第3章

IPv6邻居发现

ND(Neighbor Discovery,邻居发现)协议是IPv6的一个关键协议,它综合了IPv4中的ARP、ICMP路由器发现和ICMP重定向等协议,并对它们做了改进。作为IPv6的基础性协议,ND协议还提供了前缀发现、邻居不可达检测、重复地址检测、地址自动配置等功能。

ND协议在RFC4861—Neighbor Discovery in IPv6中定义。

通过本章的学习,应该掌握以下内容。

(1) IPv6邻居发现协议的基本功能。

(2) IPv6邻居发现协议的报文结构。

(3) IPv6地址解析过程。

(4) IPv6邻居状态机。

(5) 无状态地址自动配置过程。

(6) IPv6报文重定向基本原理。

3.1 ND协议概述

3.1.1 功能简介

IPv6的ND协议实现了IPv4中的一些协议功能,如ARP、ICMP路由器发现和ICMP重定向等,并对这些功能进行了改进。同时,作为IPv6的一个基础性协议,ND协议还提供了其他许多非常重要的功能,如前缀发现、邻居不可达检测、重复地址检测、无状态地址自动配置等,如图3-1所示。

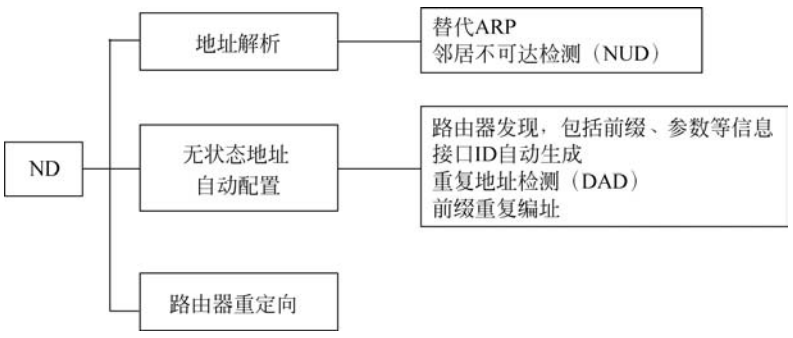

图3-1 ND协议功能组成

图3-1中提到的术语、概念解释如下。

(1) 地址解析:地址解析是一种确定目的节点的链路层地址的方法。ND中的地址解析功能不仅代替了原IPv4中的ARP协议,同时还用邻居不可达检测(NUD)方法来维护邻居节

点之间的可达性状态信息。

(2) 无状态地址自动配置：ND 协议中特有的地址自动配置机制，包括一系列相关功能，如路由器发现、接口 ID 自动生成、重复地址检测(DAD)等。通过无状态自动配置机制，链路上的节点可以自动获得 IPv6 全球单播地址。

① 路由器发现：路由器在与其相连的链路上发布网络参数信息，主机捕获此信息后，可以获得全球单播 IPv6 地址前缀、默认路由、链路参数(链路 MTU)等信息。

② 接口 ID 自动生成：主机根据 EUI-64 规范或其他方式为接口自动生成接口标识符。

③ 重复地址检测(DAD)：根据前缀信息生成 IPv6 地址或手动配置 IPv6 地址后，为保证地址的唯一性，在这个地址可以使用之前，主机需要检验此 IPv6 地址是否已经被链路上其他节点所使用。

④ 前缀重新编址：当网络前缀变化时，路由器在与其相连的链路上发布新的网络参数信息，主机捕获这些新信息，重新配置前缀、链路 MTU 等地址相关信息。

(3) 路由器重定向：当在本地链路上存在一个到达目的网络的更好的路由器时，路由器需要通告节点来进行相应配置改变。

3.1.2 ND 协议报文

在 IPv4 的地址解析中，ARP 报文直接封装在以太帧中，其以太网协议类型为 0x0806，代表 ARP 报文。ARP 被看作工作在 2.5 层的协议。而 ND 协议本身基于 ICMPv6 实现，因此 ND 协议是在第三层上实现的。ND 协议报文的以太网协议类型为 0x86DD，即 IPv6 报文。IPv6 的下一个报头协议类型为 58，表示是 ICMPv6 报文。上述两者的对比如图 3-2 所示。

图 3-2 ARP 与 ND 协议报文封装

ND 协议定义了 5 种 ICMPv6 报文类型，包括 RS、RA、NS、NA 和 Redirect 报文，如表 3-1 所示。

表 3-1 ICMPv6 报文类型

ICMPv6 类型	消 息 名 称
Type=133	RS(Router Solicitation，路由器请求)
Type=134	RA(Router Advertisement，路由器公告)
Type=135	NS(Neighbor Solicitation，邻居请求)
Type=136	NA(Neighbor Advertisement，邻居公告)
Type=137	Redirect(重定向报文)

NS/NA 报文主要用于地址解析，RS/RA 报文主要用于无状态地址自动配置，Redirect 报文用于路由器重定向。

3.1.3 重要概念

节点根据IPv6地址是否存在于指定链路的某个接口上，把这些地址划分为On-link或Off-link。同时，邻居之间对用于通信的IPv6地址，还维护一个可达性状态信息。在维护邻居可达性(Reachability)状态信息的交互报文中，使用了目标(Target)地址的概念，来指明查询的对象。

1. On-link

On-link表示这个IPv6地址存在于指定链路的某个接口上。遇到以下四种情况时，节点可以认为这样的IPv6地址是On-link的。

(1) 这个地址中的前缀属于指定链路上的某个前缀。

(2) 这个地址被邻居路由器指定，作为重定向报文中的目标地址。

(3) 节点收到了从这个地址发出的NA报文(这个地址是NS报文中的目标地址)。

(4) 节点从这个地址收到了ND协议报文。

2. Off-link

相对于On-link，即表示这个地址不存在于指定链路的某个接口上。

3. 可达性(Reachability)

表明邻居节点的IP层是否可达。

4. 目标地址

在地址解析中，表示哪个地址寻求解析信息；在重定向中，表示报文被重定向到新的第一跳地址。此外，在DAD和NUD中也用到了目标地址。

3.1.4 主机数据结构

主机数据结构(Conceptual Data Structures)是在RFC4861中定义的。为使相邻节点间的交互更为方便，IETF建议节点维护以下表项。

1. 邻居缓存表(Neighbor Cache)

邻居缓存表是由近期发送过数据流的邻居信息组成的表项。邻居缓存表内记录了每个邻居的IP地址、相应的链路层地址、可达性状态等信息，类似于IPv4中的ARP表项。

邻居缓存表可以根据RS、NS和NA报文动态更新，同时也可以通过命令进行静态配置。

2. 前缀列表(Prefix List)

前缀列表是主机根据接收到的RA报文中的前缀信息建立的表项，记录了与前缀相关的参数信息，如前缀地址、前缀长度、有效时间、优先时间等。

3. 默认路由器表(Default Router List)

默认路由器表包含了本地链路上默认路由器的信息。表项的内容可从RA报文中提取，或者通过手动配置。

4. 目的缓存表(Destination Cache)

由已发送报文的目的地址所组成的表项，是主机发送报文时查找的第一张表。在数据转发初始阶段，节点会查询邻居缓存表、前缀列表和默认路由器表来建立该表，同时还根据重定向报文进行更新。目的缓存表记录了目的IP地址、对应下一跳地址、目的路径MTU等信息。

3.2 IPv6 地址解析

地址解析在报文转发过程中具有至关重要的作用。当一个节点需要得到同一链路上另外一个节点的链路层地址时，需要进行地址解析。IPv4 中使用 ARP 协议实现了这个功能，IPv6 使用 ND 协议实现了这个功能，但功能有所增强。

IPv6 的地址解析过程包括两部分，一部分解析了目的 IP 地址所对应的链路层地址；另一部分是邻居可达性状态的维护过程，即邻居不可达检测。

3.2.1 地址解析

1. IPv6 地址解析的优点

IPv6 地址解析技术在基本思想上仍然与 IPv4 的 ARP 类似，但是 IPv6 地址解析相比 IPv4 的 ARP 最大的一个不同是，IPv6 地址解析工作在 OSI 模型的网络层，与链路层协议无关。这是一个很显著的优点，它的益处如下。

(1) 加强了地址解析协议与底层链路的独立性。对每一种链路层协议都使用相同的地址解析协议，无须再为每一种链路层协议定义一个新的地址解析协议。

(2) 增强了安全性。ARP 攻击、ARP 欺骗是 IPv4 中严重的安全问题。在第三层实现地址解析，可以利用三层标准的安全认证机制来防止这种 ARP 攻击和 ARP 欺骗。

(3) 减小了报文传播范围。在 IPv4 中，ARP 广播必须泛滥到二层网络中每台主机。IPv6 的地址解析利用三层组播寻址限制了报文的传播范围，仅将地址解析请求发送到待解析地址所属的被请求节点(Solicited-node)组播组，减小了报文传播范围，节省了网络带宽。

2. IPv6 地址解析过程

IPv6 中，ND 协议通过在节点间交互 NS 和 NA 报文完成 IPv6 地址到链路层地址的解析，解析后用得到的链路层地址和 IPv6 地址等信息来建立相应的邻居缓存表项，如图 3-3 所示，NodeA 的链路层地址为 00E0-FC00-0001，全局地址为 1::1:A；NodeB 的链路层地址为

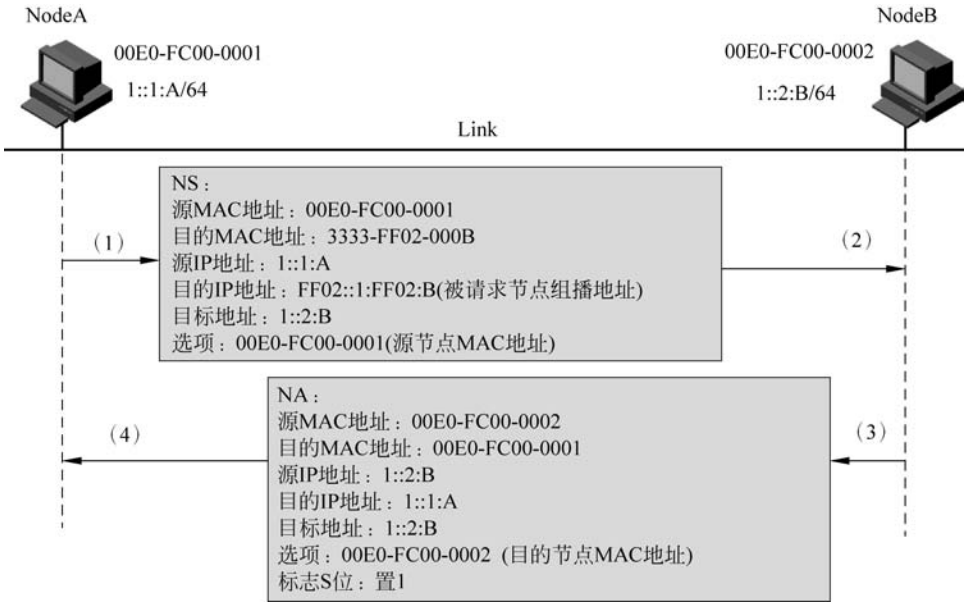

图 3-3 地址解析

00E0-FC00-0002，全局地址为 1::2:B。当 NodeA 要发送数据报文到 NodeB 时，如果不知道 NodeB 的链路层地址，则需要 ND 协议完成以下地址解析过程。

(1) NodeA 发送一个 NS 报文到链路上，目的 IPv6 地址为 NodeB 对应的被请求节点组播地址(FF02::1:FF02:B)，选项字段中携带了 NodeA 的链路层地址 00E0-FC00-0001。

(2) NodeB 接收到该 NS 报文后，由于报文的目的地址 FF02::1:FF02:B 是 NodeB 的被请求节点组播地址，所以 NodeB 会处理该报文；同时，根据 NS 报文中的源地址和源链路层地址选项更新自己的邻居缓存表项。

(3) NodeB 发送一个 NA 报文来应答 NS，同时在消息的目标链路层地址选项中携带自己的链路层地址 00E0-FC00-0002。

(4) NodeA 接收到 NA 报文后，根据报文中携带的 NodeB 链路层地址，创建一个到目标节点 NodeB 的邻居缓存表项。

通过交互，NodeA 和 NodeB 就获得了对方的链路层地址，建立起到达对方的邻居缓存表项，从而可以相互通信。

当一个节点的链路层地址发生改变时，以所有节点组播地址 FF02::1 为目的地址发送 NA 报文，通知链路上的其他节点更新邻居缓存表项。

3.2.2 NUD(邻居不可达检测)

NUD(Neighbor Unreachability Detection，邻居不可达检测)是节点确定邻居可达性的过程。邻居不可达检测机制通过邻居可达性状态机描述邻居的可达性。邻居可达性状态机之间满足一定的条件时，可相互迁移。

1. 邻居可达性状态机

邻居可达性状态机保存在邻居缓存表中，共有以下五种。

(1) Incomplete(未完成)状态：表示正在解析地址，邻居的链路层地址尚未确定。当节点第一次发送 NS 报文到邻节点时，会同时在邻居缓存表中创建一个到此邻节点的新表项，此时表项状态就是 Incomplete。

(2) Reachable(可达)状态：表示地址解析成功，该邻居可达。节点可以与处于 Reachable 状态的邻节点互相通信。不过 Reachable 状态伴随有一个 Reachable_Time 定时器，它并不是一个稳定的状态。在 Reachable_Time 定时器超时后，会转化到 Stale(失效)状态。

(3) Stale(失效)状态：表示未确定邻居是否可达。Stale 状态是一个稳定的状态。

(4) Delay(延迟)状态：表示未确定邻居是否可达。Delay 状态也不是一个稳定的状态，而是一个延时等待状态。Delay 状态下，节点需要收到“可达性证实信息”后，才能进入 Reachable 状态。

(5) Probe(探测)状态：同样表示未确定邻居是否可达。节点会向处于 Probe 状态的邻居持续发送 NS 报文，直到接收到“可达性证实信息”后，才能进入 Reachable 状态。

在 Stale 和 Probe 状态时，节点需要收到“可达性证实信息”后，才能进入 Reachable 状态。“可达性证实信息”的来源有以下两种。

(1) 来自上层连接协议的暗示：如果邻节点之间有 TCP 连接，且收到了对端节点发出的确认消息，则表明邻节点之间可达。

(2) 来自不可达探测回应：节点发送 NS 报文后，收到邻节点回应的 S 置位的 NA 报文，则会认为邻节点可达。S 置位的 NA 报文表明这个 NA 报文是专门响应 NS 报文的。

图 3-4 表示了邻居缓存表项中状态机的变化。为描述方便，在五种状态的基础上再增加 Empty 状态，表示节点上没有相关邻节点的邻居缓存表项。

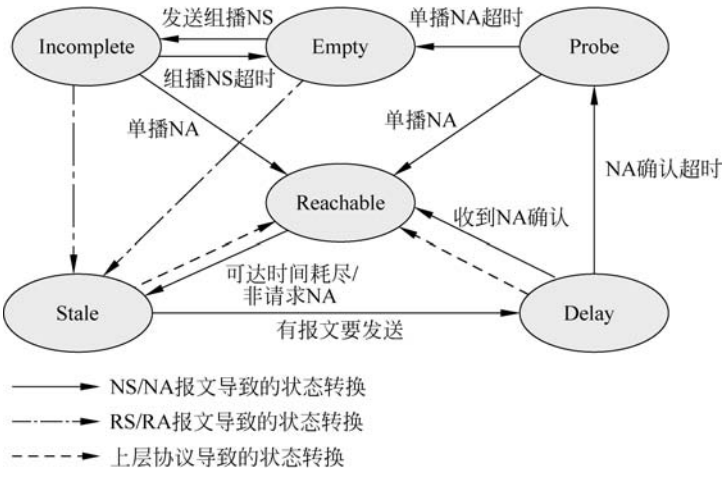

图 3-4 邻居状态机

图 3-4 中实线箭头表示由 NS/NA 报文导致的状态转换，各状态间的相互转换如下。

(1) 在 Empty 状态时，如果有报文要发送给邻节点，则在本地邻居缓存表建立关于该邻节点的表项，并将该表项置于 Incomplete 状态，同时向邻节点以组播方式发送 NS 报文。

(2) 节点收到邻居回应的单播 NA 回应后，将处于 Incomplete 状态的邻居缓存表项转化为 Reachable 状态。如果地址解析失败(发出的组播 NS 超时)，则删除该表项。

(3) 处在 Reachable 状态的表项，如果在 Reachable_Time 时间内没有收到关于该邻居的“可达性证实信息”，则进入 Stale 状态。此外，如果该节点收到邻节点发出的非 S 置位 NA 报文，并且链路层地址有变化，相关表项会进入 Stale 状态。还有一种情况，当节点在 Empty 状态时，收到某邻节点的初次 NS 报文时，会根据报文中的源链路层地址建立该邻节点的缓存表项，并将该表项置于 Stale 状态。

(4) 处在 Stale 状态的表项，当有报文发往该邻居时，这个报文会利用缓存的链路层地址进行封装，使该表项进入 Delay 状态，并等待收到“可达性证实信息”。

(5) 进入 Delay 状态后，如果在 Delay_First_Probe_Time 时间内还未能收到关于该邻居的“可达性证实信息”，则该表项进入 Probe 状态。

(6) 在 Probe 状态，节点会周期性地用 NS 报文来探测邻居的可达性，探测最大时间间隔为 Retrans_Timer，在最多尝试 Max_Unicast_Solicit 次后，如果仍未收到邻居回应的 NA 报文，则认为该邻居已不可达，该表项将被删除。

图 3-4 中虚线箭头表示由上层协议导致的状态转换。只要上层协议报文交互仍在进行中，则相关表项就会始终保持 Reachable 状态。同时，每当上层协议表示要开始传输数据时，表项中的 Reachable_Time 就会被刷新，并转到 Reachable 状态。

图 3-4 中点虚线箭头表示由 RS/RA 报文导致的状态转换。当在 Emtpy 或者 Incomplete 状态时，节点只要收到 RS 或者 RA 报文，就会转到 Stale 状态。

需要说明的是，在协议实现中，任何时刻邻居缓存表项都可以从其他状态进入 Emtpy 状态。

2. NUD 检测过程

图 3-5 显示了 NUD 检测过程，该过程与图 3-3 所描述的地址解析过程类似。

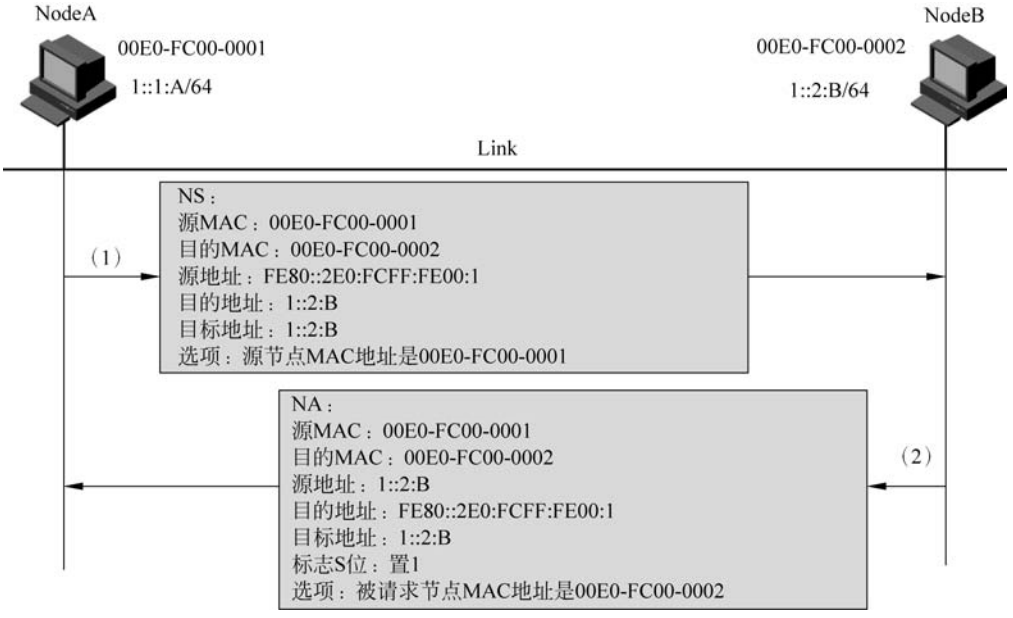

图 3-5 NUD 检测过程

在 NodeA 上，有关 NodeB 的表项在 Reachable 状态经过 Reachable_Time(默认为 30 秒)后，变为 Stale 状态。此时，当 NodeA 有报文要发送给 NodeB 时，且没有上层协议能够提供到 NodeB 的“可达性证实信息”时，NodeA 需要重新验证到 NodeB 的可达性。

NUD 过程与地址解析过程的主要不同之处在于以下两点。

(1) NUD 的 NS 报文的目的 MAC 是目的节点的 MAC 地址；目的 IPv6 地址为 NodeB 的单播地址，而不是被请求节点的组播地址。

(2) NA 报文中的 S 标记必须置位，表示是可达性确认报文，即这个 NA 报文是专门响应 NS 报文的。

需要注意的是，邻居的可达性仅代表了同一链路上相邻节点的可达性，并不能代表网络中端到端的可达性。如果源到目标之间的路径跨越了路由器等第三层设备，NUD 则仅仅验证了到目标路径上第一跳的可达性。

此外，邻居的可达性是单向的。在图 3-5 所示的不可达性检测中，一个请求和应答的过程仅仅使 NodeA(请求发送者)得到了 NodeB(被请求者)的可达性信息，NodeB 并没有获得 NodeA 的可达性信息。此时如果要达到“双向”可达，还需 NodeB 发送 NS 探测报文，NodeA 给 NodeB 回应 S 标志置位的 NA 报文。

3.2.3 地址解析交互报文

1. NS 报文

NS 报文是 ICMPv6 中类型为 135 的报文，如图 3-6 所示。

其中部分字段含义如下。

(1) Target Address：待解析的 IPv6 地址，16 字节长。Target Address 不能是组播地址，

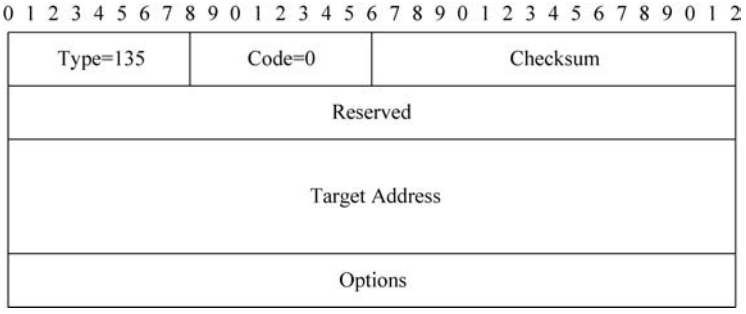

图 3-6 NS 报文

可以是链路本地地址、站点本地地址和全球单播地址。

(2) Options：地址解析中只使用了链路层地址选项(Link-layer Address Option)，是发送NS 报文的节点的链路层地址。链路层地址选项的格式如图 3-7 所示。

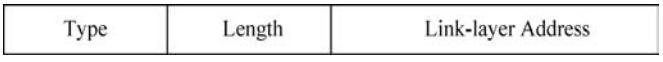

图 3-7 链路层地址选项的格式

其中部分字段含义如下。

(1) Type：选项类型，在链路层地址选项中包括如下两种。

① Type 值为 1，表明链路层地址为 Source Link-layer Address(源链路层地址)，在 NS、RS、Redirect 报文中使用。

② Type 值为 2，表明链路层地址为 Target Link-layer Address(目标链路层地址)，在NA、Redirect 报文中使用。

(2) Length：选项长度，以 8 字节为单位。

(3) Link-layer Address：链路层地址。长度可变，对于以太网为 6 字节。

2. NA 报文

NA 报文是 ICMPv6 中类型为 136 的报文，如图 3-8 所示。

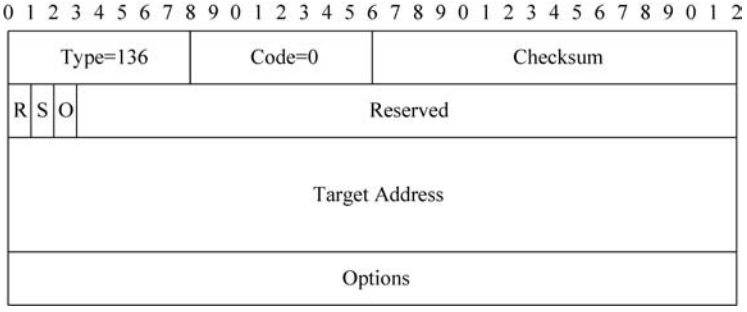

图 3-8 NA 报文

其中部分字段含义如下。

(1) R：路由器标记(Router Flag)位，表示 NA 报文发送者的角色。置为“1”表示发送者是路由器，置为“0”表示发送者为主机。

(2) S：请求标记(Solicited Flag)位。置为“1”表示该 NA 报文是对 NS 报文的响应。

(3) O：覆盖标记(Override Flag)位。置为“1”表示节点可以用 NA 报文中携带的目标链路层地址选项中的链路层地址覆盖原有的邻居缓存表项。置为“0”表示只有在链路层地址未知时，才能用目标链路层地址选项来更新邻居缓存表项。

(4) Target Address：待地址重复检测或地址解析的 IPv6 地址。如果 NA 报文是响应 NS 报文的，则该字段直接复制 NS 报文中的 Target Address。

(5) Options：只能是 Type 值为 2 的 Target Link-layer Address，是被解析节点的链路层地址。

3.3 无状态地址自动配置

IPv6 同时定义了无状态与有状态地址自动配置机制。有状态地址自动配置使用 DHCPv6 协议来给主机动态分配 IPv6 地址，无状态地址自动配置通过 ND 协议来实现。在无状态地址自动配置中，主机通过接收链路上的路由器发出的 RA 消息，结合接口的标识符而生成一个全球单播地址。

无状态地址自动配置的优点如下。

(1) 真正的即插即用。节点连接到没有 DHCP 服务器的网络时，无须手动配置地址等参数便可访问网络。

(2) 网络迁移方便。当一个站点的网络前缀发生变化，主机能够方便地进行重新编址而不影响网络连接。

(3) 地址配置方式选择灵活。系统管理员可根据情况决定使用何种配置方式——有状态、无状态还是两者兼有。

无状态自动配置涉及三种机制：路由器发现、DAD 检测和前缀重新编址。路由器发现可以使节点获得链路上可用的前缀及路由器信息；DAD 检测保证了配置的每个 IPv6 地址在链路上的唯一性；前缀重新编址则是在前面两个机制的基础上，重新通告前缀，完成网络前缀的切换。

3.3.1 路由器发现

路由器发现是指主机怎样定位本地链路上的路由器和确定其配置信息的过程，主要包含以下三方面的内容。

(1) 路由器发现(Router Discovery)：主机发现邻居路由器以及选择哪一个路由器作为默认网关的过程。

(2) 前缀发现(Prefix Discovery)：主机发现本地链路上的一组 IPv6 前缀，生成前缀列表。该列表用于主机的地址自动配置和 On-link 判断。

(3) 参数发现(Parameter Discovery)：主机发现相关操作参数的过程，如链路最大传输单元(MTU)、报文的默认跳数限制(Hop Limit)、地址配置方式等信息。

在路由器通告报文 RA 中承载着路由器的相关信息，ND 协议通过 RS 和 RA 的报文交互完成路由器发现、前缀发现和参数发现三大功能。协议交互主要有两种情况：主机请求触发路由器通告和路由器周期性发送路由器通告。

1. 主机请求触发路由器通告

当主机启动时，主机会向本地链路范围内所有的路由器发送 RS 报文，触发链路上的路由器响应 RA 报文。主机接收到路由器发出的 RA 报文后，自动配置默认路由器，建立默认路由

器列表、前缀列表和设置其他的配置参数。

图 3-9 为 RS 报文触发 RA 报文的过程。图中 NodeA 的链路层地址为 0014-22D4-91B7，链路本地地址为 FE80::214:22FF:FED4:91B7；路由器的链路层地址为 000F-E248-406A，链路本地地址为 FE80::20F:E2FF:FE48:406A。NodeA 以自己的链路本地地址作为源地址，发送一个 RS 报文到所有路由器的组播地址 FF02::2；路由器 RT 收到该报文后，用它的链路本地地址作为源地址，发送 RA 报文到所有节点的组播地址 FF02::1，NodeA 从而获得了路由器上的相关配置信息。

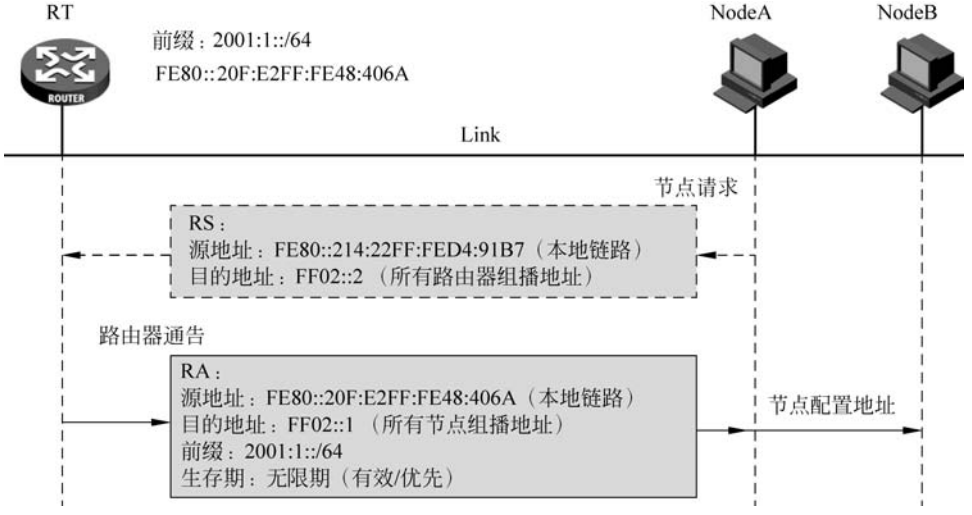

图 3-9 路由器通告过程

注意：为了避免链路上的 RS 报文泛滥，启动时每个节点最多只能发送 3 个 RS 报文。

2. 路由器周期性发送路由器通告

路由器周期性地发送 RA 报文，使主机节点发现本地链路上的路由器及其配置信息，主机节点根据这些内容来维护默认路由器列表、前缀列表和配置其他参数。

图 3-9 中，路由器 RT 用它的本地链路地址 FE80::20F:E2FF:FE48:406A 作为源地址，所有节点的组播地址 FF02::1 作为目的地址，周期性（默认值为 200 秒）地发送 RA 报文，通告自己的前缀（2001:1::/64）等配置信息。然后，监听到该消息的 NodeA 和 NodeB 可以据此配置自己的 IPv6 全球单播地址或者站点本地地址。

3.3.2 重复地址检测

DAD（Duplicate Address Detection，重复地址检测）是节点确定即将使用的地址是否在链路上唯一的过程。所有的 IPv6 单播地址，包括自动配置或手动配置的单播地址，在节点使用之前必须要通过重复地址检测。

DAD 机制通过 NS 和 NA 报文实现，如图 3-10 所示，NodeA 发送的 NS 报文，其源地址为未指定地址，目的地址为接口配置的 IPv6 地址对应的被请求节点组播地址，NA 报文的目标地址字段为待检测的这个 IPv6 地址（图中为 2001:1::1:A/64）。在 NS 报文发送到链路上（默认发送一次 NS 报文）后，如果在规定时间内没有收到应答的 NA 报文，则认为这个单播地址在链路上是唯一的，可以分配给接口；反之，如果收到应答的 NA 报文，则表明这个地址已经被其他节点所使用，不能配置到接口。节点所回应的 NA 报文的源地址为该节点的发送报