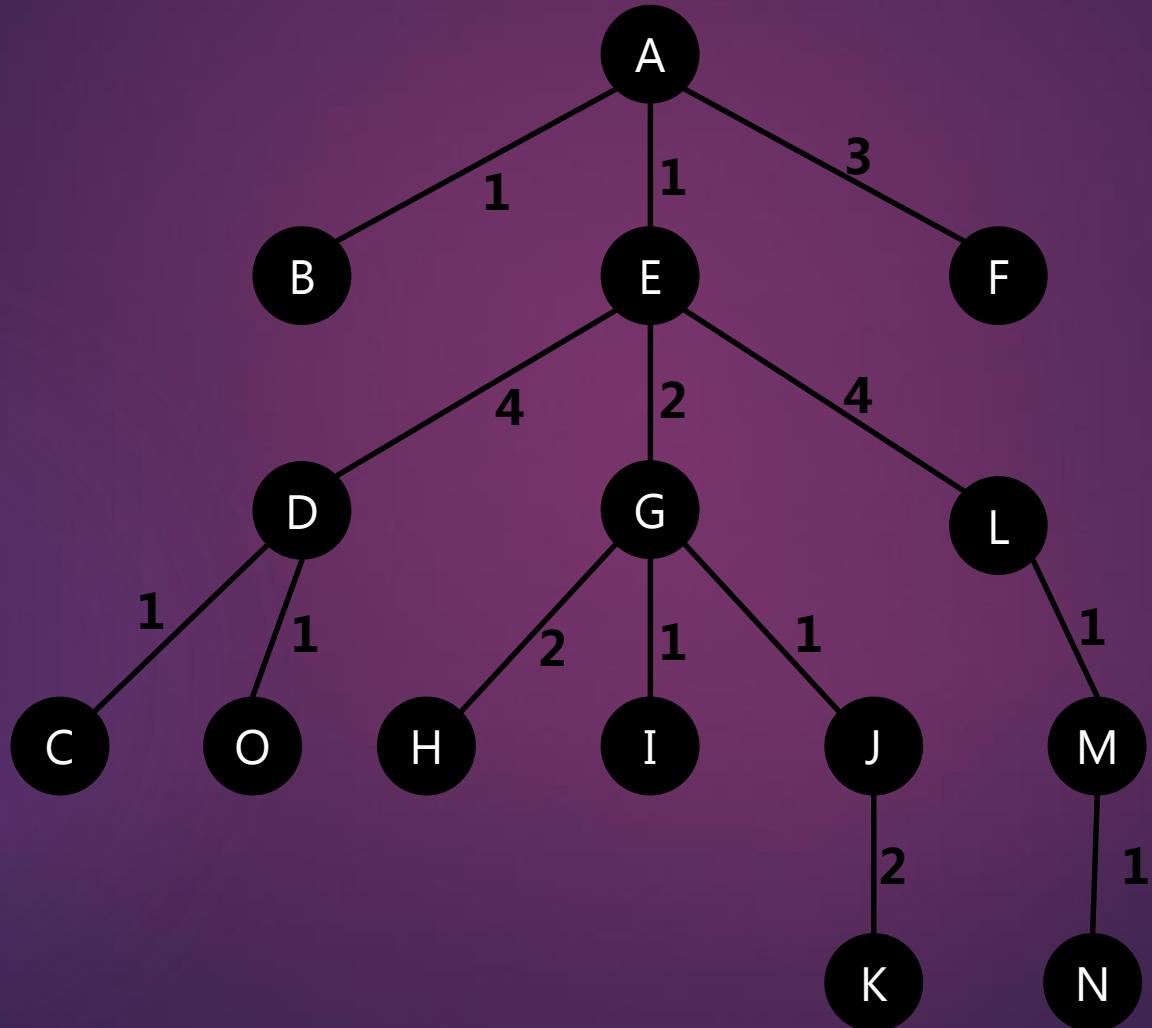
1.3 最短路径优先算法 (Dijkstra算法)

源节点→目的节点	路径	距离
A→B	A→B	1
A→C	A→E→D→C	6
A→D	A→E→D	5
A→E	A→E	1
A→F	A→F	3
A→G	A→E→G	3
A→H	A→E→G→H	5
A→I	A→E→G→I	4
A→J	A→E→G→J	4
A→K	A→E→G→J→K	6
A→L	A→E→L	5
A→M	A→E→L→M	6
A→N	A→E→L→M→N	7
A→O	A→E→D→O	6

Dijkstra算法会算出每个节点到其他任意节点的最短路径，但左表只给出节点 A 到其他所有节点的最短路径，其他节点自己可以类比得出。

下一页给出最短路径树供参考，和左表是一样的。

最短路径优先算法 (Dijkstra 算法)



自治系统(Autonomous System)

在互联网中，一个自治系统是一个有权自主决定在本系统中应采用何种路由协议的小型单位。这个网络单位可以是一个简单的网络也可以是一个由一个或多个普通的网络管理员来控制的网络群体，它是一个单独的可管理的网络单元（例如一所大学，一个企业或者一个公司个体）

一个自治系统将会分配一个全局的唯一的16位号码，有时我们把这个号码叫做自治系统（ASN）。

在一个自治系统中的所有路由器必须相互连接，运行相同的路由协议

自治系统之间的路由使用外部网关协议，例如BGP。

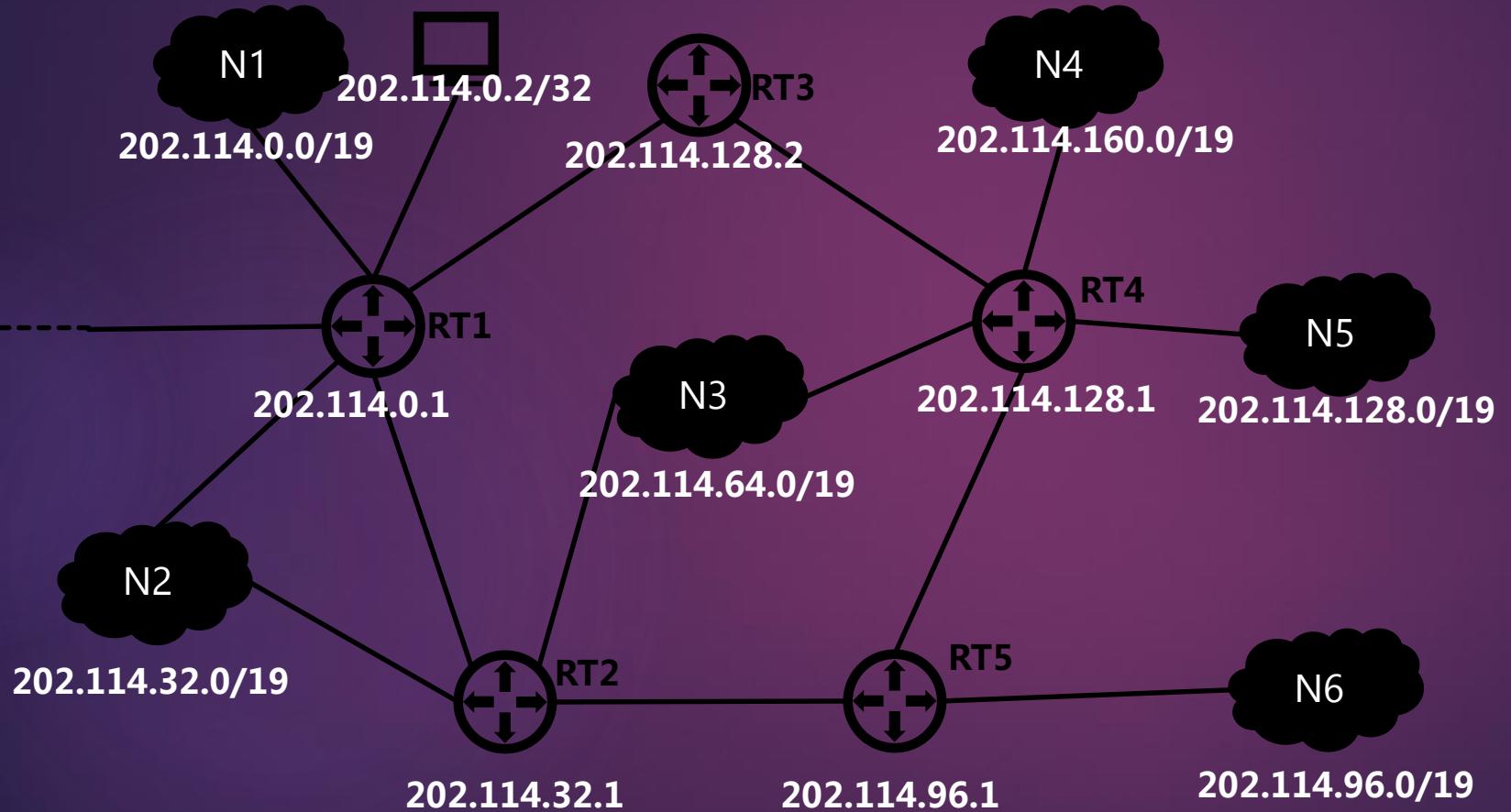
路由表是路由器工作的核心所在，由一条条路由组成。路由器根据网络报文的目的 IP 地址检索路由表，从而知道该报文应从哪个接口转发。

直连路由：直连路由不需要配置，当接口存在 IP 地址且状态正常时，由路由进程自动生成。特点是开销小，配置简单，无须人工维护，但只能发现本接口所属网段的路由。

静态路由：由管理员手动配置而成的路由称为静态路由。特点是无开销，配置简单，适合简单拓扑结构的网络；但网络发生故障后静态路由不会自动修正，必须人工排查。

动态路由：由动态路由协议发现的路由，当网络拓扑结构十分复杂时，配置静态路由工作量大且容易出错，这时适合使用动态路由协议发现路由。特点是开销大，配置复杂。

路由
来源



左图是一个自治系统网络示例，该网络构成一个子网号为 $202.114.0.0/32$ 的子网。

此网络有唯一出口，即通过路由器RT1的一个接口连接到外部网络。

为避免图显得混乱，这里没有标明链路开销。

下面几页即以此为例，解释路由表和路由过程。

1.5

路由 (RT1的路由表)

编号	目的地址/掩码	协议类型	优先级	开销	下一跳	出接口
1	0.0.0.0/0	Static	60	10	204.112.0.1	E0/1
2	202.114.0.0/19	Direct	0	0	-----	E0/2
3	202.114.0.2/32	Direct	0	0	-----	E0/3
4	202.114.32.0/19	Direct	0	0	-----	E0/4
5	202.114.128.2/32	Direct	0	0	-----	E0/5
6	202.114.32.1/32	Direct	0	0	-----	E0/6
7	202.114.64.0/19	OSPF	10	5	202.114.32.1	E0/6
8	202.114.96.0/19	OSPF	10	6	202.114.32.1	E0/6
9	202.114.128.0/19	OSPF	10	5	202.114.128.2	E0/5
10	202.114.160.0/19	OSPF	10	5	202.114.128.2	E0/5
11	202.114.96.1/32	OSPF	10	4	202.114.32.1	E0/6
12	202.114.128.1/32	OSPF	10	3	202.114.128.2	E0/5
13	202.114.128.0/19	OSPF	10	5	202.114.128.1	E0/5

路由来源：RT1路由表中路由的来源可以是多样的，同时存在静态路由，直连路由和动态路由协议发现的路由。

在一个比较复杂的网络中不推荐全部使用静态路由，然而在某些比较理想的网络下，配置若干条关键的静态路由可以为路由器节省大量开销。如此例的自治系统出接口只有一个，所有由此网络发送到外部网络的报文经过RT1时不存在选路的问题，直接配置一条静态路由就省了许多事。

动态路由的来源可以不仅仅是OSPF，可以是任何动态路由协议。

路由开销：图中的开销是从本路由器到目的网络的整条路径的开销，不要误解为到下一跳的开销。

最长匹配原则：路由匹配时遵从最长匹配原则，如目的IP 202.114.0.2进行路由匹配时，有两条匹配的路由项，分别是2和3，然而3比2匹配的更精确，故路由器将按照3的接口进行转发。

默认路由：当路由器根据目的 IP 地址检索此路由表找不到匹配项时，将丢弃IP包。如果配置了默认路由，将使用默认路由，即路由1，这是一条静态路由。实际上路由1能够和任何地址进行匹配。

路由迭代：当所匹配的路由项的下一跳地址不和此路由器直接相连，路由器还需要对路由表进行迭代查找，直到找出最终的下一跳，这叫路由迭代。例如路由13的下一跳路由器是RT4，不和RT1直连，所以继续根据下一跳IP查找路由，找到路由12，从E0/5接口转发。

静态路由： 1

直连路由： 2~6

动态路由： 7~13

□ 直连路由开销均为0，不能更改

根据掩码长度不同，可以把路由项分为以下类型：

- **主机路由**：掩码长度是32位的路由，标明此路由匹配单一IP地址，路由（3,5,6,11,12）
- **子网路由**：掩码长度小于32但大于0，标明此路由匹配一个子网，路由（2,4,7,8,9,10,13）
- **默认路由**：掩码长度为0，标明此路由匹配全部IP地址，路由（1）

路由优先级

不同路由协议考虑的因素不同，会出现到同一目的地址存在多条不同路由的情况，这时候路由器需要依据路由优先级选择将哪一条路由加入路由表。优先级越高，其数值越小。下表是H3C路由器默认优先级，供参考。

路由来源	优先级	路由来源	优先级
Direct	0	OSPF ASE	150
OSPF	10	OSPF NSSA	150
IS-IS	15	IBGP	255
STATIC	60	EBGP	255
RIP	100	UNKNOWN	255

- 直连路由优先级永远是0
- 各动态路由优先级可以手工配置
- 每条静态路由优先级可以不同

1.6

路由协议分类

IGP (Interior Gateway Protocol , 内部网关协议) :

在自治系统内部运行，常见的IGP协议包括RIP , OSPF , IS-IS

EGP (Exterior Gateway Protocol , 外部网关协议) :

运行于不同自治系统之间 BGP是最常见的

单播路由协议 : 包括RIP , OSPF , BGP和 IS-IS 等

组播路由协议 : 包括PIM-SM , PIM-DM等

根据作用
范围划分

根据使用
算法划分

根据目的
地址类型
划分

根据 IP
协议版本
划分

距离矢量协议 (Distance-Vector) : 包括 RIP 和 BGP

链路状态协议 (Link-State) : 包括 OSPF 和 IS-IS

IPv4路由协议 : 包括RIP , OSPF , BGP和 IS-IS 等

IPv6路由协议 : 包括OSPFv3 , IPv6 BGP和IPv6 IS-IS等

- 1. OSPF
- 2. 本地回环接口
- 3. Router-ID
- 4. 链路开销
- 5. OSPF 的四种网络类型

2.1 OSPF (Open Shortest Path First)

OSPF 是由 IETF 开发的基于链路状态的自治系统内部路由协议，
目前通用的 OSPF 协议第二版由 RFC2328 定义。

OSPF 特点：

- 适应范围广：支持各种规模的网络，最多可支持几百台路由器。
- 快速收敛：网络拓扑结构发生变化后立即发送更新报文，使这一变化在自治系统中同步。
- 无自环：最短路径优先算法保证了不会生成路由环路。
- 区域划分：允许自治系统的网络被划分成区域来管理。路由器链路状态数据库的减小降低了内存的消耗和CPU的负担；区域间传送路由信息的减少降低了网络带宽的占用。
- 等价路由：支持到同一目的地址的多条等价路由。
- 路由分级：使用4类不同的路由，按优先顺序来说分别是：区域内路由、区域间路由、第一类外部路由、第二类外部路由。
- 支持验证：支持基于接口的报文验证，以保证报文交互和路由计算的安全性。
- 组播发送：在某些类型的链路上以组播地址发送协议报文，减少对其他设备的干扰。
- 网络开销小，路由器CPU和内存开销大，适合企业中小型网络
- 直接使用 IP 包传递数据，协议号为89。由于 IP 协议本身无连接不可靠，所以可靠性由协议自身满足。

RFC
2328

OSPF 定位：

- 内部网关协议 (IGP)
- 链路状态协议
- 单播路由协议
- 使用最短路径优先算法计算路由

本地回环接口（Loopback Interface）

本地回环接口特点：

- loopback口是给路由器赋予一个具有IP地址的逻辑接口，这个接口的特点是总是up，不会随着物理接口的状态而变化
- 环回接口由于独占一个IP地址，子网掩码一般建议设为255.255.255.255
- 路由器默认没有任何环回接口，但是它们很容易创建，在网络设备上可以通过配置命令来创建一个或多个环回接口
- 在一个网络中，不同设备的环回接口地址以及同一设备上的不同环回接口地址应该统一规划，避免重复

在Windows系统中，采用127.0.0.1作为本地回环地址，此地址不能向主机外发送报文。

可允许运行在同一台主机上的程序和服务器程序通过TCP/IP进行通讯。

建立BGP邻居：因为loopback口会一直保持UP，BGP会话中使用loopback口可以提高网络的健壮性。

路由器管理地址：管理员完成网络规划之后，为了方便管理，会为每一台路由器创建一个loopback接口，并在该接口上单独指定一个IP地址作为管理地址，管理员会使用该地址对路由器远程登陆（telnet），该地址实际上起到了类似设备名称一类的功能。

建立BGP邻居：因为loopback口会一直保持UP，BGP会话中使用loopback口可以提高网络的健壮性。

作为Router-ID：OSPF协议中，要求自治系统内每台路由器有个唯一的标识，即RouterID，32位无符号整数。loopback口恰好满足自治系统内的唯一性，也是32位的，故作为RouterID是最佳选择。

建立虚拟隧道：在建立IPSec或GRE之类的虚拟隧道时，使用loopback接口可以保证整个隧道的稳定性。

在自治系统内，所有路由器之间要相互通信，每个路由器必须有自己唯一的标识，即Router-ID，一个32位无符号整数，OSPF 路由器发出的链路状态都会写上自己的Router-ID。每一台OSPF路由器只有一个Router-ID，Router-ID 使用 IP 地址的形式来表示。

确定Router-ID的流程：

1. 手工指定Router-ID
2. 如果当前设备配置了Loopback接口，将选取所有Loopback接口上数值最大的 IP 地址作为Router-ID
3. 如果没有活动的Loopback接口，则选择活动物理接口 IP 地址最大的
4. Router-ID 只在 OSPF 启动时计算，或者重置 OSPF 进程后计算

链路开销 (COST)

OSPF 协议使用接口带宽来计算COST，不同路由协议计算出的COST不存在比较关系。

计算COST的规则：

1. 计算方式：用10000 0000除以以bit为单位的接口带宽
2. COST值和接口带宽成反比，带宽越高，COST值越小
3. COST值必须为整数，所以最小为1，不能整除的进行取整
4. 如果路由要经过两跳才能到达目标网络，那么很显然，两跳COST值要累加起来，才算 是到达目标网络的路由的COST值
5. OSPF会自动计算接口上的COST值，但也可以通过手工指定该接口的COST值，手工指 定的优先于自动计算的值
6. 到达目标相同COST值的路径，可以执行负载均衡，最多6条链路同时执行负载均衡
7. COST是单向的，即两台路由器之间，A到B与B到A的COST值可能不一样

两种外部路由距离：

OSPF支持两种外部路由距离，类型1与 OSPF距离使用同样的计量单位，处理起 来与OSPF内部路由信息一样；类型2的距 离被认为大于AS任何内部的路径距离， 处理时不会与AS内部路径距离叠加，因 而不受AS内部路由信息的影响，而仅仅 是选择其数值较小者。在一个 AS 中，类型 1 和类型 2 的外部距离可以同时存在。这 时，类型 1 将始终被优先选择。

OSPF 的四种网络类型

OSPF 协议中，根据路由器间连接方式的不同，可以分为四种类型的网络。
四种网络的特点，都以运行 OSPF 协议为前提。

	点对点网络 (Point-to-point)
	广播网络 (Broadcast)
	NBMA网络 (Non-Broadcast Multi-Access , 非广播多路访问)
	点对多点网络 (Point-to-MultiPoint)

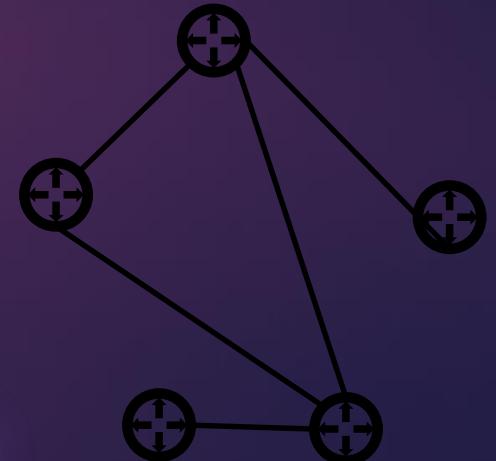
点对点网络

点对点网络 (Point-to-point) : 两台路由器之间通过专线连接而组成的网络。当链路层协议是PPP、HDLC时，OSPF缺省认为网络类型是P2P。

点对点网络特点：

- 以组播形式 (224.0.0.5) 发送协议报文。
- 自动发现邻居
- 不选举DR/BDR
- 网络上的有效邻居总是可以形成邻接关系
- hello时间间隔为10s

右图路由器之间都有一条物理链路进行连接，它们构成了一个点对点网络。

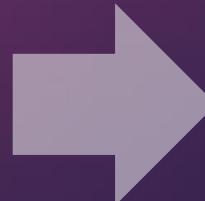


广播网络 (Broadcast) : 路由器共享链路时即构成广播网络。当链路层协议是Ethernet (以太网) 、 FDDI 时 , OSPF 缺省认为网络类型是 Broadcast 。

广播网络特点 :

- 广播网络上的每一对路由器都被认为可以直接通讯
- 通常以组播形式 (224.0.0.5 和 224.0.0.6) 发送协议报文
- 自动发现邻居
- 选举 DR/BDR
- hello 时间间隔为 10s

右图路由器连接在一条总线上 , 每两台路由器都可以直接通讯 , 一台路由器发出的广播报文可被其他所有路由器接收 , 此即广播网络。



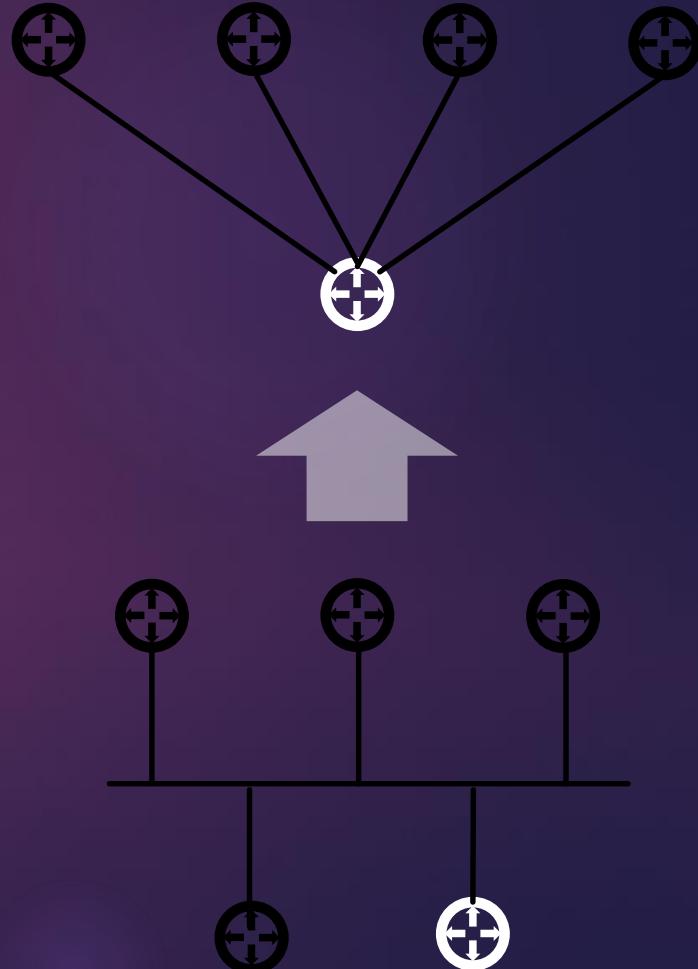
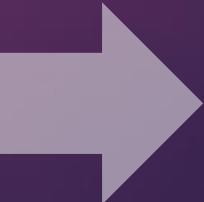
NBMA网络

NBMA网络 (Non-Broadcast Multi-Access , 非广播多路访问) : 路由器共享链路 , 但此网络却不能发送广播报文 , 只能通过单播报文来模拟广播网络。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。

NBMA网络特点 :

- 没有广播能力
- 不能自动发现邻居 , 要在DR上配置所有其他路由器的地址信息 , 由DR根据此地址表向其他所有路由器发送单播报文来模拟广播效果 , 从而发现邻居。
- 选举DR/BD
- hello时间间隔为30s

拓扑结构和广播网一样 , 但由于链路层协议的缘故 , 此网络不支持发送广播报文 , 所以只能DR根据配置好的地址表轮流发送单播报文 , 达到模拟广播的效果 , 此即NBMA网络。其运行时的逻辑拓扑见上图 , 其实也是邻居关系图。



2.5

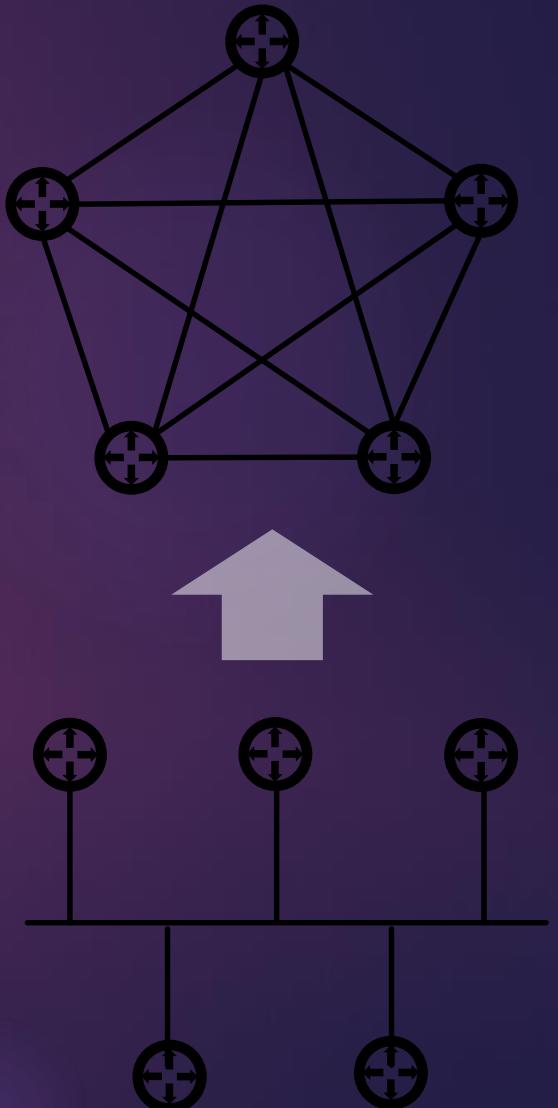
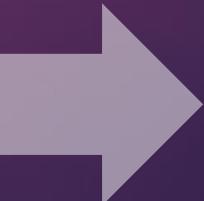
点对多点网络 (Point-to-MultiPoint)

点对多点网络 (Point-to-MultiPoint) : 没有一种链路层协议会被缺省的认为是P2MP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。

点到多点网络特点：

- 没有广播能力
- OSPF认为每两个路由器之间的连接都是点对点连接
- 以组播224.0.0.5地址发送协议报文
- 自动发现邻居
- 不选举DR/BDR
- hello时间间隔为30s

拓扑结构和NBMA网络一样，因为每两个路由器之间都直接相连，所以此网络逻辑上可以看成上面两两互连的点对多点网络，其运行方式和点对点网络一样。



NBMA & 点对多点网络

NBMA & 点对多点网络是 OSPF 中同一网络类型（非广播网络）的两种不同的运行模式，到底选用哪一种取决于 OSPF 的配置。

NBMA 网络：

- 一个非常重要的限制，即要求NBMA网络的每台路由器直连，这在很多情况下是无法满足的，如右图。
- 当出现上面所说有两台路由器无法直连的情况时，可以把此网络划分为多个NBMA网络，但这无疑增加了管理负担，也容易出错。
- 这时适合使用点对多点模式，其逻辑拓扑如下面的图。

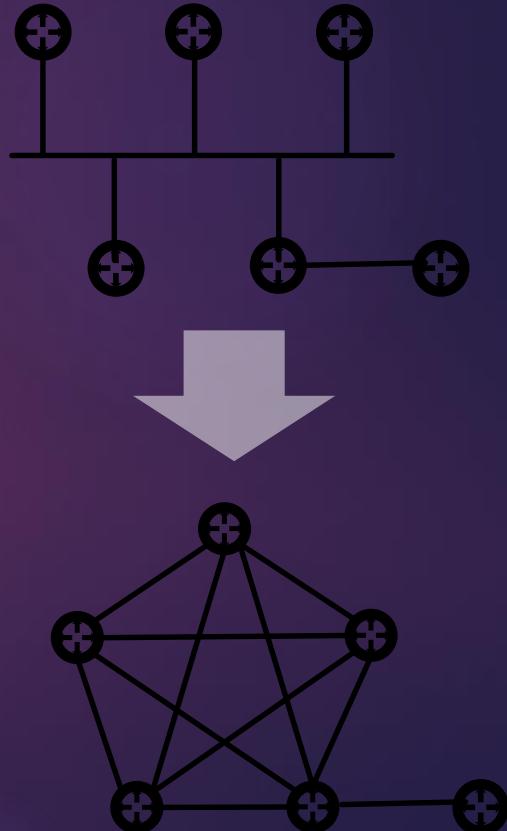


点对多点网络：

- 这种模式的网络最大的缺点是开销太大。图中轻易可见路由器间连接数量太多，从而邻居关系也多，导致交换链路信息时网络开销大。
- 相比之下，NBMA的连接数量就少得多，故效率也是最高的。



点对多点网络何以能够组播？
那为何NBMA网络就不组播呢？



OSPF 过程解析

- 1. OSPF 过程概览
- 2. 划分区域
- 3. hello报文&发现邻居
- 4. 选举DR/BDR
- 5. 建立邻接
- 6. 交换链路状态
- 7. 计算路由

3.1

OSPF 过程概览



划分区域：AS可能路由太多，如果每个路由器都保存一份相同的AS整网链路信息，无疑开销太大，宜采用分治思想划分区域，区域内享有一定程度的自治。

选举DR/BDR：在广播网和NBMA网络中，路由器是两两互连的，在交换链路信息时这种两两互连的邻接关系会导致大量无谓的开销。选举DR/BDR会让邻接关系成为星形拓扑，提升效率。

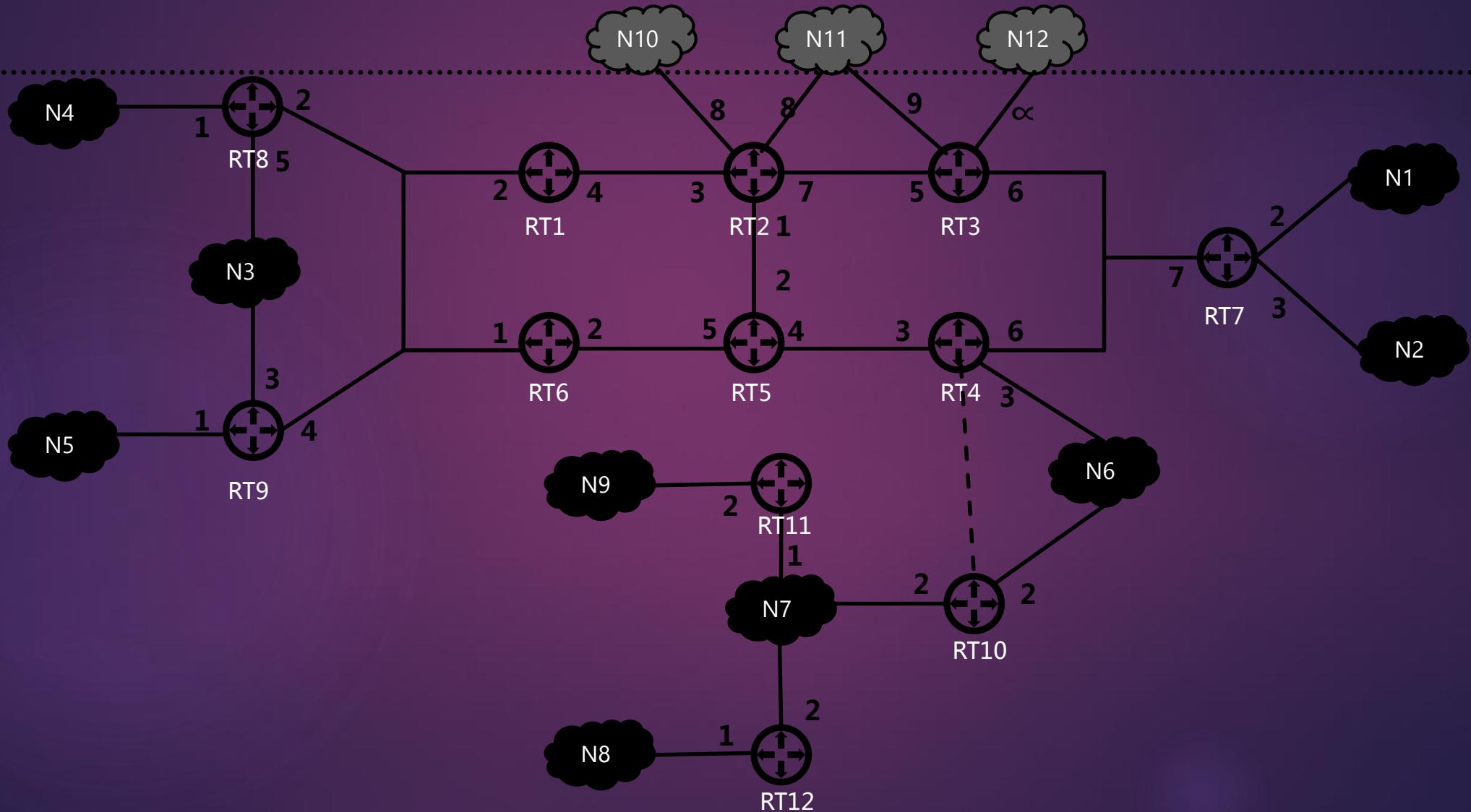
发现邻居：周期性的向周围发送hello报文，发现邻居。

建立邻接：只有邻居路由器才有可能建立邻接关系，具有邻接关系的路由器才可以互相交换链路信息。

交换链路状态：邻接路由器之间互相通告自己拥有的链路状态，最终导致所有路由器拥有相同的AS整网链路状态数据库，这个数据库即描绘了整个AS的拓扑信息（如果没有划分区域的话）。

计算路由：每个路由器根据自己的链路状态数据库，采用最短路径优先算法，以自己为根结点，计算出到所有网络的路由。

构造AS网络示例



上一页的网络图描述的是一个AS，本章后续所有 OSPF 协议运行过程的阐述都以此图为例



灰色的网络图标表示是AS外部的网络，不属于本AS

有一条到AS外部网络的链路开销为 ∞ ，说明此COST是第二种外部路由距离。

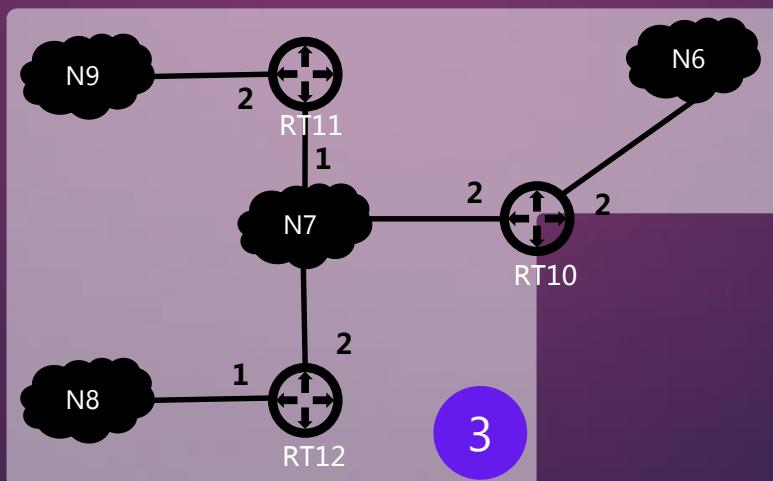
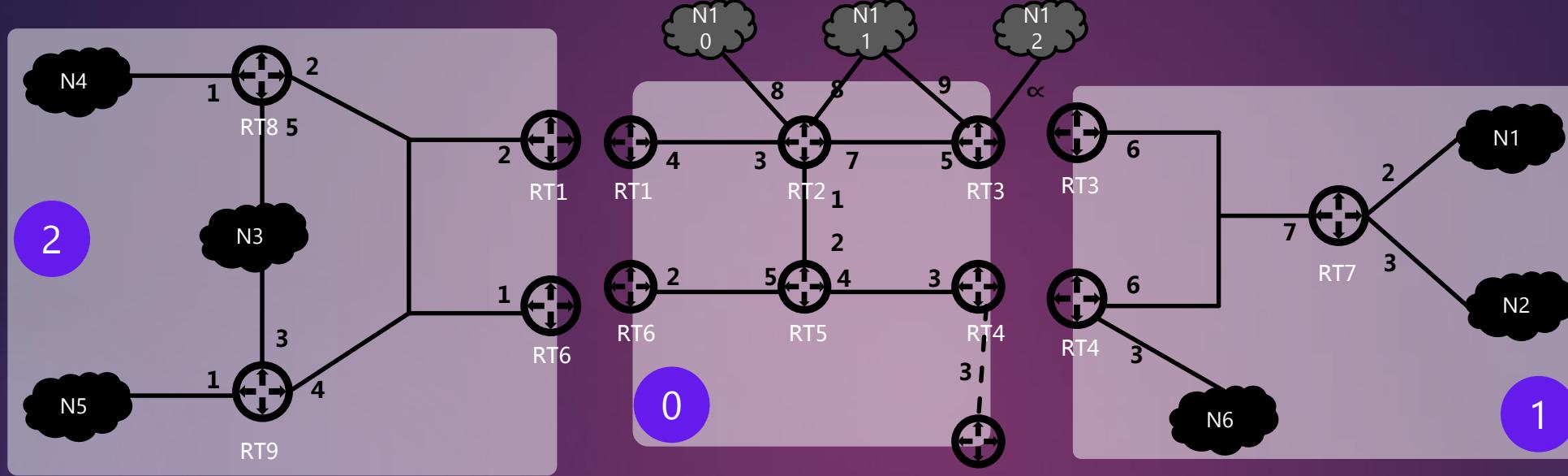
所有路由器都有路由器标号，网络也有网络标号，每条链路都标明有COST值。

为了使图不显得混乱拥挤，这里并没有给出 IP 地址，但足以供后面阐述。

RT4 与 RT10 不直接相连，但它们之间建立了一条虚链接，所以它们可以成为邻居或邻接关系。

3.2

划分区域



- 0 骨干区域 , 点对点网络
- 1 区域1 , 广播网络
- 2 区域2 , NBMA网络
- 3 区域3 , P2MP网络

划分区域

此AS可以划分为 4 个区域。AS并不是一定要划分区域的，这样链路信息交换完毕后每个路由器都有一份一模一样的数据库，此数据库包含整个AS的拓扑信息。但当AS很大的时候，划分区域会减少很大开销，每个路由器的链路信息数据库会有差别，也比不划分区域时要小得多。

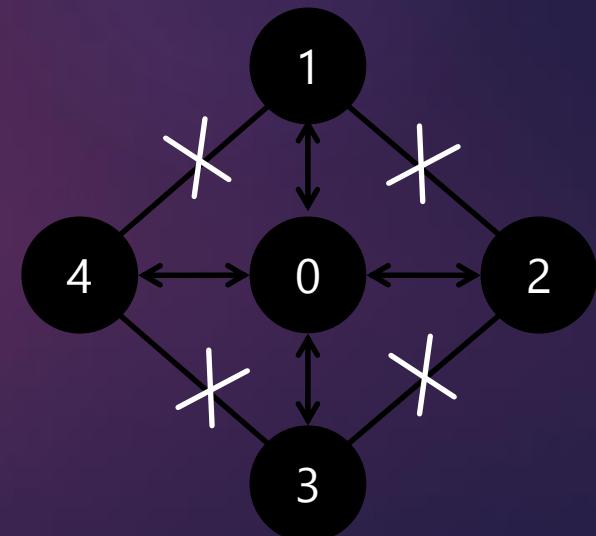
区域号 (Area ID)：每个区域用区域号标识，骨干区域的区域号必须是 0

在本例中，一个区域只有四种网络类型中的一种，这是纯粹是为了保持简单。然而一个区域是可以同时存在多种不同的网络类型的，可以按需划分。

本例中每一个区域都很小，在实际组网中一个区域可能包含更多路由器。

Hub-Spoke拓扑架构：区域的划分为了能够尽量设计成无环网络，所以采用了Hub-Spoke的拓扑架构，也就是采用核心与分支的拓扑。所有其他常规区域与骨干区域（核心）直接相连，常规区域之间不能直接通信，必须通过骨干区域交换链路信息。

骨干区域不能被分隔，即不能有骨干路由器和其他骨干路由器孤立，不然AS 中的一部分网络就会变为不可到达。骨干区域的分隔可以通过配置虚链接来修复。



划分区域

四种类型路由器：

划分区域前，AS所有路由器功能一样，地位等同，包含的链路信息数据库也一样。划分区域后，路由器将分为如下四种类型。

区域内路由器（IR : Internal Router）：该类路由器的所有接口都属于同一个OSPF区域

区域边界路由器（ABR : Area Border Router）：该类路由器可以同时属于两个以上的区域，但其中一个必须是骨干区域。ABR用来连接骨干区域和非骨干区域，它与骨干区域之间既可以是物理连接，也可以是逻辑上的连接。

骨干路由器（BR : Backbone Router）：该类路由器至少有一个接口属于骨干区域。因此，所有的ABR和位于Area0的内部路由器都是骨干路由器。

自治系统边界路由器（ASBR : AS Border Router）：与其他AS交换路由信息的路由器称为ASBR。ASBR并不一定位于AS的边界，它有可能是区域内路由器，也有可能是ABR。只要一台OSPF路由器引入了外部路由的信息，它就成为ASBR。

RT : 2, 5, 7, 8, 9, 11, 12

RT : 1, 3, 4, 6, 10

RT : 1, 2, 3, 4 ,5, 6, 10

RT : 2, 3

3.3

Hello报文 & 发现邻居

每台路由器周期性的向四周发送Hello报文以发现邻居，维持邻接。P2P和P2MP网中，路由器间N条连接可形成N对邻居；广播网中，N台路由器可形成 $N(N-1)/2$ 对邻居；NBMA网中，N台路由器可形成 $2(N-1)-1$ 对邻居。

0	7	15	31
Version	1	Packet length	
		Router ID	
		Area ID	
Checksum		AuType	
	Authentication		
	Authentication		
	Network Mask		
HelloInterval	Options	Rtr Pri	
	RouterDeadInterval		
	Designated router		
	Backup designated router		
	Neighbor		
	...		

主要字段解释	
Router ID	路由器ID
Area ID	区域号
Authentication	认证信息
Network Mask	发送Hello报文的接口所在网络的掩码，若相邻两台路由器掩码不同，则不能建立邻居关系。在无编号点对点网络和虚拟连接上，该域应当被设为 0.0.0.0。
HelloInterval	发送Hello报文的时间间隔。若相邻两台路由器Hello间隔时间不同，则不能建立邻居关系。
Rtr Pri	路由器优先级。如果设置为0，则该路由器接口不能成为DR/BDR
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。如果相邻两台路由器的失效时间不同，则不能建立邻居关系。
DR	指定路由器的接口的IP地址
BDR	备份指定路由器的接口的IP地址
Neighbor	一系列邻居路由器的Router ID

3.3

Hello报文 & 发现邻居—NBMA

NBMA网络不具备广播能力，无法自动发现邻居，需要在每台可能成为DR的路由器上静态配置所有其他路由器的信息。

NBMA网络发现邻居流程：

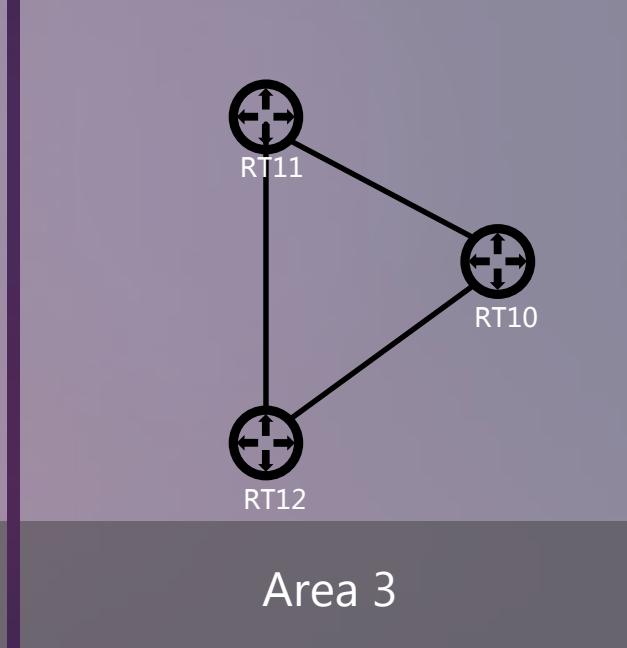
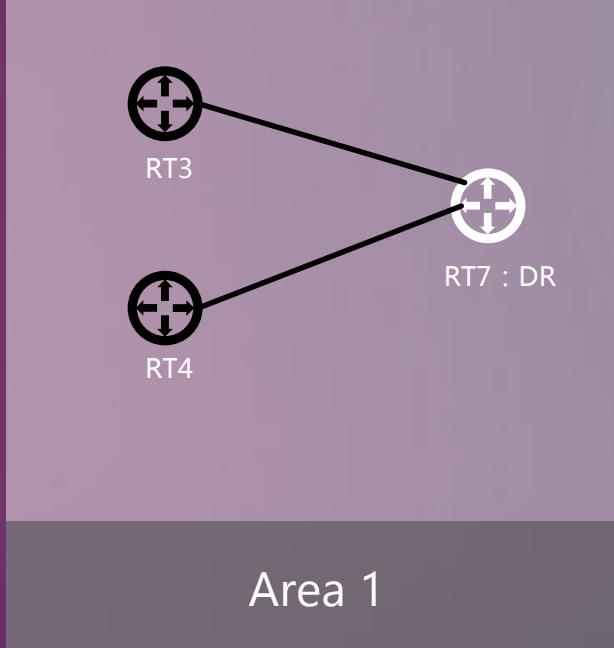
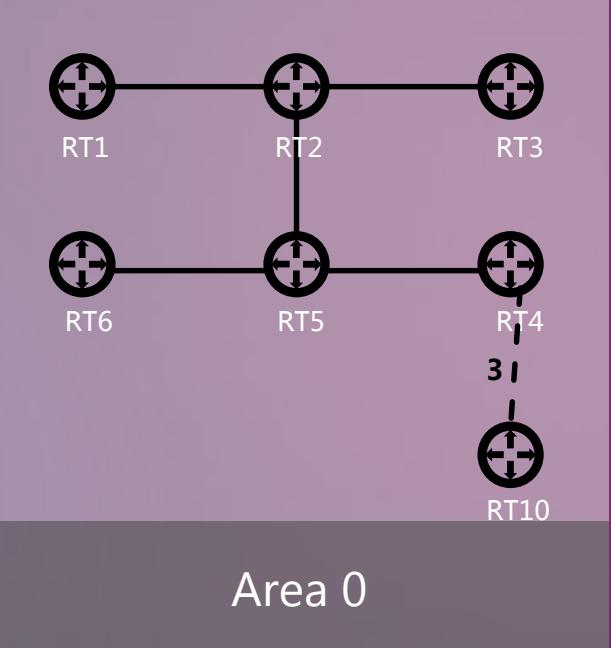
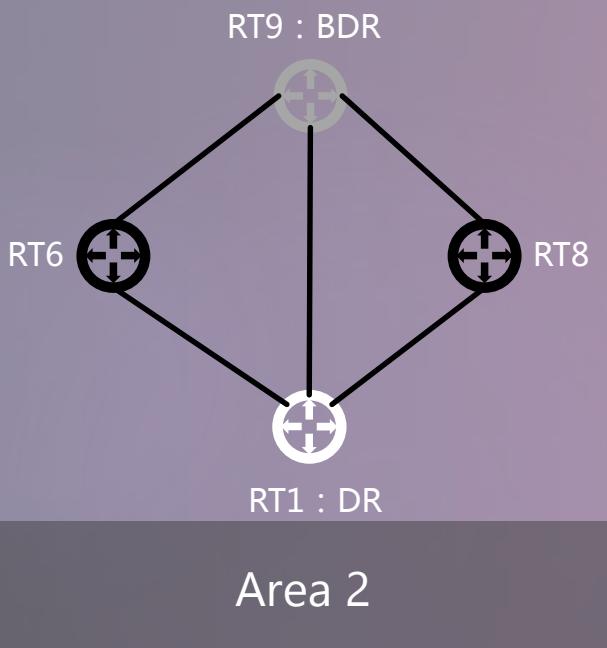
1. 每台有能力成为DR/BDR的路由器之间周期性单播发送hello报文
2. 有能力成为DR/BDR的路由器之间依次通过hello报文选举出DR/BDR
3. DR和BDR周期性的向所有其他路由器单播发送hello报文
4. 一般路由器需要周期性的向DR/BDR单播发送hello报文
5. 邻居关系建立完毕

- 所有有能力成为DR/BDR的路由器间始终交换 hello 包
- 为了减少 hello 包的发送，应当控制 NBMA 网络上有能力成为 DR/BDR 的路由器数量

Area 2 中， RT1上配置的其他路由器信息列表	
路由器接口IP地址	是否可能成为 DR/BDR
RT6	否
RT8	否
RT9	是

3.3

Hello报文 & 发现邻居—邻居图



选举DR/BDR

区域1 是广播网，区域2 是NBMA网，这两个区域中，认为每两台路由器都可以形成邻接关系，但这样是很不必要的，同一链路信息会重复传递多次，会导致处理器和网络资源的浪费。故需要选举DR/BDR以提升效率。

DR/BDR选举过程：

- 初始阶段，hello报文的 DR和BDR字段都填 0.0.0.0，即表示没有DR/BDR
- 选举BDR
 - 1. 如果有多个路由器的 hello 报文宣告自己为BDR，则按右边规则筛选
 - 2. 如果没有路由器宣告自己为BDR，则按右边规则从所有有资格的路由器中筛选
- 选举DR
 - 1. 如果有多个路由器的 hello 报文宣告自己为DR，则按右边规则筛选
 - 2. 如果没有路由器宣告自己为DR，则选当前BDR为DR，并重新选举BDR
 - 3. 选举BDR时，当前DR不会参与选举
- 如果路由器接口优先级配置为0，则没有资格成为DR/BDR
- 被选择的 DR 不一定是拥有最高优先级的路由器；被选择的 BDR 也不一定是第二高优先级的路由器
- 当DR/BDR已经选取完毕，就算一台具有更高优先级的路由器变为有效，也不会替换该网段中已经选取的 DR/BDR成为新的DR/BDR
- 如果路由器 X 是网络上唯一可能成为 DR 的路由器，它将选择自己为 DR，而没有 BDR

DR/BDR筛选规则：

- 1. 比较路由器接口优先级，大者胜（优先级范围0~255）
- 2. 优先级相同则比较Router ID，大者胜

BDR：因为DR选举过程复杂，DR失效后很长时间无法传递链路信息，所以需要选举出BDR，即Backup DR（备份DR），以保证高可靠性，当DR失效后BDR能立即升级接替DR的工作。

建立邻接

只有邻接路由器之间才可以交换链路信息，路由器间具有邻居关系是前提条件，然而并非所有邻居路由器都可以形成邻接关系。

在邻居状态机中有两处需要作出是否形成邻接的判断：

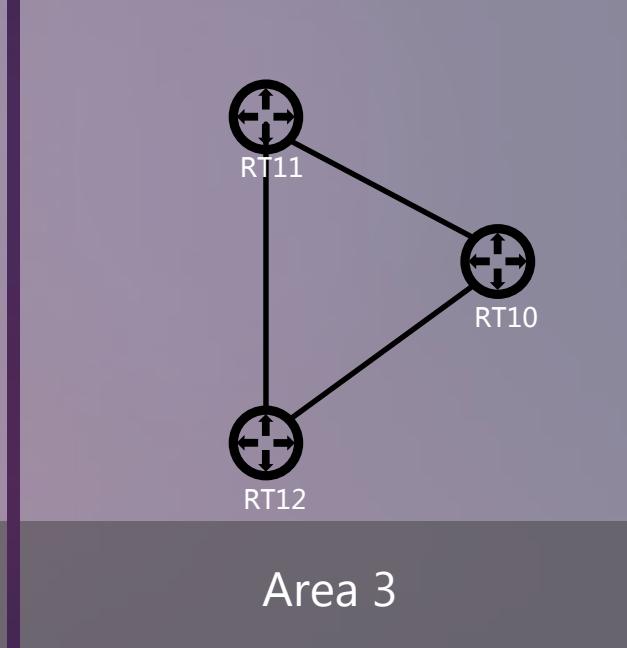
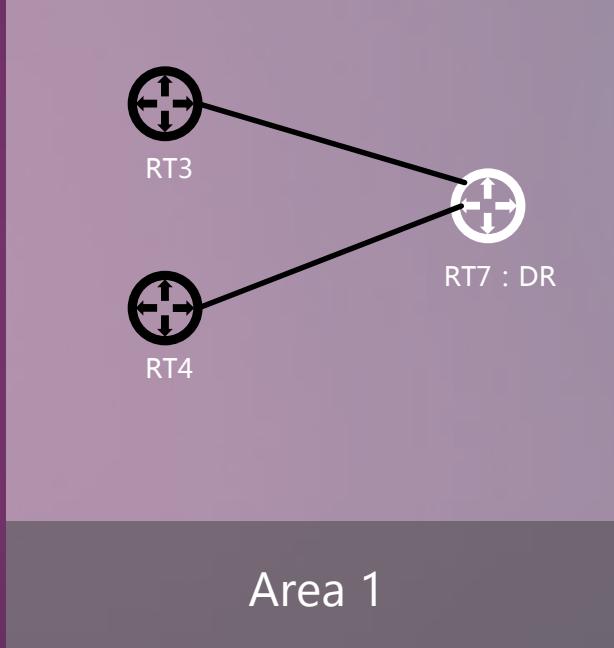
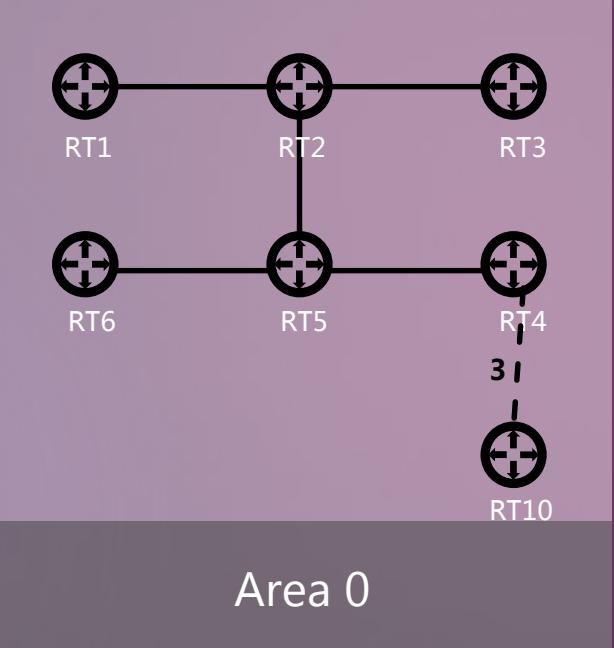
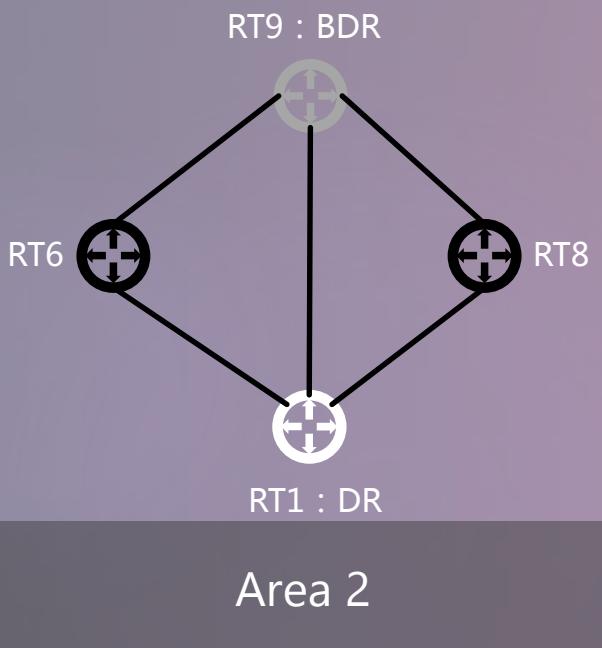
- 当与邻居的双向通讯初次建立
- 当所接入网络上的 DR (BDR) 发生改变

邻居路由器间建立邻接的规则：

- 网络类型为点对点网络或点对多点网络
- 网络广播网和NBMA网络上，所有路由器和DR/BDR形成邻接
- 路由器间由虚拟通道连接

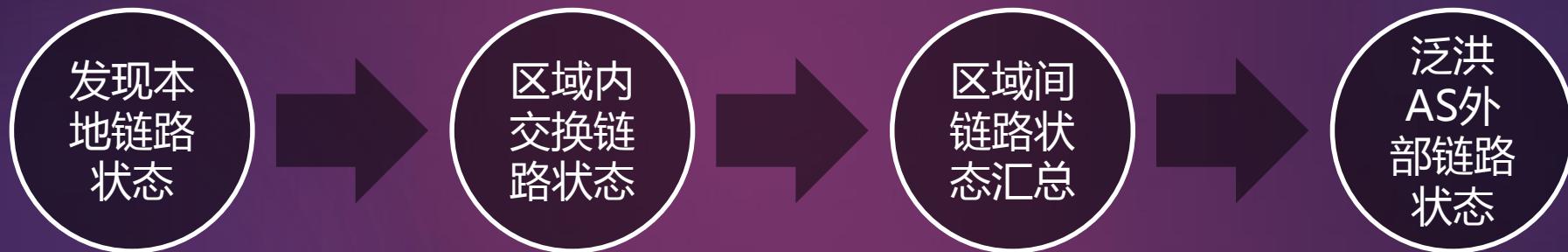
3.5

建立邻接—邻接图



3.6

交换链路状态



3.6

交换链路状态—发现本地链路状态

路由器接口只要正常工作，就可以自动发现“本地链路状态”。

为方便只介绍原理，这里只关注链路状态的三种属性：

- 本路由器
- 对端网络或或路由器
- COST值

“本地链路状态”在正规文档里面并没有定义这个词语，只是我在描述一开始与路由器直接相连的链路状态时，苦于没有现成的词语定义，故杜撰了这个词。希望不要受到误导。直连路由即根据“本地链路状态”而得。

RT6	2
RT8	2
RT9	2

RT1

RT1	1
RT8	1
RT9	1

RT6

RT1	2
RT6	2
RT9	2
N4	1
N3	5

RT8

RT1	4
RT6	4
RT8	4
N3	3
N5	1

RT9

3.6

交换链路状态—区域内交换链路状态

区域内每台路由器都有自己的“本地链路状态”，接下来需要通过洪泛，互通有无，让区域内的所有路由器都构建自己的链路状态数据库；区域内所有路由器的链路状态数据库一样，且完整描述了整个区域的链路状态。

		from													
		RT1	RT2	RT3	RT4	RT5	RT6	RT10		RT10	RT11	RT12			
to	RT1		3			1			RT10	1	2				
	RT6	2		5					RT10	1	2				
	RT8	2	1		6				RT11	2		2			
	RT9	2	1	2		4			RT12	2	1				
	N3			5					N6	2					
	N4			1					N7	2	1	2			
	N5				1				N8			1			
	RT10							2	N9	2					
	N6														

Area2
Area0
Area1
Area3

3.6

交换链路状态—链路状态汇总

区域内链路状态交换完毕，这时ABR同时拥有一般区域和骨干区域的链路状态数据库。链路状态汇总分下面两步：

1. ABR将一般区域的链路状态精简后洪泛至骨干区
2. ABR将骨干区链路状态数据库精简后在一般区域洪泛

		from			
		RT3	RT4	RT1	RT6
to	N1	8	8	N3	5
	N2	9	9	N4	3
	N6	9	3	N5	3
Area1		Area2		Area3	

各ABR需要向骨干区洪泛的区域内链路状态汇总信息

	RT3	RT4		RT1	RT6
N3	13	12	N1	17	14
N4	11	10	N2	18	15
N5	11	10	N6	12	9
N7	11	5	N7	14	11
N8	12	6	N8	15	12
N9	13	7	N9	16	13

ABR向各区域洪泛骨干区链路状态汇总信息

第一步洪泛完毕，骨干区将拥有到AS中所有目的网络的链路状态信息。

第二步洪泛完毕，各区域将拥有到AS中所有目的网络的链路状态信息。



咦！为什么不向区域3洪泛链路状态汇总信息？



因为区域3被配置为存根区域，所以特殊对待！

3.6

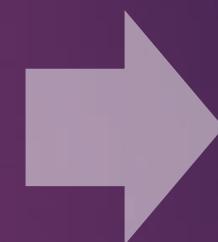
交换链路状态—链路状态汇总

汇总后 Area 1
链路状态数据库



	RT3	RT4	RT7
RT3		6	7
RT4	6		7
RT7	6	6	
N1			2
N2			3
N6		3	
N3	13	12	
N4	11	10	
N5	11	10	
N7	11	5	
N8	12	6	
N9	13	7	

汇总后 Area 2
链路状态数据库



	RT1	RT6	RT8	RT9
RT1		1	2	4
RT6	2		2	4
RT8	2	1		4
RT9	2	1	2	
N3			5	3
N4			1	
N5				1
N1	17	14		
N2	18	15		
N6	12	9		
N7	14	11		
N8	15	12		
N9	16	13		

交换链路状态—洪泛外部链路状态

外部链路状态会被原样洪泛至整个AS（除存根区域）：

1. 外部链路状态洪泛至骨干区
2. 外部链路状态洪泛至所有一般区域

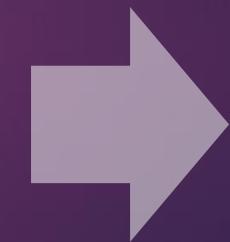
到ASBR的链路状态也会被一并洪泛

在骨干区洪泛的
外部链路状态



	from	
	RT2	RT3
N10	8	
N11	8	9
N12		∞

在Area2 洪泛的
外部链路状态



	RT1	RT6	RT2	RT3
RT2	4	4		
RT3	11	11		
N10			8	
N11			8	9
N12				∞

3.6

交换链路状态—洪泛外部链路状态

最终 Area 1 链
路状态数据库



	RT3	RT4	RT7	RT2
RT2	5			
RT3		6	7	
RT4	6		7	
RT7	6	6		
N1			2	
N2				3
N6		3		
N3	13	12		
N4	11	10		
N5	11	10		
N7	11	5		
N8	12	6		
N9	13	7		
N10				8
N11				8
N12		α		

最终 Area 2 链
路状态数据库



	RT1	RT6	RT8	RT9	RT2	RT3
RT1		1	2	4		
RT6	2		2	4		
RT8	2		1		4	
RT9	2		1	2		
RT2	4		4			
RT3	11	11				
N3				5	3	
N4				1		
N5					1	
N1	17		14			
N2	18		15			
N6	12		9			
N7	14		11			
N8	15		12			
N9	16		13			
N10					8	
N11					8	9
N12						α

交换链路状态—存根区域 (Stub Area)

OSPF允许将某些区域配置为存根区域，当有如下网络条件时，存根区域路由器的链路状态数据库会大大缩小，从而节省内存开销：

- AS链路状态数据库主要包含的是外部链路状态
- 区域只有单一出口，即一个ABR，或者不需要按每条外部路径来选择离开区域的出口

存根区域的限制条件：

- 在存根区域中不能配置有虚拟通道
- ASBR 也不能存在于存根区域中

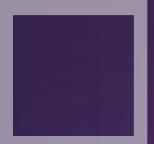
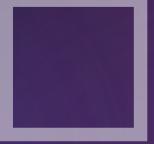
存根区域内必须使用默认路径，默认路径是在区域内洪泛汇总链路状态时，由一台或多台ABR宣告并在此区域内洪泛的。此后在洪泛外部链路状态阶段，外部链路状态不会被洪泛至存根区域。也就是说，存根区域用一条默认路径代替了大量的外部链路状态，由此存根区域内路由器的链路状态数据库得以减小。

在Area3 的情况下，只有一个ABR，即区域的出接口只有一个，这时将其配置为存根区域是十分划算的。不仅所有AS外部链路状态都可以代替为默认路径，甚至只要是区域外的链路状态都可以用一条默认路径代替。

from	
to	
	RT10
0.0.0.0/0	2

Area 3 中，由ABR RT10洪泛的默认路径，这会在区域三的路由器中生成一条默认路由。距离值是RT10的出接口COST。

交换链路状态—四种链路状态

	Router-LSA：即由路由器发现并在区域内交换的链路状态
	Network-LSA：即DR中配置的路由器列表，也需要在其所在广播网或NBMA网络洪泛
	Summary-LSA：汇总链路状态，在区域间交换并在区域内洪泛
	AS-external- LSA：AS外部链路状态，在AS内洪泛

计算路由

AS每个路由器都存有一份链路状态数据库，如果没有划分区域，它们应该是相同的；如果划分了区域，则不同区域的链路状态数据库有所差别。

每个路由器独立计算自己的路由，计算路由分为如下两步：

1. 根据链路状态数据库计算出最短路径优先树
2. 根据最短路径优先树得出路由

1. 根据链路状态数据库计算出最短路径优先树

这一步骤是与交换链路状态阶段是相互交叉的。在汇总链路状态阶段，明显各区域需要计算出最短路径优先树，然后才能将链路状态进行精简。

1. 根据最短路径优先树得出路由

略

A young woman with long brown hair and black-rimmed glasses is smiling warmly at the camera. She is wearing a light blue denim jacket over a white top. She is holding a white ceramic coffee cup with both hands, positioned towards the bottom left of the frame. The background is a blurred indoor setting, possibly a cafe or a lounge, with warm lighting and some greenery visible.

RELAX

参 考 资 料

RFC 2328

RFC 2328 (中文版)

H3C网络学院路由交换技术第一卷 (下册) : 第8篇第36章

OSPF百度百科

[H3C OSPF技术介绍](#)

Thanks