技术盛宴丨IPv6系列基础篇（下）——邻居发现协议NDP

原创 杨万里 锐捷网络 2019-01-07

通过上一期文章（IPv6基础篇（上）——地址与报文格式），相信大家对于IPv6的背景、地址和报文格式有了一定了解,接下来大家可能对于终端访问IPv6网络资源的过程原理更感兴趣。那么一个终端如果要访问IPv6的资源，关键的步骤是什么呢？当然是它需要一个IPv6的地址。那么这个地址又从何而来？是不是只能像IPv4一样手动配置或者通过DHCP服务器下发？其实不然，IPv6有更加简洁的地址分配方式，可以通过邻居发现协议实现IPv6地址的自动分配。并且IPv6邻居发现协议远不止这一项功能，这一期将对IPv6邻居发现协议做展开讲解。

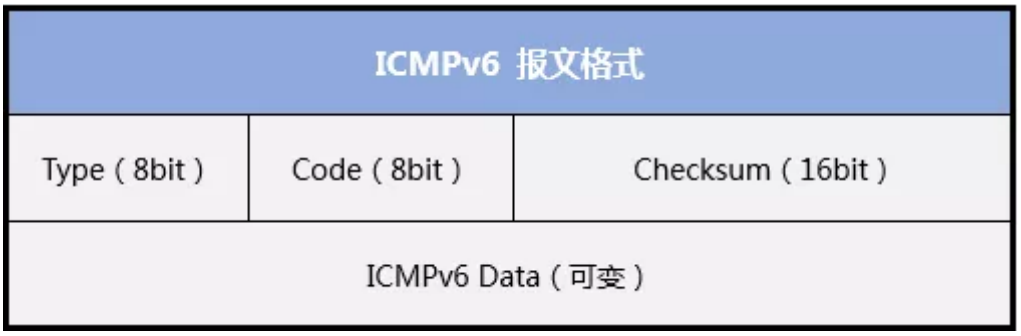
NDP协议概述

NDP（Neighbor Discovery Protocol，邻居发现协议）是IPv6协议体系中一个重要的基础协议。通过使用ICMPv6报文实现以下丰富的功能：

- 无状态自动配置（简化版的DHCP）：路由器发现、前缀发现、参数发现；
- 重复地址检测（DAD），相当于IPv4的免费ARP；
- 地址解析，相当于IPv4的ARP；
- 邻居不可达检测（NUD）；
- 路由器重定向。

为NDP定义的ICMPv6消息

ICMPv6（Internet Control Message Protocol Version 6，互联网控制报文协议版本6）是IPv6的基础协议之一。ICMPv6的协议类型号（IPv6报文中的Next Header字段的值）为58。ICMPv6的报文格式图1所示：



▲图1:ICMPv6报文格式

报文中字段解释如下：

- **Type**：表明消息的类型，0至127表示差错报文类型，128至255表示消息报文类型；
- **Code**：表示此消息类型细分的类型；
- **Checksum**：表示ICMPv6报文的校验和，校验的部分包括了ICMPv6数据和IPv6的报头部分（IPv6报头不含校验）；
- **Data**：ICMPv6数据。

ICMPv6消息类型中有5种是为了支持邻居发现协议而定义的，功能如图2描述：

ICMPv6 Type	消息名称	报文功能
133	路由器请求（RS）	主机发送RS要求路由器产生RA，RA信息中包含MTU以及前缀信息
134	路由器通告（RA）	
135	邻居请求（NS）	用来判断邻居的链路层地址，以及重复地址检测
136	邻居通告（NA）	
137	重定向消息	与IPv4重定向同理

▲图2: ICMPv6五种消息类型

无状态自动配置

IPv6地址配置方式

IPv6地址有128位，即使有简化书写的方式，为主机配置IPv6地址也是一件工作量不小的活儿。IPv6地址除了手工配置外，还能够自动配置，自动配置有两种方式：

1 有状态自动配置

主机通过配置协议（如DHCPv6）获取IPv6地址以及其他信息（如DNS）。状态化自动配置相比于手工配置工作效率要高得多，而相比于无状态自动配置来说更加可控，能够更加清晰地了解到主机及地址分配的相关信息。短板是需要额外部署应用服务器，如DHCPv6 Server。

2 无状态自动配置

相比于前者，无状态地址自动配置则显得更加便捷，IPv6终端使用无状态自动配置能够做到即插即用，无需部署额外的应用服务器、无需使用DHCPv6。在IPv6路由器与IPv6主机之间，利用ICMPv6协议中的路由器请求消息RS（Router Solicitation）和路由器通告RA（Router Advertisement）消息来完成无状态自动配置过程。主机通过RS消息发现链路上的IPv6路由器，而IPv6路由器通过RA消息向主机通告IPv6地址前缀信息，主机在收到IPv6前缀信息后，与自己的网卡接口ID一起构成128位的IPv6全局单播地址。

路由器通告消息

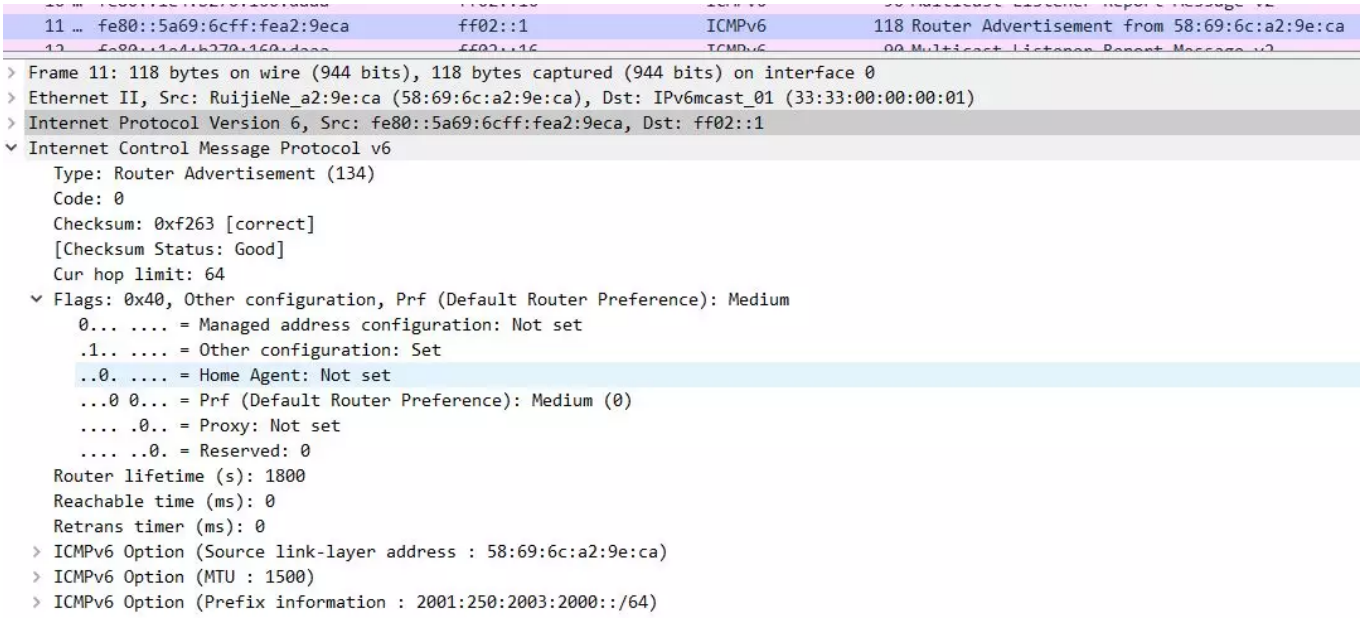
1 RA报文

每台路由器以组播方式定时发送RA报文，用于在二层网络中通告自己的存在。RA报文中会带有网络前缀信息，及另外的一些标志位信息。RA报文的Type字段值为134。

2 RS报文

主机接入网络后希望尽快获取网络前缀进行通信，那么此时主机可以立刻发送RS报文，网络上的路由器将回应RA报文。RS报文的Tpye字段值为133。

RA报文详解如图3所示：



▲图3:RA报文详解

RA报文中重要字段的解释：

- **Managed Address Configuration（M比特）**：默认为0。该标记指示主机该使用何种自动配置方式来获取IPv6单播地址。当M比特被设置为1时，收到该RA消息的主机将使用有状态配置协议（DHCPv6）来获取IPv6地址。
- **Other Configuration（O比特）**：默认为0。该标记指示主机使用何种方式来配置除了IPv6地址外的其他配置信息（如DNS）。当O比特被设置为1，则收到该RA消息的主机将使用配置协议（DHCPv6）来获取除了IPv6地址以外的其他配置信息。

通过M和O比特位的组合，我们可以更清楚地看到终端获取地址和其他配置信息的方式。下面是关于M及O比特的组合：

1. **M=0，O=0**

   应用于没有DHCPv6服务器的环境。主机使用RA消息中的前缀构造IPv6单播地址，同时使用其他方法（非DHCPv6），例如手工配置的方法设置其他配置信息（如DNS）。

2. **M=1，O=1**

   主机使用DHCPv6来配置IPv6单播地址以及其他配置信息。这种应用也称为DHCPv6 Stateful。

3. **M=0，O=1**

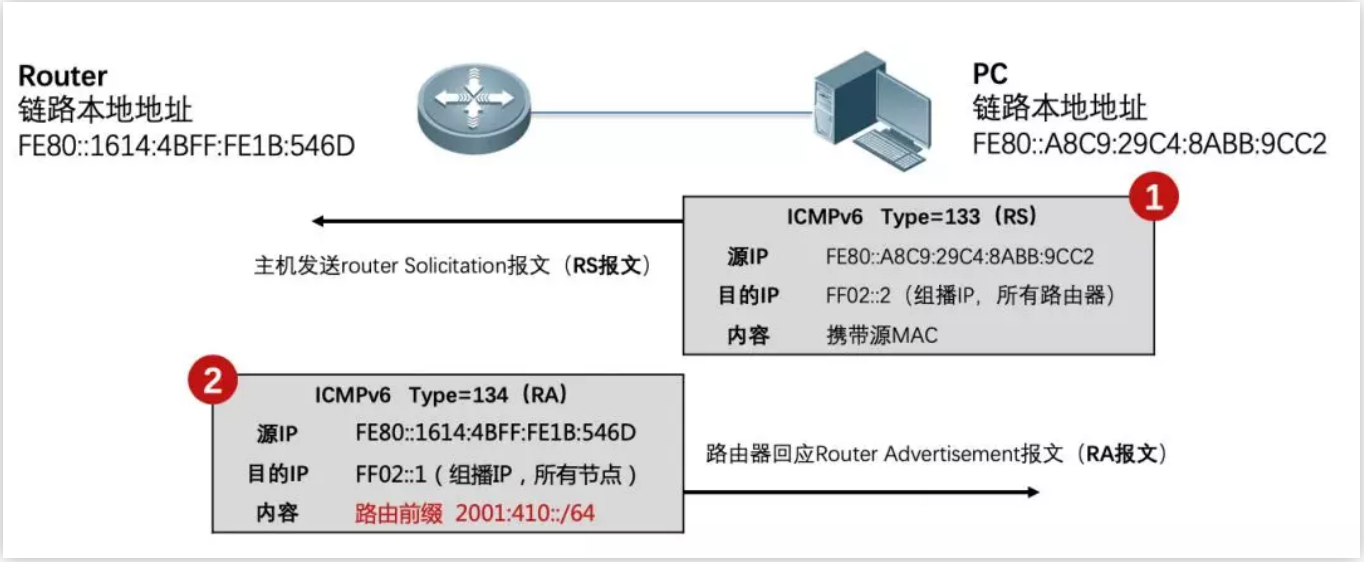
   主机使用RA消息获得的IPv6前缀构造IPv6地址，同时使用DHCPv6来获取除了地址之外的其他配置信息。这种应用也被称为DHCPv6 Stateless。

4 **M=1，O=0**

主机仅仅使用DHCPv6来获取IPv6地址，至于其他配置信息则并不通过DHCPv6获得。

无状态自动配置过程

IPv6主机无状态自动配置的过程如图4所示：



▲图4: IPv6主机无状态自动配置的过程

1. 根据接口标识产生链路本地地址。
2. 发出邻居请求，进行重复地址检测。
3. 如果检测到此链路本地地址已在链路中使用，即地址冲突，则停止自动配置，需要手工配置。
4. 如不冲突，链路本地地址生效，节点具备本地链路通信能力。
5. 主机发送RS报文（或接收到路由器定期发送的RA报文）。
6. 根据RA报文中的前缀信息和通过EUI-64规范生成的接口标识获得IPv6全局单播地址。

重复地址检测DAD

基本概念

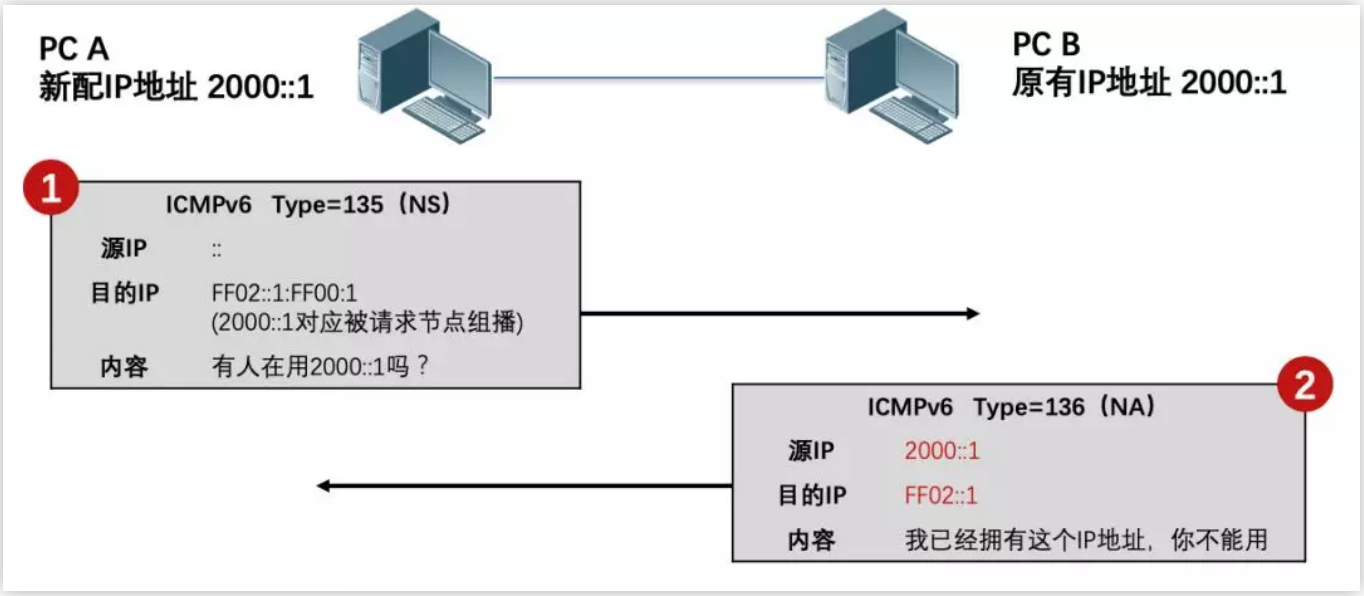
DAD（Duplicate Address Detect，重复地址检测）是在接口使用某个IPv6单播地址之前进行的，主要是为了探测是否有其它的节点使用了该地址。所有的IPv6单播地址，包括自动配置和手动配置的单播地址，在节点使用之前都要通过重复地址检测。

一个IPv6单播地址在分配给一个接口之后且通过重复地址检测之前，称为试验地址（Tentative Address）。此时该接口不能使用这个试验地址进行单播通信，但是仍然会加入两个组播组：All-Nodes组播组和试验地址所对应的Solicited-Node组播组。

IPv6重复地址检测技术和IPv4中的免费ARP类似：节点向试验地址所对应的Solicited-Node组播组发送NS报文。NS报文中目的地址即为该试验地址。如果收到某个其他站点回应的NA报文，就证明该地址已被网络上使用，节点将不能使用该试验地址通讯。

重复地址检测过程

IPv6主机重复地址检测的过程如图5所示：



▲图5: IPv6主机重复地址检测的过程

PC A的IPv6地址2000::1为新配置地址，即2000::1为PC A的试验地址。PC A向2000::1的Solicited-Node组播地址FF02::1:FF00:1发送一个以2000::1为请求的目标地址的NS报文进行重复地址检测，由于2000::1并未正式指定，所以NS报文的源地址为未指定地址。当PC B收到该NS报文后，有两种处理方法：

▶ 如果PC B发现2000::1是自身的一个试验地址，则PC B放弃使用这个地址作为接口地址，并且不会发送NA报文。

▶ 如果PC B发现2000::1是一个已经正常使用的地址，PC B会向FF02::1发送一个NA报文，该消息中会包含2000::1。这样，PC A收到这个消息后就会发现自身的试验地址是重复的，从而弃用该地址。

地址解析

替代IPv4的ARP

在IPv4中，当主机需要和目标主机通信时，需要先通过ARP协议获得目的主机的MAC地址。在IPv6中，同样需要从IP地址解析到MAC地址的功能，邻居发现协议实现了这个功能。

但是IPv6取消了ARP协议，而是通过邻居请求报文NS（Neighbor Solicitation）和邻居通告报文NA（Neighbor Advertisement）来解析三层地址对应的链路层地址。

1. **NS报文**

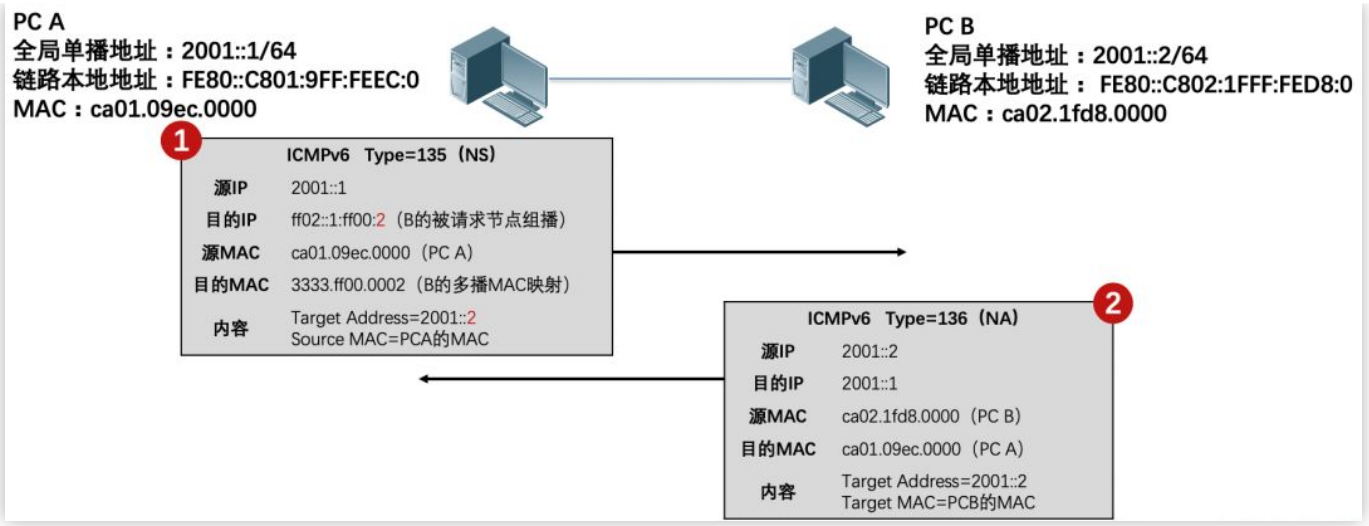
   Type字段值为135，Code字段值为0，在地址解析中的作用类似于IPv4中的ARP请求报文。

2. **NA报文**

   Type字段值为136，Code字段值为0，在地址解析中的作用类似于IPv4中的ARP应答报文。

地址解析过程

IPv6主机地址解析的过程如图6所示：



▲图6: IPv6主机地址解析的过程

① PC A在向PC B发送报文之前它要先解析出PC B的MAC地址，所以首先PC A会发送一个NS报文，其中源IP地址为PC A的IPv6地址，目的IP地址为PC B的被请求节点组播地址（前缀F02::1:F/104，并结合请求IPv6地址中的低24位，具体细节请参阅上一期），需要解析的目标IP为PC B的IPv6地址，这就表示PC A想要知道PC B的MAC地址。同时，NS报文还携带了PC A的MAC地址。

② 当PC B接收到了NS报文之后，就会回应NA报文，其中源地址为PC B的IPv6地址，目的地址为PC A的IPv6地址（使用NS报文中的PC A的MAC地址进行单播），同时包含PC B的MAC地址。这样就完成了一次地址解析的过程。

邻居不可达检测NUD

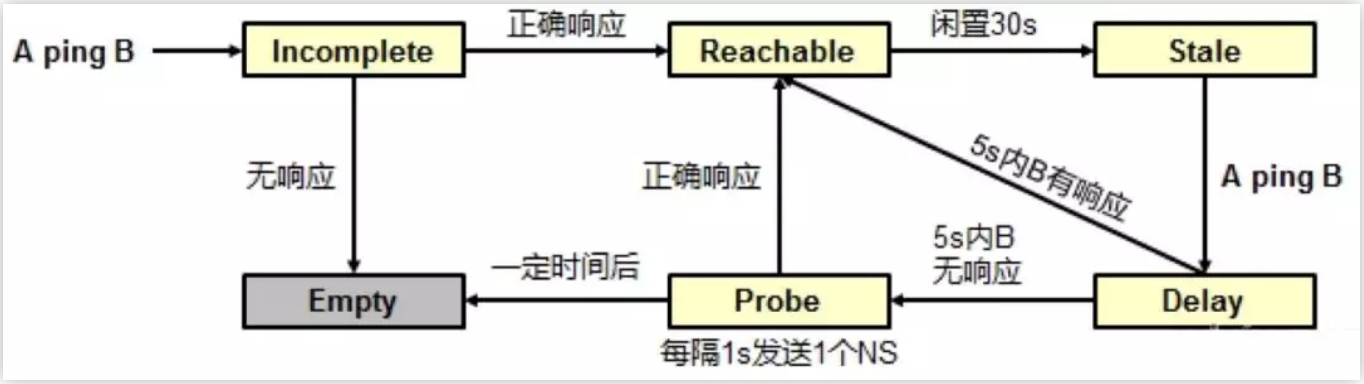
IPv6的邻居状态

NDP的一个重要特征是检测同一链路上以前相连通的两个节点现在是否依然连通，这是通过NUD（Neighbor Unreachability Detection，邻居不可达检测）完成的。NUD帮助维护多个邻居条目组成的邻居缓存，每个邻居都有相应的状态，状态之间可以迁移。
RFC2461中定义了5种邻居状态，分别是：

- **未完成(Incomplete)**：表示正在解析地址，但邻居链路层地址尚未确定。
- **可达(Reachable)**：表示地址解析成功，该邻居可达。
- **陈旧(Stale)**：表示可达时间耗尽，未确定邻居是否可达。
- **延迟(Delay)**：邻居可达性未知。Delay状态不是一个稳定的状态，而是一个延时等待状态。
- **探测(Probe)**：邻居可达性未知,正在发送邻居请求探针以确认可达性.

邻居状态迁移过程

邻居状态的具体迁移过程如图7所示：



▲图7: 邻居状态迁移的具体过程

下面以A、B两个邻居节点之间相互通信过程中A节点的邻居状态变化为例（假设A、B之前从未通信），说明邻居状态迁移的过程。

1. A先发送NS报文，并生成缓存条目，此时，邻居状态为Incomplete。
2. 若B回复NA报文，则邻居状态由Incomplete变为Reachable，否则固定时间后邻居状态由Incomplete变为Empty，即删除表项。
3. 经过邻居可达时间，邻居状态由Reachable（默认30s）变为Stale，即未知是否可达。
4. 如果在Reachable状态，A收到B的非请求NA报文，且报文中携带的B的链路层地址和表项中不同，则邻居状态马上变为Stale。
5. 在Stale状态若A要向B发送数据，则邻居状态由Stale变为Delay，并发送NS请求。
6. 在经过一段固定时间后，邻居状态由Delay（默认5s）变为Probe（每隔1s发送一次NS报文，连续发送3次），其间若有NA应答，则邻居状态由Delay变为Reachable。
7. 在Probe状态，A每隔1s发送单播NS，发送3次后，有应答则邻居状态变为Reachable，否则邻居状态变为Empty，即删除表项。

总结

邻居发现协议NDP是IPv6协议体系中一个重要的基础协议，替代了IPv4的ARP（Address Resolution Protocol）和ICMP路由器发现（Router Discovery），定义了使用ICMPv6报文实现地址解析、跟踪邻居状态、重复地址检测、路由器发现以及重定向等功能。锐捷网络的主流产品，包括交换机、路由器、无线等均支持NDP协议。

更多IPv6相关技术讲解，敬请期待《技术盛宴》后续的IPv6系列文章。

本期作者：杨万里
锐捷网络互联网系统部行业咨询