过时记录从地址表中删除的时间一般为几分钟，因此在过时记录被删除之前，依然可能有发给该设备的包到达交换机。这时，交换机会将包转发到老的端口，通信就会发生错误，这种情况尽管罕见，但的确也有可能发生。不过大家不必紧张，遇到这样的情况，只要重启一下交换机，地址表就会被清空并更新正确的信息，然后网络就又可以正常工作了。

总之，交换机会自行更新或删除地址表中的记录，不需要手动维护[①]。当地址表的内容出现异常时，只要重启一下交换机就可以重置地址表，也不需要手动进行维护。

3.2.3 特殊操作

上面介绍了交换机的基本工作方式，下面来看一些特殊情况下的操作。比如，交换机查询地址表之后发现记录中的目标端口和这个包的源端口是同一个端口。当像图 3.9 这样用集线器和交换机连接在一起时就会遇到这样的情况，那么这种情况要怎么处理呢？首先，计算机 A 发送的包到达集线器后会被集线器转发到所有端口上，也就是会到达交换机和计算机 B（图 3.9 ①）。这时，交换机转发这个包之后，这个包会原路返回集线器（图 3.9 ②），然后，集线器又把包转发到所有端口，于是这个包又到达了计算机 A 和计算机 B。所以计算机 B 就会收到两个相同的包，这会导致无法正常通信。因此，当交换机发现一个包要发回到原端口时，就会直接丢弃这个包。

还有另外一种特殊情况，就是地址表中找不到指定的 MAC 地址。这可能是因为具有该地址的设备还没有向交换机发送过包，或者这个设备一段时间没有工作导致地址被从地址表中删除了。这种情况下，交换机无法判断应该把包转发到哪个端口，只能将包转发到除了源端口之外的所有端口上，无论该设备连接在哪个端口上都能收到这个包。这样做不会产生什么问题，因为以太网的设计本来就是将包发送到整个网络的，然后只有相应的接收者才接收包，而其他设备则会忽略这个包。

① 具备管理功能的高端交换机是提供手动维护地址表的功能的，但一般的低端机型中没有这个功能，想手动维护也不行。

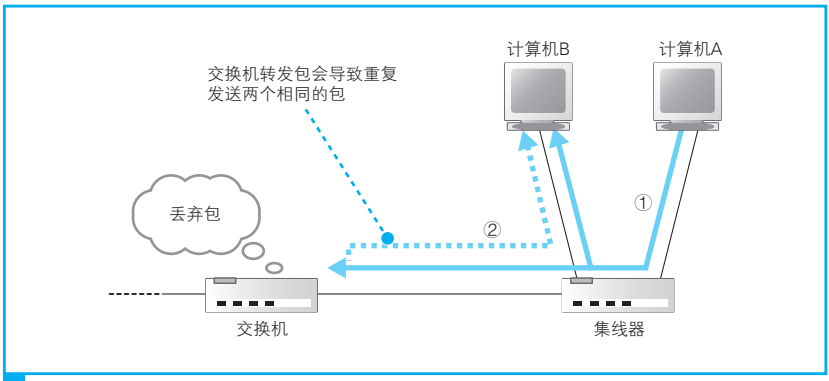

图 3.9 不向源端口转发网络包

有人会说：“这样做会发送多余的包，会不会造成网络拥塞呢？”其实完全不用过于担心，因为发送了包之后目标设备会作出响应，只要返回了响应包，交换机就可以将它的地址写入地址表，下次也就不需要把包发到所有端口了。局域网中每秒可以传输上千个包，多出一两个包并无大碍。

此外，如果接收方 MAC 地址是一个广播地址[①]，那么交换机会将包发送到除源端口之外的所有端口。

3.2.4 全双工模式可以同时进行发送和接收

全双工模式是交换机特有的工作模式，它可以同时进行发送和接收操作，集线器不具备这样的特性。

使用集线器时，如果多台计算机同时发送信号，信号就会在集线器内部混杂在一起，进而无法使用，这种现象称为碰撞，是以太网的一个重要特征。不过，只要不用集线器，就不会发生碰撞。

① 广播地址（broadcast address）是一种特殊的地址，将广播地址设为接收方地址时，包会发送到网络中所有的设备。MAC 地址中的 FF:FF:FF:FF:FF:FF 和 IP 地址中的 255.255.255.255 都是广播地址。

而使用双绞线时，发送和接收的信号线是各自独立的[①]，因此在双绞线中信号不会发生碰撞。网线连接的另一端，即交换机端口和网卡的 PHY（MAU）模块以及 MAC 模块，其内部发送和接收电路也是各自独立的，信号也不会发生碰撞。因此，只要不用集线器，就可以避免信号碰撞了。

如果不存在碰撞，也就不需要半双工模式中的碰撞处理机制了。也就是说，发送和接收可以同时进行。然而，以太网规范中规定了在网络中有信号时要等该信号结束后再发送信号，因此发送和接收还是无法同时进行。于是，人们对以太网规范进行了修订，增加了一个无论网络中有没有信号都可以发送信号的工作模式，同时规定在这一工作模式下停用碰撞检测（图 3.10）。这种工作模式就是全双工模式。在全双工模式下，无需等待其他信号结束就可以发送信号，因此它比半双工模式速度要快[②]。由于双方可以同时发送数据，所以可同时传输的数据量也更大，性能也就更高。

交换机的全双工模式可以同时发送和接收信号。

3.2.5　自动协商：确定最优的传输速率

随着全双工模式的出现，如何在全双工和半双工模式之间进行切换的问题[③]也产生了。在全双工模式刚刚出现的时候，还需要手动进行切换，但这样实在太麻烦，于是后来出现了自动切换工作模式的功能。这一功能可以由相互连接的双方探测对方是否支持全双工模式，并自动切换成相应的

① 1000BASE-T 规格的千兆以太网中，发送和接收信号线不是独立的，而是在同一条线上同时传输两个方向的信号，但 PHY（MAU）模块可以将发送和接收的信号进行分离，因此两者也不会发生碰撞。

② 如果网络中包的数量很少，不会出现等待其他传输结束的情况，那么全双工模式和半双工模式的速度是一样的。

③ 原本以太网并没有工作模式的概念，也没有表示工作模式的专门术语。后来在全双工模式出现时，人们将通信技术中一直使用的全双工、半双工两个词搬到了网络技术中，于是就出现了这样的叫法。

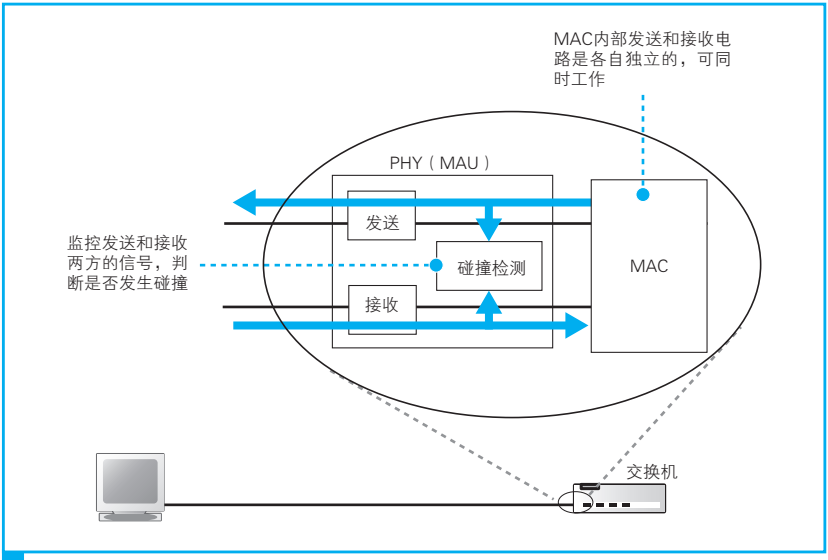

图 3.10 全双工模式的工作方式
在 MAU 的发送和接收电路之间有一个检测信号碰撞的模块，当网络在全双工模式下工作时，发送和接收可同时进行，这一模块就失效了。

工作模式。此外，除了能自动切换工作模式之外，还能探测对方的传输速率并进行自动切换。这种自动切换的功能称为自动协商。

在以太网中，当没有数据在传输时，网络中会填充一种被称为连接脉冲的脉冲信号。在没有数据信号时就填充连接脉冲，这使得网络中一直都有一定的信号流过，从而能够检测对方是否在正常工作，或者说网线有没有正常连接。以太网设备的网线接口周围有一个绿色的 LED 指示灯，它表示是否检测到正常的脉冲信号。如果绿灯亮，说明 PHY（MAU）模块以及网线连接正常[①]。

在双绞线以太网规范最初制定的时候，只规定了按一定间隔发送脉冲信号，这种信号只能用来确认网络是否正常。后来，人们又设计出了如图 3.11 这样的具有特定排列的脉冲信号，通过这种信号可以将自身的状态告知对

① MAC 模块、缓冲区、内存和总线部分的异常无法通过这个指示灯来判断。

方。自动协商功能就利用了这样的脉冲信号，即通过这种信号将自己能够支持的工作模式[①]和传输速率相互告知对方，并从中选择一个最优的组合[②]。

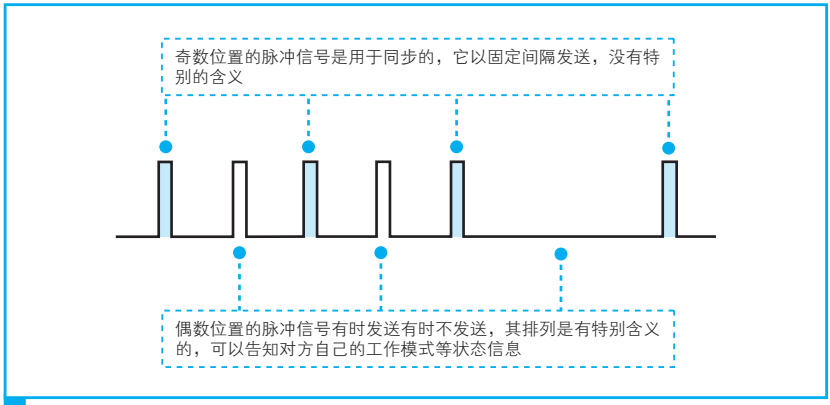

图 3.11　没有传输数据时网络中的信号

下面来看一个具体的例子。假设现在连接双方的情况如表 3.2 所示，网卡一方支持所有的速率和工作模式，而交换机只支持到 100 Mbit/s 全双工模式。当两台设备通电并完成硬件初始化之后，就会开始用脉冲信号发送自己支持的速率和工作模式。当对方收到信号之后，会通过读取脉冲信号的排列来判断对方支持的模式，然后看看双方都支持的模式有哪些。表 3.2 是按照优先级排序的，因此双方都支持的模式就是第 3 行及以下的部分。越往上优先级越高，因此在本例中 100 Mbit/s 全双工模式就是最优组合，于是双方就会以这个模式开始工作。

① 即是否支持全双工模式，以及是否支持半双工模式。

② 自动协商功能是后来才写入以太网规范中的，因此会出现支持这一功能的设备和不支持这一功能的设备混用的情况。在这样的情况下，不支持自动协商的设备由于其所发送的脉冲信号不具备规定的排列，无法正确告知工作模式，所以会引发故障。自动协商规格本身也存在一定的缺陷，这些缺陷有时也会引发故障。因此，尽管有人不喜欢这个功能，但只要正确理解和使用它，就可以防止上述故障。当然，现在基本上已经没有不支持自动协商的旧设备了，因此一般也不会出问题。

表 3.2 自动协商的示例
如果双方设备为以下组合，则最优模式为 100 Mbit/s 全双工。

传输速率 / 工作模式	网 卡	交 换 机
1 Gbit/s 全双工	○	×
1 Gbit/s 半双工	○	×
100 Mbit/s 全双工	○	○
100 Mbit/s 半双工	○	○
10 Mbit/s 全双工	○	○
10 Mbit/s 半双工	○	○

3.2.6 交换机可同时执行多个转发操作

交换机只将包转发到具有特定 MAC 地址的设备连接的端口，其他端口都是空闲的。如图 3.7 中的例子所示，当包从最上面的端口发送到最下面的端口时，其他端口都处于空闲状态，这些端口可以传输其他的包，因此交换机可以同时转发多个包。

相对地，集线器会将输入的信号广播到所有的端口，如果同时输入多个信号就会发生碰撞，无法同时传输多路信号，因此从设备整体的转发能力来看，交换机要高于集线器。

3.3 路由器的包转发操作

3.3.1 路由器的基本知识

网络包经过集线器和交换机之后，现在到达了路由器，并在此被转发到下一个路由器。这一步转发的工作原理和交换机类似，也是通过查表判断包转发的目标。不过在具体的操作过程上，路由器和交换机是有区别的。因为路由器是基于 IP 设计的，而交换机是基于以太网设计的[①]。IP 和以太网的区

① 1.2.1 节和 2.5.1 节对 IP 进行过简单介绍。

别在很多地方都会碰到，我们稍后再具体讲，现在先来看看路由器的概况。

首先，路由器的内部结构如图 3.12 所示。这张图已经画得非常简略了，大家只要看明白路由器包括转发模块和端口模块两部分就可以了。其中转发模块负责判断包的转发目的地，端口模块负责包的收发操作。这一分工模式在第 2 章介绍计算机内部结构的时候也出现过，换句话说，路由器转发模块和端口模块的关系，就相当于协议栈的 IP 模块和网卡之间的关系。因此，大家可以将路由器的转发模块想象成 IP 模块，将端口模块想象成网卡。

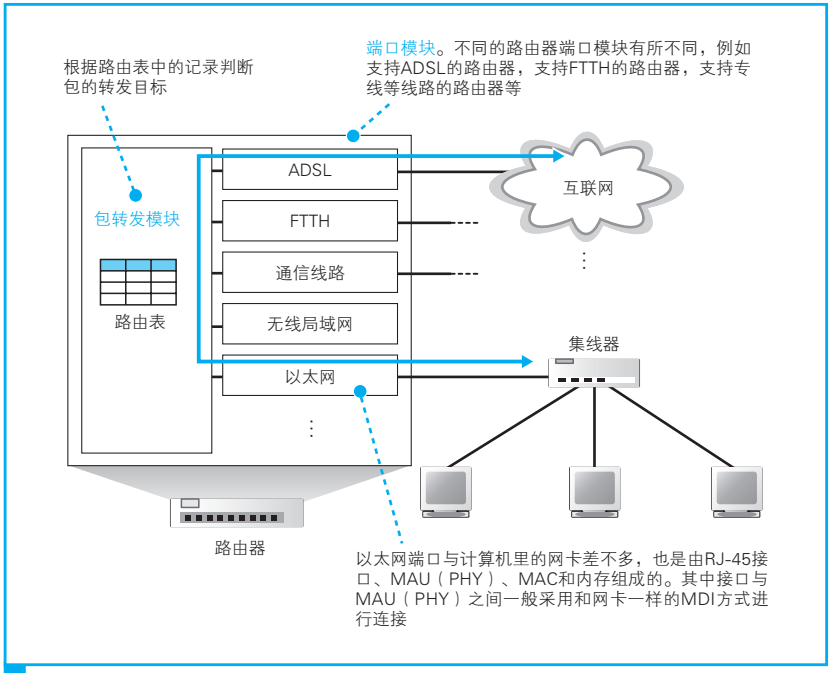

图 3.12　路由器的结构

通过更换网卡，计算机不仅可以支持以太网，也可以支持无线局域网，路由器也是一样。如果路由器的端口模块安装了支持无线局域网的硬件，就可以支持无线局域网了。此外，计算机的网卡除了以太网和无线局域网之外很少见到支持其他通信技术的品种，而路由器的端口模块则支持除局

域网之外的多种通信技术，如 ADSL、FTTH，以及各种宽带专线等，只要端口模块安装了支持这些技术的硬件即可[①]。

看懂了内部结构之后，大家应该能大致理解路由器的工作原理了吧。路由器在转发包时，首先会通过端口将发过来的包接收进来，这一步的工作过程取决于端口对应的通信技术。对于以太网端口来说，就是按照以太网规范进行工作，而无线局域网端口则按照无线局域网的规范工作，总之就是委托端口的硬件将包接收进来。接下来，转发模块会根据接收到的包的 IP 头部中记录的接收方 IP 地址，在路由表中进行查询，以此判断转发目标。然后，转发模块将包转移到转发目标对应的端口，端口再按照硬件的规则将包发送出去，也就是转发模块委托端口模块将包发送出去的意思。

这就是路由器的基本原理，下面再做一些补充。刚才我们讲到端口模块会根据相应通信技术的规范来执行包收发的操作，这意味着端口模块是以实际的发送方或者接收方的身份来收发网络包的。以以太网端口为例，路由器的端口具有 MAC 地址[②]，因此它就能够成为以太网的发送方和接收方[③]。端口还具有 IP 地址，从这个意义上来说，它和计算机的网卡是一样的。当转发包时，首先路由器端口会接收发给自己的以太网包[④]，然后查询转发目标，再由相应的端口作为发送方将以太网包发送出去。这一点和交换机是不同的，交换机只是将进来的包转发出去而已，它自己并不会成为发送方或者接收方。

路由器的各个端口都具有 MAC 地址和 IP 地址。

① 从原理上说，计算机只要安装相应的适配器，也可以支持各种通信技术，但现实中除了局域网之外几乎没有其他需求，因此一般市场上也没有这样的产品。

② 和网卡一样，MAC 地址也是在生产时写入端口的 ROM 中的。

③ 但端口并不会成为 IP 的发送方和接收方。

④ 端口是按照以太网规范接收包的，即当端口的 MAC 地址和包的接收方 MAC 地址一致时，端口才接受这个包，否则就丢弃包。

3.3.2　路由表中的信息

在“查表判断转发目标”这一点上，路由器和交换机的大体思路是类似的，不过具体的工作过程有所不同。交换机是通过 MAC 头部中的接收方 MAC 地址来判断转发目标的，而路由器则是根据 IP 头部中的 IP 地址来判断的。由于使用的地址不同，记录转发目标的表的内容也会不同。

关于细节我们留到后面再讲，现在先来大致介绍一下。路由器中的表叫作路由表，其中包含的信息如图 3.13 所示[①]。

路由器根据“IP 地址”判断转发目标。

最左侧的目标地址列记录的是接收方的信息。这里可能不是很容易理解，实际上这里的 IP 地址只包含表示子网的网络号部分的比特值，而表示主机号部分的比特值全部为 0[②]。路由器会将接收到的网络包的接收方 IP 地址与路由表中的目标地址进行比较，并找到相应的记录。交换机在地址表中只匹配完全一致的记录，而路由器则会忽略主机号部分，只匹配网络号部分。打个比方，路由器在转发包的时候只看接收方地址属于哪个区，×× 区发往这一边，×× 区发往那一边。

在匹配地址的过程中，路由器需要知道网络号的比特数，因此路由表中还有一列子网掩码。子网掩码的含义和第 1 章的图 1.9（b）中介绍的子网掩码基本相同，通过这个值就可以判断出网络号的比特数。

---

① 无论是路由器的路由表，还是我们在第 2 章的图 2.18 中展示的计算机中的路由表，它们的结构和功能都是相同的，只不过每一列的名称可能会有所不同，这是由于厂商和型号的不同导致的。不过，为了便于大家理解，我们在这张图上所使用的列名已经和图 2.18 中的列名进行了统一，因此和实际的路由器中的路由表会有所差异。

② 图 3.13 中也有一些 IP 地址的主机号不是全部为 0，关于这些地址我们稍后会解释，现在请大家先忽略。

路由器的路由表

目标地址（Destination）	子网掩码（Netmask）	网关（Gateway）	接口（Interface）	跃点数（Metric）
10.10.1.0	255.255.255.0	——	e2	1
10.10.1.101	255.255.255.255	——	e2	1
192.168.1.0	255.255.255.0	——	e3	1
192.168.1.10	255.255.255.255	——	e3	1
0.0.0.0	0.0.0.0	192.0.2.1	e1	1

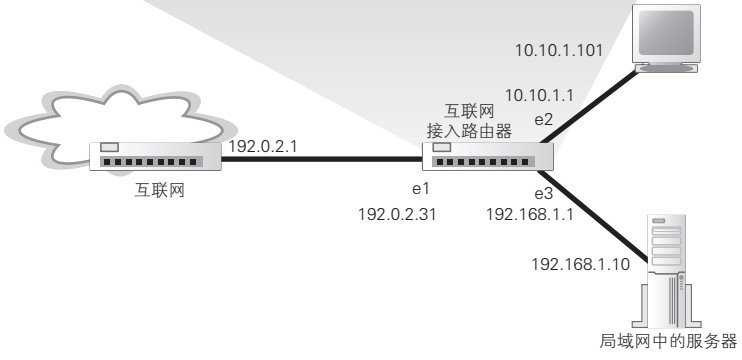

图 3.13 路由器根据路由表对包进行转发

路由器会忽略主机号，只匹配网络号。

上面这些介绍可以帮助大家大致理解路由器的工作方式，如果要进一步深入，还需要再思考一些问题。刚才我们说过，目标地址列中的 IP 地址表示的是子网，但也有一些例外，有时地址本身的子网掩码和路由表中的子网掩码是不一致的，这是路由聚合的结果。路由聚合会将几个子网合并成一个子网，并在路由表中只产生一条记录。要搞清楚这个问题，我们还是看一个例子。如图 3.14 所示，我们现在有 3 个子网，分别为

10.10.1.0/24、10.10.2.0/24、10.10.3.0/24，路由器 B 需要将包发往这 3 个子网。在这种情况下，路由器 B 的路由表中原本应该有对应这 3 个子网的 3 条记录，但在这个例子中，无论发往任何一个子网，都是通过路由器 A 来进行转发，因此我们可以在路由表中将这 3 个子网合并成 10.10.0.0/16，这样也可以正确地进行转发，但我们减少了路由表中的记录数量，这就是路由聚合。经过路由聚合，多个子网会被合并成一个子网，子网掩码会发生变化，同时，目标地址列也会改成聚合后的地址。

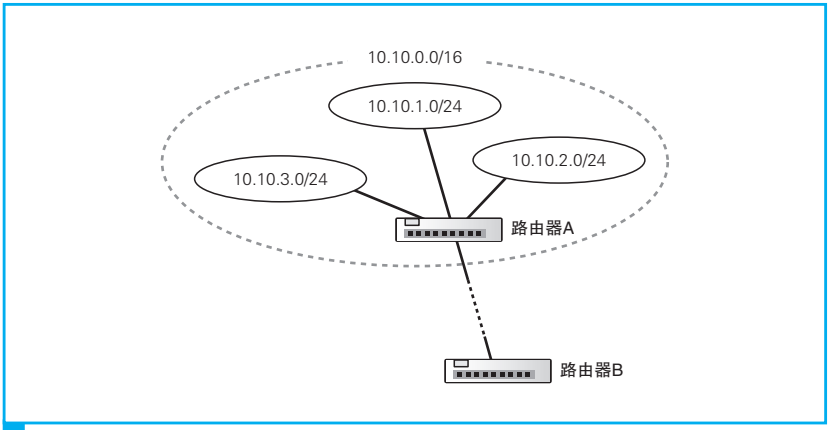

图 3.14　路由聚合

相对地，还有另外一些情况，如将一个子网进行细分并注册在路由表中，然后拆分成多条记录。

从结果上看，路由表的子网掩码列只是用来在匹配目标地址时告诉路由器应该匹配多少个比特。而且，目标地址中的地址和实际子网的网络号可能并不完全相同，但即便如此，路由器依然可以正常工作。

此外，通过上述方法，我们也可以将某台具体计算机的地址写入路由表中，这时的子网掩码为 255.255.255.255，也就是说地址中的全部 32 个比特都为 1。这样一来，主机号部分比特全部为 0 可以表示一个子网，主机号部分比特不全部为 0 可以表示某一台计算机，两种情况可以用相同的

规则来处理[①]。

路由表的子网掩码列只表示在匹配网络包目标地址时需要对比的比特数量。

关于目标地址和子网掩码我们先讲到这里。接下来在子网掩码的右边还有网关和接口两列，它们表示网络包的转发目标。根据目标地址和子网掩码匹配到某条记录后，路由器就会将网络包交给接口列中指定的网络接口（即端口）[②]，并转发到网关列中指定的 IP 地址。

最后一列是跃点计数，它表示距离目标 IP 地址的距离是远还是近。这个数字越小，表示距离目的地越近；数字越大，表示距离目的地越远。

路由表记录维护的方式和交换机也有所不同。交换机中对 MAC 地址表的维护是包转发操作中的一个步骤[③]，而路由器中对路由表的维护是与包转发操作相互独立的，也就是说，在转发包的过程中不需要对路由表的内容进行维护。

对路由表进行维护的方法有几种，大体上可分为以下两类。

（a）由人手动维护路由记录

（b）根据路由协议机制，通过路由器之间的信息交换由路由器自行维护路由表的记录

其中（b）中提到的路由协议有很多种，例如 RIP、OSPC、BGP 等都属于路由协议。

① 图 3.13 的第 2 行和第 4 行就是这样的例子。

② 在路由器的范畴中，接口和端口表示同一个意思，但从历史上的惯用法来看，接口和端口两种叫法都有。

③ 3.2.2 节对交换机如何在地址表中更新转发目标信息进行了介绍。

3.3.3　路由器的包接收操作

下面我们来看一看路由器的整个工作过程。首先，路由器会接收网络包。路由器的端口有各种不同的类型，这里我们只介绍以太网端口是如何接收包的。以太网端口的结构和计算机的网卡基本相同，接收包并存放到缓冲区中的过程也和网卡几乎没有区别。

首先，信号到达网线接口部分，其中的 PHY（MAU）模块和 MAC 模块将信号转换为数字信息，然后通过包末尾的 FCS 进行错误校验，如果没问题则检查 MAC 头部中的接收方 MAC 地址，看看是不是发给自己的包，如果是就放到接收缓冲区中，否则就丢弃这个包。如果包的接收方 MAC 地址不是自己，说明这个包是发给其他设备的，如果接收这个包就违反了以太网的规则。

> 路由器的端口都具有 MAC 地址，只接收与自身地址匹配的包，遇到不匹配的包则直接丢弃。

3.3.4　查询路由表确定输出端口

完成包接收操作之后，路由器就会丢弃包开头的 MAC 头部。MAC 头部的作用就是将包送达路由器，其中的接收方 MAC 地址就是路由器端口的 MAC 地址。因此，当包到达路由器之后，MAC 头部的任务就完成了，于是 MAC 头部就会被丢弃。

> 通过路由器转发的网络包，其接收方 MAC 地址为路由器端口的 MAC 地址。

接下来，路由器会根据 MAC 头部后方的 IP 头部中的内容进行包的转发操作。转发操作分为几个阶段，首先是查询路由表判断转发目标。关于

具体的工作过程，我们还是来看一个实际的例子，如图 3.13 的情况，假设地址为 10.10.1.101 的计算机要向地址为 192.168.1.10 的服务器发送一个包，这个包先到达图中的路由器。判断转发目标的第一步，就是根据包的接收方 IP 地址查询路由表中的目标地址栏，以找到相匹配的记录。就像前面讲过的一样，这个匹配并不是匹配全部 32 个比特，而是根据子网掩码列中的值判断网络号的比特数，并匹配相应数量的比特[①]。例如，图 3.13 的第 3 行，子网掩码列为 255.255.255.0，就表示需要匹配从左起 24 个比特。网络包的接收方 IP 地址和路由表中的目标地址左起 24 个比特的内容都是 192.168.1，因此两者是匹配的，该行记录就是候选转发目标之一。

按照这样的规则，我们可能会匹配到多条候选记录。在这个例子中，第 3、4、5 行都可以匹配[②]。其中，路由器首先寻找网络号比特数最长的一条记录[③]。网络号比特数越长，说明主机号比特数越短，也就意味着该子网内可分配的主机数量越少，即子网中可能存在的主机数量越少，这一规则的目的是尽量缩小范围，所以根据这条记录判断的转发目标就会更加准确。我们来看图 3.13 中的例子。

第 3 行 192.168.1.0/255.255.255.0 表示一个子网，第 4 行 192.168.1.10/255.255.255.255 表示一台服务器。相比服务器所属的子网来说，直接指定服务器本身的地址时范围更小，因此这里应该选择第 4 行作为转发目标。按照最长匹配原则筛选后，如果只剩一条候选记录，则按照这条记录的内容进行转发。

然而，有时候路由表中会存在网络号长度相同的多条记录，例如考虑到路由器或网线的故障而设置的备用路由就属于这种情况。这时，需要根据跃点计数的值来进行判断。跃点计数越小说明该路由越近，因此应选择跃点计数较小的记录。

---

① 前面讲过，路由表中的子网和实际的子网并非完全一致，因此说“匹配的是网络号”可能并不准确，但这样说可能更容易理解。

② 第 5 行为什么会匹配的原因我们稍后再解释，大家先知道可以匹配就好了。

③ 这一规则称为“最长匹配”原则。

如果在路由表中无法找到匹配的记录，路由器会丢弃这个包，并通过ICMP[①]消息告知发送方[②]。这里的处理方式和交换机不同，原因在于网络规模的大小。交换机连接的网络最多也就是几千台设备的规模，这个规模并不大[③]。如果只有几千台设备，遇到不知道应该转发到哪里的包，交换机可以将包发送到所有的端口上，虽然这个方法很简单粗暴，但不会引发什么问题。然而，路由器工作的网络环境就是互联网，它的规模是远远大于以太网的，全世界所有的设备都连接在互联网上，而且规模还在持续扩大，未来的互联网里到底会有多少设备，我们谁都说不准。在如此庞大的网络中，如果将不知道应该转发到哪里的包发送到整个网络上，那就会产生大量的网络包，造成网络拥塞。因此，路由器遇到不知道该转发到哪里的包，就会直接丢弃。

3.3.5　找不到匹配路由时选择默认路由

既然如此，那么是不是所有的转发目标都需要配置在路由表中才行呢？如果是公司或者家庭网络，这样的做法也没什么问题，但互联网中的转发目标可能超过 20 万个，如果全部要配置在路由表中实在是不太现实。

其实，大家不必担心，因为之前的图 3.13 路由表中的最后一行的作用就相当于把所有目标都配置好了。这一行的子网掩码为 0.0.0.0，关键就在这里，子网掩码 0.0.0.0 的意思是网络包接收方 IP 地址和路由表目标地址的匹配中需要匹配的比特数为 0，换句话说，就是根本不需要匹配。只要将子网掩码设置为 0.0.0.0，那么无论任何地址都能匹配到这一条记录，这样就不会发生不知道要转发到哪里的问题了。

---

① ICMP：Internet Control Message Protocol，Internet 控制报文协议。当包传输过程中发生错误时，用来发送控制消息。

② 2.5.11 节介绍过 ICMP。

③ 这里几千台设备的规模指的是以太网的规模。交换机本身的设计并不需要按照这个规模，但由于它是基于以太网进行工作的，因此其规模和以太网的规模是一致的。