IPv6邻居发现及状态迁移机制

H3C交换机基础维护篇

IPv6进阶篇

上一期我们讲述了IPv6地址的分类、表示、组成和配置，那么当交换机正确获取到IPv6地址之后，是如何与IPv6邻居进行通信的呢？这就是今天我们想要介绍的内容：

1. IPv6邻居发现机制
2. IPv6邻居状态机制

IPv6邻居发现机制

邻居发现协议（Neighbor Discovery Protocol）是IPv6协议的一个基本组成部分，使用了下表中的五种类型ICMPv6消息：

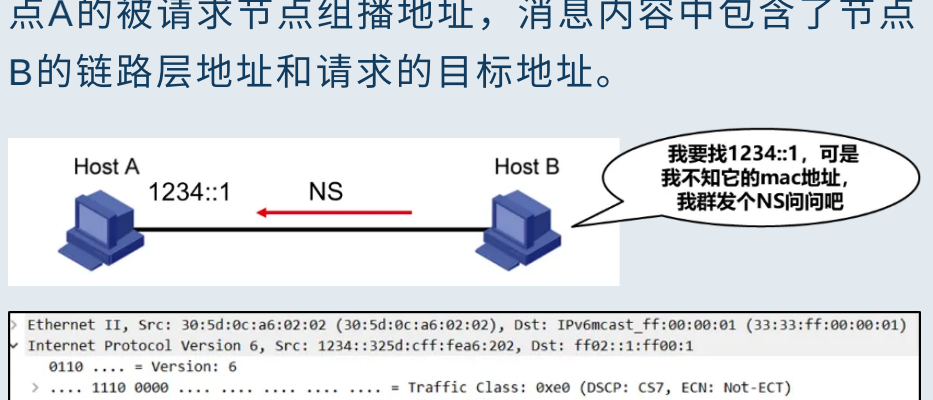
ICMPv6消息	类型号	作用
邻居请求消息NS（Neighbor Solicitation）	135	获取邻居的链路层地址
		验证邻居是否可达
		进行重复地址检测
邻居通告消息NA（Neighbor Advertisement）	136	对NS消息进行响应
		节点在链路层变化时主动发送NA消息，向邻居节点通告本节点的变化信息
路由器请求消息RS（Router Solicitation）	133	节点启动后，通过RS消息向路由器发出请求，请求前缀和其他配置信息，用于节点的自动配置
路由器通告消息RA（Router Advertisement）	134	对RS消息进行响应
		在没有抑制RA消息发布的条件下，路由器会周期性地发布RA消息，其中包括前缀信息选项和一些标志位的信息
重定向消息（Redirect）	137	当满足一定的条件时，缺省网关通过向源主机发送重定向消息，使主机重新选择正确的下一跳地址进行后续报文的发送

使用表格中的五种ICMPv6消息，可以实现在IPv4中的地址解析协议（ARP）、控制报文协议（ICMP）中的路由器发现部分、重定向协议的所有功能，并具有邻居不可达检测机制，下面我们就来一一展开介绍下。

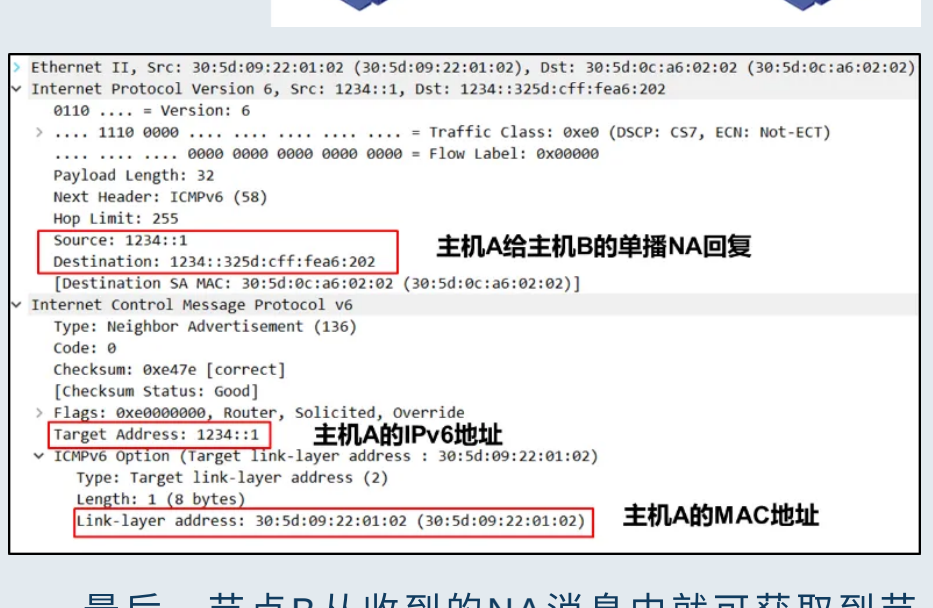
1. 地址解析过程

通过邻居请求消息NS和邻居通告消息NA，可以获取同一链路上邻居节点的链路层地址（与IPv4的ARP功能相同）。

首先，节点B以组播方式发送NS消息。NS消息的源地址是节点B的接口IPv6地址，目的地址是节点A的被请求节点组播地址，消息内容中包含了节点B的链路层地址和请求的目标地址。



接着，节点A收到NS消息后，判断报文的目标地址是否为自己的IPv6地址。如果是，则节点A可以学习到节点B的链路层地址，并以单播方式返回NA消息，其中包含了自己的链路层地址。

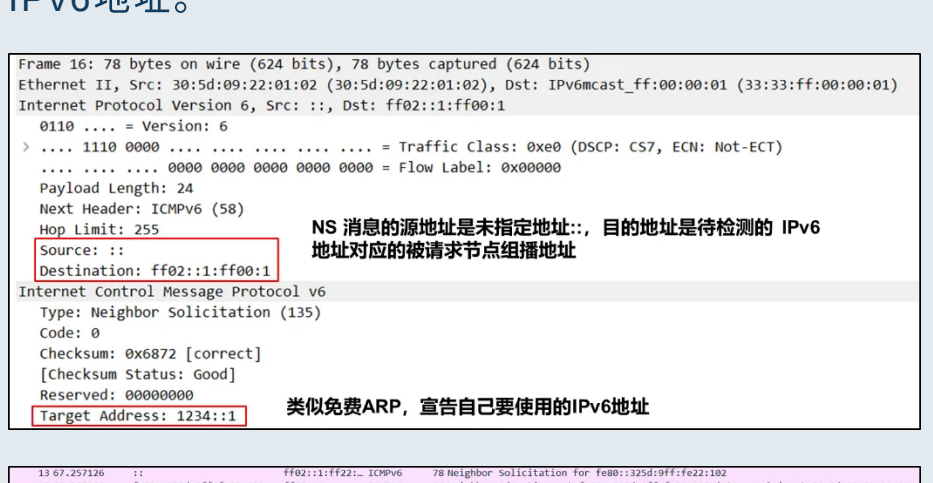


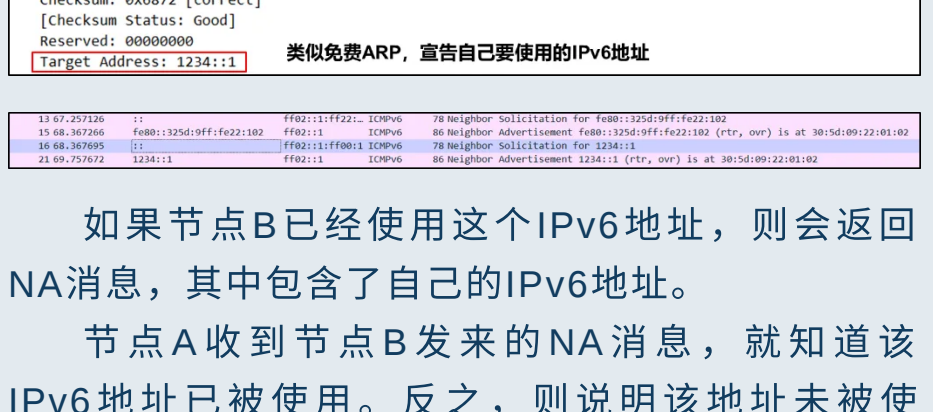
最后，节点B从收到的NA消息中就可获取到节点A的链路层地址。

2. 重复地址检查

当节点获取到一个IPv6地址后，需要使用重复地址检查功能确定该地址是否已被其他节点使用（与IPv4的免费ARP功能相似），避免冲突。通过NS和NA实现重复地址检测的过程为：

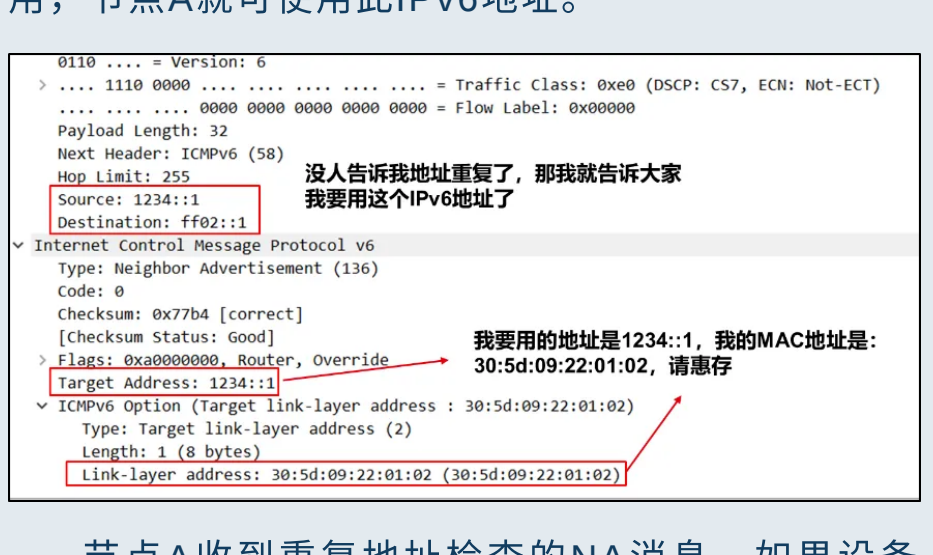
首先，节点A发送NS消息，NS消息的源地址是未指定地址::，目的地址是待检测的IPv6地址对应的被请求节点组播地址，消息内容中包含了待检测的IPv6地址。





如果节点B已经使用这个IPv6地址，则会返回NA消息，其中包含了自己的IPv6地址。

节点A收到节点B发来的NA消息，就知道该IPv6地址已被使用。反之，则说明该地址未被使用，节点A就可使用此IPv6地址。

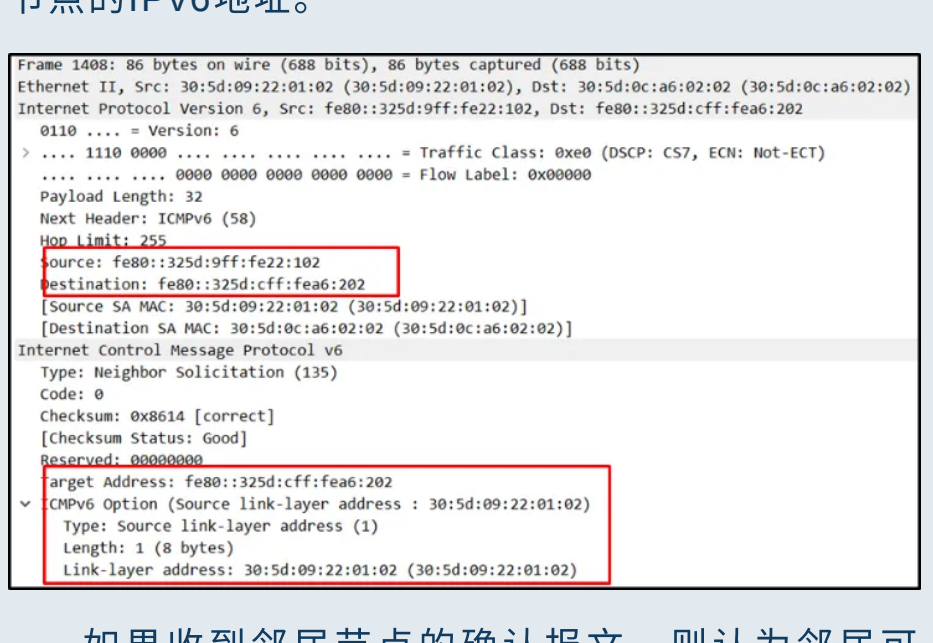


节点A收到重复地址检查的NA消息，如果设备上没有表项，不会进行学习；如果自己已有表项，则进行更新。

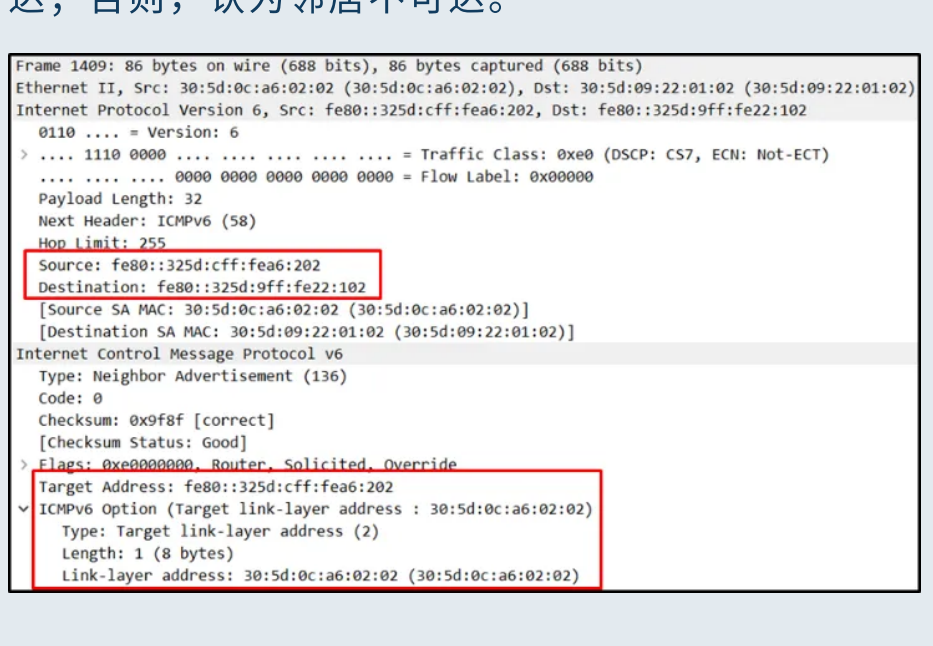
3. 邻居可达性检测

在获取到邻居节点的链路层地址后，通过邻居请求消息NS和邻居通告消息NA可以验证邻居节点是否可达。

首先，节点发送NS消息，其中目的地址是邻居节点的IPv6地址。



如果收到邻居节点的确认报文，则认为邻居可达；否则，认为邻居不可达。

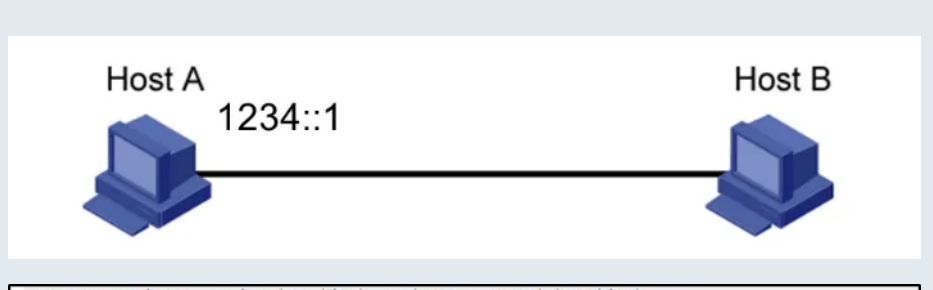


4. 路由器发现/前缀发现及地址无状态自动配置

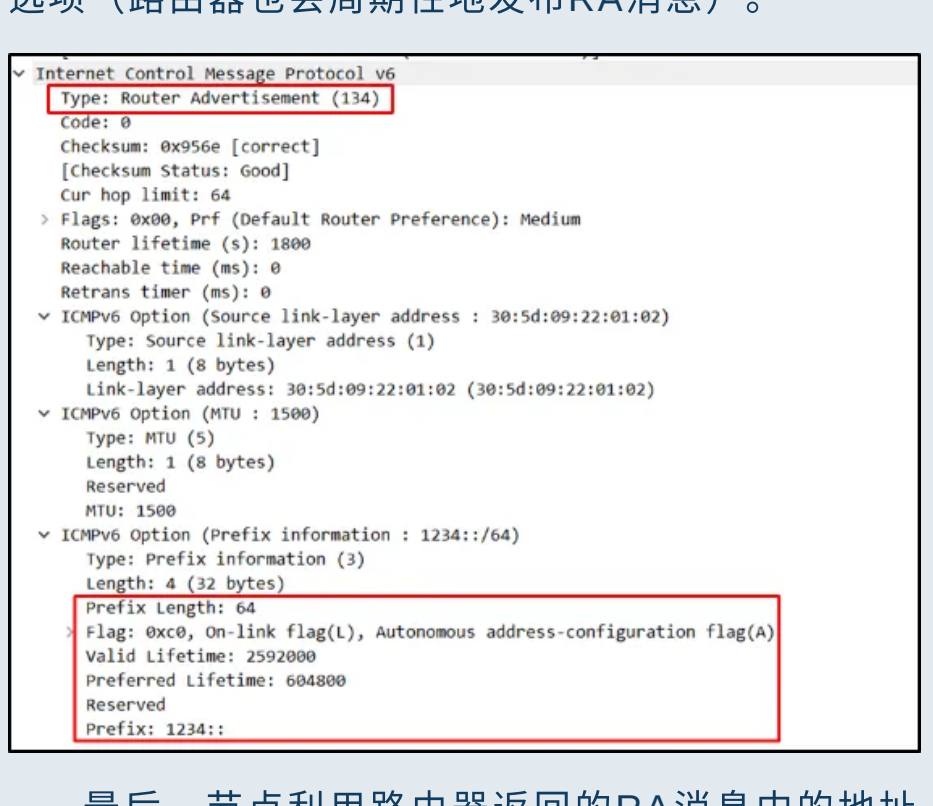
路由器发现/前缀发现是指节点从收到的RA消息中获取邻居路由器及所在网络的前缀，以及其他配置参数。地址无状态自动配置是指节点根据路由器发现/前缀发现所获取的信息，自动配置IPv6地址。

路由器发现/前缀发现通过路由器请求消息RS和路由器通告消息RA来实现，具体过程如下：

首先，节点启动时，通过RS消息向路由器发出请求，请求前缀和其他配置信息，以便用于节点的配置。



接着，路由器返回RA消息，其中包括前缀信息选项（路由器也会周期性地发布RA消息）。



最后，节点利用路由器返回的RA消息中的地址前缀及其他配置参数，自动配置接口的IPv6地址及其他信息。

IPv6邻居状态迁移

RFC4861定义了5种IPv6邻居状态，分别是：Incomplete、Reachable、Stale、Delay、Probe，其中只有Stale状态是稳定状态。

（1）Incomplete（未完成状态）：表示正在解析地址，但邻居链路层地址尚未确定。

（2）Reachable（可达状态）：表示地址解析成功，该邻居可达。

（3）Stale（失效状态）：表示可达时间耗尽，未确定邻居是否可达。

（4）Delay（延迟状态）：表示未确定邻居是否可达。Delay状态不是一个稳定的状态，而是一个延时等待状态。

（5）Probe（探测状态）：节点会向处于Probe状态的邻居持续发送NS报文。

下面以实验测试过程，展现IPv6邻居状态迁移机制。

Empty➡INCMP

没有表项时，如果有流量触发建立新的邻居表项，设备将发送NS报文去获取邻居MAC地址，此时建立的邻居表项是INCMP状态，MAC全零。

```
通过Ping操作触发流量：
<SW>ping ipv6 1994::2
Ping6(56 data bytes) 1994::1 --> 1994::2, press CTRL+C to break

通过debug观察表项变化过程：
*Jan 17 20:33:40:246 2011 S6820-2023 ND/7/ND_ENTRY:
Added INCOMPLETE NB entry: 1994::2 on interface Vlan30

查看设备邻居表项MAC全为0：
<SW>display ipv6 neighbors all
Type: S-Static   D-Dynamic   O-Openflow   R-Rule   IS-Invalid static
IPv6 address          MAC address     VID  Interface      State T  Aging
1994::2               0000-0000-0000  30   --             INCMP D  #
```

INCMP➡Empty

发送NS报文探测对端邻居是否可达后，若尝试3次对端都不响应，则删除表项，探测间隔默认为1s。

```
通过debug观察NS报文发送及邻居表项变化过程：
*Jan 17 20:33:40:246 2011 S6820-2023 ND/7/ND_PACKET:
 Sent NS packet: (第一次发送NS报文)
 Interface: Vlan30          First VLAN ID: 0   Second VLAN ID: 0
 SrcEthMAC: 4077-a9f0-38f3   SrcIP: 1994::1
 DstEthMAC: 0000-0000-0000   DstIP: ff02::1:ff00:2
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:33:41:294 2011 S6820-2023 ND/7/ND_PACKET:
 Sent NS packet: (第二次发送NS报文)
 Interface: Vlan30          First VLAN ID: 0   Second VLAN ID: 0
 SrcEthMAC: 4077-a9f0-38f3   SrcIP: 1994::1
 DstEthMAC: 0000-0000-0000   DstIP: ff02::1:ff00:2
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:33:42:294 2011 S6820-2023 ND/7/ND_PACKET:
 Sent NS packet: (第三次发送NS报文)
 Interface: Vlan30          First VLAN ID: 0   Second VLAN ID: 0
 SrcEthMAC: 4077-a9f0-38f3   SrcIP: 1994::1
 DstEthMAC: 0000-0000-0000   DstIP: ff02::1:ff00:2
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:33:43:294 2011 S6820-2023 ND/7/ND_ENTRY:
 deleted INCOMPLETE NB entry: 1994::2 on interface Vlan30
```

此时设备上查看display ipv6 neighbors all已没有对应表项。

INCMP➡REACH

若目的邻居可达并回复了正确的NA报文，则本端邻居表项状态刷新为REACH，并且将MAC更新。

```
通过debug观察报文收发及邻居表项变化过程：
*Jan 17 20:41:18:586 2011 S6820-2023 ND/7/ND_PACKET: -Slot=2;
 Received NA packet:
 Interface: Vlan30          First VLAN ID: 30   Second VLAN ID: 0
 SrcEthMAC: 80e4-55f1-6920   SrcIP: 1994::2
 DstEthMAC: 4077-a9f0-38f3   DstIP: 1994::1
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:41:18:586 2011 S6820-2023 ND/7/ND_ENTRY: -Slot=2;
 Added REACHABLE NB entry: 1994::2 on interface Vlan30

<SW>display ipv6 neighbors all
Type: S-Static   D-Dynamic   O-Openflow   R-Rule   IS-Invalid static
IPv6 address                MAC address     VID  Interface   State T  Aging
1994::2                     80e4-55f1-6920  30   FGE2/3/25   REACH D  12
FE80::82E4:55FF:FEF1:6920   80e4-55f1-6920  30   FGE2/3/25   REACH D  2
```

REACH➡STALE

REACH状态将维持30s，后自动切换成STALE状态。

```
通过debug观察邻居表项变化过程：
*Jan 17 20:41:48:677 2011 S6820-2023 ND/7/ND_ENTRY: -Slot=2;
 REACHABLE->STALE : 1994::2 on interface Vlan30

<SW>display ipv6 neighbors all
Type: S-Static   D-Dynamic   O-Openflow   R-Rule   IS-Invalid static
IPv6 address                MAC address     VID  Interface   State T  Aging
1994::2                     80e4-55f1-6920  30   FGE2/3/25   STALE D  17
FE80::82E4:55FF:FEF1:6920   80e4-55f1-6920  30   FGE2/3/25   STALE D  7
```

STALE➡DELAY

STALE状态维持的时间又称为老化时间，默认是240min，老化后表项切换成DELAY状态。

```
通过debug观察邻居表项变化过程：
*Jan 17 20:45:24:253 2011 S6820-2023 ND/7/ND_ENTRY: -Slot=2;
 STALE->DELAY : 1994::2 on interface Vlan30

<SW>display ipv6 neighbors all
Type: S-Static   D-Dynamic   O-Openflow   R-Rule   IS-Invalid static
IPv6 address                MAC address     VID  Interface   State T  Aging
1994::2                     80e4-55f1-6920  30   FGE2/3/25   DELAY D  2
FE80::82E4:55FF:FEF1:6920   80e4-55f1-6920  30   FGE2/3/25   STALE D  112
```

DELAY➡PROBE

DELAY状态延时5s，自动切换为PROBE状态。

```
通过debug观察邻居表项变化过程：
*Jan 17 20:45:29:258 2011 S6820-2023 ND/7/ND_ENTRY: -Slot=2;
 DELAY->PROBE : 1994::2 on interface Vlan30

[SW] display ipv6 neighbors all
Type: S-Static   D-Dynamic   O-Openflow   R-Rule   IS-Invalid static
IPv6 address                MAC address     VID  Interface   State T  Aging
1994::2                     80e4-55f1-6920  30   FGE2/3/25   PROBE D  1
FE80::82E4:55FF:FEF1:6920   80e4-55f1-6920  30   FGE2/3/25   STALE D  116
```

PROBE➡Empty

邻居表项在PROBE状态时，发送NS报文探测对端是否可达，若尝试3次对端都不响应，则删除表项。

```
通过debug观察NS报文发送及邻居表项变化过程：
*Jan 17 20:51:12:908 2011 S6820-2023 ND/7/ND_PACKET: -Slot=2;
 Sent NS packet: (第一次发送NS报文)
 Interface: Vlan30          First VLAN ID: 0   Second VLAN ID: 0
 SrcEthMAC: 4077-a9f0-38f3   SrcIP: fe80::4277:a9ff:fef0:38f3
 DstEthMAC: 0000-0000-0000   DstIP: 1994::2
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:51:13:908 2011 S6820-2023 ND/7/ND_PACKET: -Slot=2;
 Sent NS packet: (第二次发送NS报文)
 Interface: Vlan30          First VLAN ID: 0   Second VLAN ID: 0
 SrcEthMAC: 4077-a9f0-38f3   SrcIP: fe80::4277:a9ff:fef0:38f3
 DstEthMAC: 0000-0000-0000   DstIP: 1994::2
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:51:14:908 2011 S6820-2023 ND/7/ND_PACKET: -Slot=2;
 Sent NS packet: (第三次发送NS报文)
 Interface: Vlan30          First VLAN ID: 0   Second VLAN ID: 0
 SrcEthMAC: 4077-a9f0-38f3   SrcIP: fe80::4277:a9ff:fef0:38f3
 DstEthMAC: 0000-0000-0000   DstIP: 1994::2
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:51:15:908 2011 S6820-2023 ND/7/ND_ENTRY: -Slot=2;
 deleted PROBE NB entry: 1994::2 on interface Vlan30
```

PROBE➡REACH

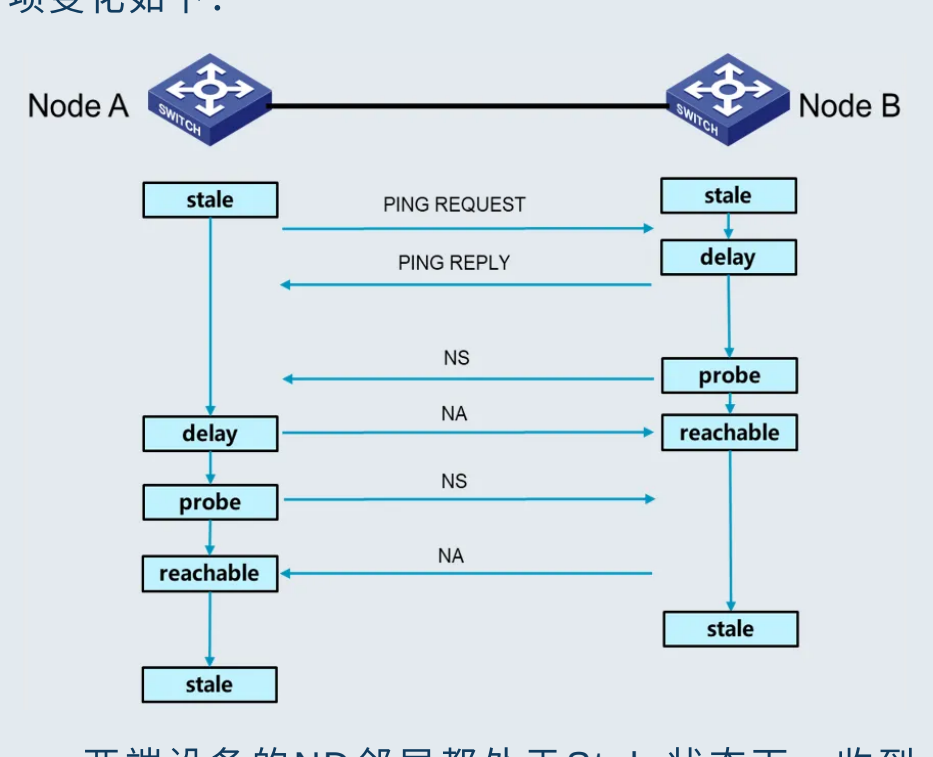
邻居表项在PROBE状态时，发送NS报文探测对端邻居是否可达，若邻居可达并回复NA报文，则本端邻居表项状态刷新为REACH。

```
通过debug观察NA报文接收及邻居表项变化过程：
*Jan 17 20:45:29:259 2011 S6820-2023 ND/7/ND_PACKET: -Slot=2;
 Received NA packet:
 Interface: Vlan30          First VLAN ID: 30   Second VLAN ID: 0
 SrcEthMAC: 80e4-55f1-6920   SrcIP: 1994::2
 DstEthMAC: 4077-a9f0-38f3   DstIP: fe80::4277:a9ff:fef0:38f3
 LinkId: 0xffff   VsiIndex: 0xffffffff

*Jan 17 20:45:29:259 2011 S6820-2023 ND/7/ND_ENTRY: -Slot=2;
 PROBE->REACHABLE : 1994::2 on interface Vlan30

[SW]display ipv6 neighbors all
Type: S-Static   D-Dynamic   O-Openflow   R-Rule   IS-Invalid static
IPv6 address                MAC address     VID  Interface   State T  Aging
1994::2                     80e4-55f1-6920  30   FGE2/3/25   REACH D  14
FE80::82E4:55FF:FEF1:6920   80e4-55f1-6920  30   FGE2/3/25   REACH D  4
```

关于Stale状态

讲完了5种IPv6邻居状态机及其迁移机制，可能有同学要举手提问了，唯一的稳定状态Stale字面意思是陈旧的，那邻居表项为Stale状态时邻居是否可达呢？

通过一个小实验可以发现，Stale状态下的邻居虽然在设备看来不确定是否可达，但如果有流量通过时还可以互访。一次常规的Ping操作两端邻居表项变化如下：



两端设备的ND邻居都处于Stale状态下，收到流量的节点B会马上将节点A的邻居表项迁移为Delay，延迟5S变为Probe状态并发送NS报文。

1. 节点A收到对端的NS报文后回复NA，同时将对端的表项迁移至Delay。

2. 节点B收到NA回复后将对端表项迁移为Reachable。

3. 节点A上对端的表项在Delay延迟5S后迁移至Probe，也发送了NS报文；收到NA回复后表项迁移至Reachable。

4. 节点A、B上的邻居表项在Reachable状态下持续30S再次恢复为Stale。

总结

今天简单讲述了IPv6进阶篇之邻居发现及状态迁移机制，让我们回顾下重点内容：

1. ICMPv6消息包含NS、NA、RS、RA和Redirect五种。

2. 通过NS和NA消息，可以实现IPv6地址解析、重复地址检测和邻居可达性检测功能。

3. 通过RS和RA消息，可以实现路由器发现/前缀发现及地址无状态自动配置功能。

4. IPv6邻居状态包含Incomplete、Reachable、Stale、Delay、Probe五种，其中只有Stale状态是稳定状态。

—— end ——