第五章 网络层

一、本章知识点

1、概念

（1） 网络层的功能

网络层的任务是将源计算机发出的数据分组（数据报）经过适当的路径送到目的地计算机，从源端到目的端可能要经过若干中间节点。这一功能与数据链路层有很大的区别，数据链路层仅把数据帧从线缆或信道的一端传到另一端。因此，网络层是处理计算机网络中端到端数据传输的最低层。

（2） 存储转发分组交换机制（Store-and-Forward Packet Switching）

一台主机要发送一个分组（数据报），那么它将分组传送给最近的路由器，该路由器或者在它自己的 LAN 上，或者在一条通向承运商的点到点链路上。该分组被存储在路由器上，一直到它完全到达路由器为止，所以路由器可以验证它的校验和。然后它被沿路转发到下一台路由器，直到到达目标主机为止，最后在目标主机上它被递交给相应的进程。

（3） 向传输层提供的服务：虚电路和数据报

- 虚电路：网络层向传输层提供了面向连接的服务，在发送分组（数据报）之前，必须首先建立起一条从源路由器到目标路由器之间的路径。该连接成为一个 VC（virtual circuit，虚电路）。
- 分组（数据报）：网络层向传输层提供了无连接的服务，所有的分组（数据报）都被独立地传送到子网中，并且独立于路由，不需要提前建立任何辅助设施。

2、路由选择算法

（1） 最优化原则

如果路由器 J 是在从路由器 I 到路由器 K 的最优路径上，那么，从 J 到 K 的最优路径也必定沿着同样的路由路径。

（2） 最短路径选择算法（Dijkstra 最短路由搜索算法）

首先为通信子网建立一个子网图，图中的每个节点代表一个网络节点（路由器），每条弧代表一条通信新路（链路），弧上的标注代表两个相邻节点之间的权值。然后把每个节点用从源节点沿已知最佳路径到本节点的或距离来标注。

- 算法思想：设 G=(V,E)是一个带权有向图，把图中顶点集合 V 分成两组，第一组为已求出最短路径的顶点集合（用 S 表示，初始时 S 中只有一个源点，以后每求得一条最短路径 ，就将加入到集合 S 中，直到全部顶点都加入到 S 中，算法就结束了），第二组为其余未确定最短路径的顶点集合（用 U 表示），按最短路径长度的递增次序依次把第二组的顶点加入 S 中。在加入的过程中，总保持从源点 v 到 S 中各顶点的最短路径长度不大于从源点 v 到 U 中任何顶点的最短路径长度。此外，每个顶点对应一个距离，S 中的顶点的距离就是从 v 到此顶点的最短路径长度，U 中的顶点的距离，是从 v 到此顶点只包括 S 中的顶点为中间顶点的当前最短路径长度。
- 算法步骤：

A. 初始时，S 只包含源点，即 S＝{v}，v 的距离为 0。U 包含除 v 外的其他顶点，即:U={其余顶点}，若 v 与 U 中顶点 u 有边，则<u,v>正常有权值，若 u 不是 v 的出边邻接点，则<u,v>权值为∞。

B. 从 U 中选取一个距离 v 最小的顶点 k，把 k，加入 S 中（该选定的距离就是 v 到 k 的最短路径长度）。

C. 以 k 为新考虑的中间点，修改 U 中各顶点的距离；若从源点 v 到顶点 u 的距离（经过顶点 k）比原来距离（不经过顶点 k）短，则修改顶点 u 的距离值，修改后的距离值的顶点 k 的距离加上边上的权。

D. 重复步骤 B 和 C 直到所有顶点都包含在 S 中。

（3） 扩散算法（Flooding）

在扩散法中，每一个入境分组将被路由器转发到除了它进来的那条路线之外的每一条输出线路上。

由于扩散算法会产生大量的重复分组，需要改进扩散算法，避免重复的分组。

（4） 距离矢量路由选择（Distance Vector routing）

是一种动态路由算法。在一个距离矢量路由选择算法中，所有的节点都定期地将它们的距离表传送给所有与之直接邻接的节点，包括：

- 每条路径的目的地（另一节点）
- 路径的代价（距离）

所有的节点都监听从其他节点传送来的路由选择更新信息，并在下列情况下更新它们的路由选择表：

A. 被通告一条新的路径，该路由在本节点的路由表中部存在，此时本地系统加入这条新的路由；

B. 通过发送来路由信息的节点有一条到达某个目的地的路由，该路由比当前使用的路由有较短的距离（较小的代价）。在这种情况下，就用经过发送路由信息的节点的新路由替换路由表中到达那个目的地的现有路由。

C. 在本节点的现有路由表中为了到达某一目的地首先应前往的下一节点如果通告了一个较高的代价，就要使用这一新的代价更新从本节点前往同一目的地的代价。

无穷计算（Count-to-infinity）问题：距离矢量路由算法的主要缺点是网络规模的伸展性差。它对链路状态变化的响应慢，需要大尺寸的路由信息报文交换，并且报文的长度与通信子网内的个数成正比。由于距离矢量协议需要每个存储转发的节点都参与路由信息的交换，因而交换信息的交通量也可能巨大，并产生无穷计算（Count-to-infinity）问题。

（5） 链路状态路由选择

是一种动态路由算法，通常包括以下 5 个步骤：

A. 发现邻居节点

发现它的邻居节点，并知道其网络地址；

B. 测量线路开销

测量到各邻居节点的延迟或者开销；

C. 创建链路状态分组 LSP

构造一个分组，分组中包含所有它刚刚知道的信息；这个分组的内容包含了发送方的标识，以及是一个序列号（Seq）和年龄（Age），以及一个邻居列表。对于每个邻居，同时也要给出到这个邻居的延迟。

D. 泛洪链路状态分组

使用 flooding 算法，将这个分组发送给所有其他的路由器；

E. 计算最短路径

路由器根据网络拓扑计算出到每一个其他路由器的最短路径。

3、拥塞控制

（1）拥塞（congestion）

当一个子网或子网的一部分中出现太多分组的时候，网络的性能开始下降。这种情况称之为拥塞，网络资源上有太多的分组时，将会导致网络性能下降，即对资源需求的总和大于可用资源。

产生拥塞的原因：

- 低带宽线路
- 多个输入对应一个输出
- 节点缓冲容量太小
- 结点处理机速度不高

（2）拥塞控制的通用原则

包括开环的（open loop）和闭环的（close loop）方法。开环属于预防性技术，通过采用开环控制技术，预先避免拥塞的发生；闭环属于反应性技术，一旦网络发生拥塞随之做出反应，即通过采用闭环控制技术缓解或消除拥塞。

（3）拥塞控制途径：增加资源和减少负载

按照解决方案应用在不同的时间尺度从慢（预防性）到快（反应性）的顺序依次是

- 网络供给
- 流量感知路由
- 准入控制
- 流量限制：抑制分组（choke packet）、逐跳抑制分组（逐跳反压）、显示拥塞通知（IP 中的 ECN 及 TCP 中均采用）
- 负载脱落：当路由器来不及处理分组而被淹没时，只要将这些分组丢弃即可。

  负载丢弃采用葡萄酒（wine）或牛奶（milk）策略。

  随机的早期检测（Random Early Detection）算法：在实际耗尽所有的缓冲区空间之间就开始丢弃分组。

4、服务质量管理（QoS）

（1）服务质量的评价标准

- 流（flow）：从一个源到一个目标的分组流（steam）称之为流。
- 评价标准：可靠性、延迟、抖动和带宽

（2）流量整形

- 流量整形（traffic shaping）：在服务器端对流量进行平滑处理，即调节数据传输的平均速率（以及突发性）。
- **漏桶和令牌桶算法**

漏桶（leaky bucket）算法：每个主机连接到网络的接口中都包含一个漏桶，即含有一个有限长度的内部队列。如果当该队列满的时候，又有一个分组到来，那么该分组将被丢弃，亦即是，如果在一台主机上，队列中的分组数目已经达到了最大值，这是又有一个或者多个进程要发送分组，那么新发送的分组将被丢弃。

令牌桶（token bucket）算法：桶中保存的是令牌，每个令牌包含一定数量的字节的分组，令牌桶允许主机积累发送全，直至到达桶的最大尺寸，因而相对于漏桶，允许有一定的突发流量。

（3）分组调度

- 公平排队（fair queueing）：路由器为每一条输出线路使用一组单独的队列，每个流一个队列。当一条线路空闲的时候，路由器轮流扫描这些队列，从下一个队列中取出第一个分组。
- 加权的公平排队（weighted fair queueing）：通过不同的主机赋予不同的优先级进行加权调节，从而提高算法的效率。

（4）准入控制

5、网络互联

（1）互联网络的分组转发

（2）隧道技术

隧道（tunneling）：当源和目标主机位于相同类型的网络中，中间的网络属于不同类型时使用隧道方案。

（3）互联网路由

- 互联网定义了两级路由算法：在每个网络内部使用的内部网关协议（interior gateway protocol）和在网络之间使用的外部网关协议（exterior gateway protocol）。
- 自治系统（AS，Autonomous System）

自治系统：在单一的技术管理下的一组路由器，而这些路由器使用一种 AS 内部的路由选择协议和共同的度量以确定分组在该 AS 内的路由，同时还使用一种 AS 之间的路由选择协议用以确定分组在 AS 之间的路由。

现在对自治系统 AS 的定义是强调下面的事实：尽管一个 AS 使用了多种内部路由选择协议和度量，但重要的是一个 AS 对其他 AS 表现出的是一个单一的和一致的路由选择策略。

- 因特网有两大类路由选择协议

内部网关协议 IGP (Interior Gateway Protocol)：即在一个自治系统内部使用的路由选择协议。目前这类路由选择协议使用得最多，如 RIP 和 OSPF 协议。

外部网关协议 EGP (External Gateway Protocol)：若源站和目的站处在不同的自治系统中，当数据报传到一个自治系统的边界时，就需要使用一种协议将路由选择信息传递到另一个自治系统中。这样的协议就是外部网关协议 EGP。在外部网关协议中目前使用最多的是 BGP-4。

（4）分段与重组

分段：互联网协议定义一个足够小的基本分段长度值，以便于基本的分段能够通过每一个网络。

1、 Internet 上的网络层

（1）IP 协议

- IP 分组

一个 IP 数据报由首部和数据两部分组成。首部的前一部分是固定长度，共 20 字节，是所有 IP 数据报必须具有的。在首部的固定部分的后面是一些可选字段，其长度是可变的。

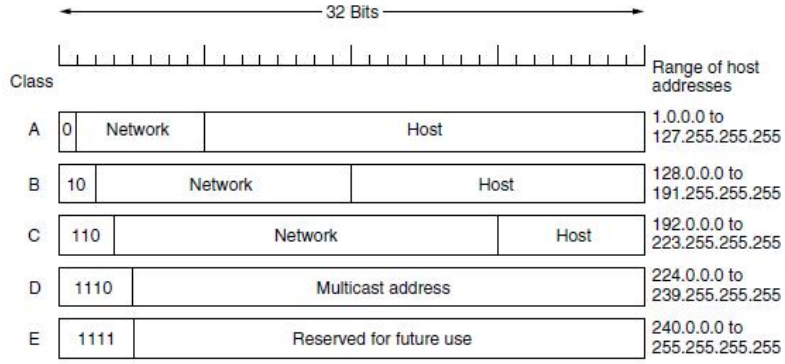
（2）IP 地址

- 分类 IP 地址

每一类地址都由两个固定长度的字段组成，其中一个字段是网络号 net-id，它标志主机（或路由器）所连接到的网络，而另一个字段则是主机号 host-id，它标志该主机（或路由器）。

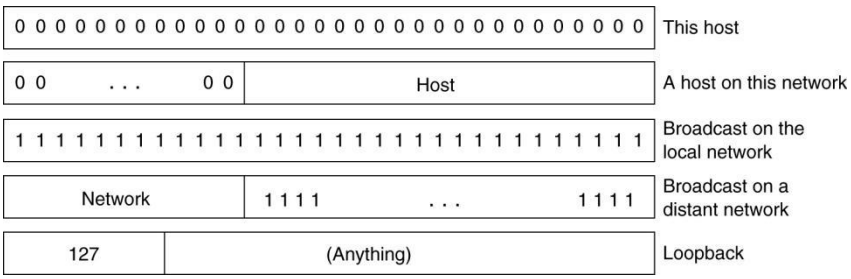
IP 地址 ::= { <网络号>, <主机号>}

如下图，分类 IP 地址的结构



-
- 内部 IP 地址：三类特殊的私有 IP 地址（注意：私有 IP 地址在公网不能用）
  - ◆ 10.0.0.0 – 10.255.255.255/8
  - ◆ 172.16.0.0 – 172.31.255.255/12 (16: 0001 0000)
  - ◆ 192.168.0.0 – 192.168.255.255/16

- 特殊 IP 地址



- IP 地址块相关问题

地址块大小：2^m

可容纳的主机数量：2^m -2（全 0 是第一个，一般作为网络号使用，全 1 是最后一个，一般作为广播地址使用，因此需要减去两个）

（3）地址解析协议 ARP

不管网络层使用的是什么协议，在实际网络的链路上传送数据帧时，最终还是必须使用硬件地址。每一个主机都设有一个 ARP 高速缓存(ARP cache)，里面有所在的局域网上的各主机和路由器的 IP 地址到硬件地址的映射表。

当主机 A 欲向本局域网上的某个主机 B 发送 IP 数据报时，就先在其 ARP 高速缓存中查看有无主机 B 的 IP 地址。如有，就可查出其对应的硬件地址，再将此硬件地址写入 MAC 帧，然后通过局域网将该 MAC 帧发往此硬件地址。

- ARP 高速缓存的作用

为了减少网络上的通信量，主机 A 在发送其 ARP 请求分组时，就将自己的 IP 地址到硬件地址的映射写入 ARP 请求分组。

当主机 B 收到 A 的 ARP 请求分组时，就将主机 A 的这一地址映射写入主机 B 自己的 ARP 高速缓存中。这对主机 B 以后向 A 发送数据报时就更方便了。

（4）子网

- **划分子网的基本思路**

划分子网纯属一个单位内部的事情。单位对外仍然表现为没有划分子网的网络。

从主机号借用若干个位作为子网号 subnet-id，而主机号 host-id 也就相应减少了若干个位。

IP 地址 ::= {<网络号>, <子网号>, <主机号>}

凡是从其他网络发送给本单位某个主机的 IP 数据报，仍然是根据 IP 数据报的目的网络号 net-id，先找到连接在本单位网络上的路由器。然后此路由器在收到 IP 数据报后，再按目的网络号 net-id 和子网号 subnet-id 找到目的子网。最后就将 IP 数据报直接交付目的主机。

- 子网掩码

从一个 IP 数据报的首部并无法判断源主机或目的主机所连接的网络是否进行了子网划分。使用子网掩码(subnet mask)可以找出 IP 地址中的子网部分。

子网掩码是一个网络或一个子网的重要属性。

路由器在和相邻路由器交换路由信息时，必须把自己所在网络（或子网）的子网掩码告诉相邻路由器。

路由器的路由表中的每一个项目，除了要给出目的网络地址外，还必须同时给出该网络的子网掩码。

若一个路由器连接在两个子网上就拥有两个网络地址和两个子网掩码。

- 分组路由转发过程

A. 从收到的分组的首部提取目的 IP 地址 D。

B. 先用各网络的子网掩码和 D 逐位相“与”，看是否和相应的网络地址匹配。若匹配，则将分组直接交付（所谓直接交付，即分组的目的 IP 地址就是下一跳的 IP 地址）。否则就是间接交付，执行(3)。

C. 若路由表中有目的地址为 D 的特定主机路由，则将分组传送给指明的下一跳路由器；否则，执行(4)。

D. 对路由表中的每一行的子网掩码和 D 逐位相“与”，若其结果与该行的目的网络地址匹配，则将分组传送给该行指明的下一跳路由器；否则，执行(5)。

E. 若路由表中有一个默认路由，则将分组传送给路由表中所指明的默认路由器；否则，执行(6)。

F. 报告转发分组出错。

（5）无分类编址（classless addressing） CIDR

- CIDR 最主要的特点

A. CIDR 消除了传统的 A 类、B 类和 C 类地址以及划分子网的概念，因而可以更加有效地分配 IPv4 的地址空间。

B. CIDR 使用各种长度的“网络前缀”(network-prefix)来代替分类地址中的网络号和子网号。

- IP 地址从三级编址（使用子网掩码）又回到了两级编址。

IP 地址 ::= {<网络前缀>, <主机号>}

CIDR 把网络前缀都相同的连续的 IP 地址组成“CIDR 地址块”

128.14.32.0/20 表示的地址块共有 212 个地址（因为斜线后面的 20 是网络前缀的位数，所以这个地址的主机号是 12 位）。这个地址块的起始地址是 128.14.32.0。
在不需要指出地址块的起始地址时，也可将这样的地址块简称为“/20 地址块”。

128.14.32.0/20 地址块的最小地址：128.14.32.0

128.14.32.0/20 地址块的最大地址：128.14.47.255

全 0 和全 1 的主机号地址一般不使用。

- 路由聚合（汇聚）

前缀长度不超过 23 位的 CIDR 地址块都包含了多个 C 类地址。

CIDR 地址块中的地址数一定是 2 的整数次幂。网络前缀越短，其地址块所包含的地址数就越多。

- 最长前缀匹配

使用 CIDR 时，路由表中的每个项目由“网络前缀”和“下一跳地址”组成。在查找路由表时可能会得到不止一个匹配结果。

应当从匹配结果中选择具有最长网络前缀的路由：最长前缀匹配(longest-prefix matching)。

网络前缀越长，其地址块就越小，因而路由就越具体(more specific) 。

最长前缀匹配又称为最长匹配或最佳匹配。

二、相关协议和设备

1、协议

IP 协议、IP 地址、ICMP 协议、OSPF、BGP、IPv6、ARP 协议、DHCP 协议

2、设备

路由器，NAT 盒子

第六章 传输层

一、解决的主要问题

本章描述传输层的基本功能和服务、协议实现的要素、以及因特网的传输层协议——TCP 和 UDP 的主要原理。

二、本章知识点

1、主要概念

（1）端到端通信：传输层通过使用端口号做标识，向应用层提供了进程间的逻辑通信的功能。

（2）端口号：因特网的传输层地址，服务器端采用固定的常用端口，客户端采用由操作系统动态分配的短暂端口。

（3）三次握手：传输层的连接建立和连接释放方式。

（4）拥塞控制：当网络中的负载超过网络资源（路由器的处理能力和缓存）时，将会出现传输时延过长甚至丢包的情况，即网络拥塞。拥塞控制是指通过采用某种策略，避免拥塞，或者在拥塞出现时缓解拥塞。

（5）最大最小公平性：在端到端的共享链路上，最大限度保证每个数据流的最小带宽需求。

（6）伪报头（Pseudo-header）：TCP 和 UDP 在计算校验和时，需增加 12 个字节的伪报头（包含源 IP 地址和目的 IP 地址等），伪报头不会传输到网络中。

2、算法

（1）TCP 的拥塞控制算法

- 慢启动：启动速率很低，拥塞窗口初始值为 1 个最大报文段长度（MSS）。在达到阈值或者发现丢包之前，拥塞窗口按指数级增长
- AI：当拥塞窗口达到阈值时，按线性增长
- MD：出现重发定时器超时，拥塞窗口降为最小值（1 个 MSS），阈值改为当前窗口的一半，重新开始新的慢启动
- 快速恢复：收到三个重复的 ACK 时，拥塞窗口降为当前窗口的一半，开始新的 AI。

（2）Nagle 算法

TCP 的发送端使用此算法来提高传输效率，一次尽可能发送较大的数据量（1 个 MSS）。

（3）Clark 算法

TCP 的接收端使用此算法来提高传输效率，只在有较大缓存（至少为 1/2 MSS）时，才发送窗口更新通知。

（4）Jacobson 算法

用于估算端到端环回时延 RTT。

$$RTT\text{估值} = \alpha \times RTT\text{估值的历史值} + (1-\alpha) \times RTT\text{的测量值}$$

3、传输层协议

（1）TCP

提供面向连接的、可靠的字节流服务，不保证应用层的消息边界；采用三次握手建立连接、四步释放连接，默认采用 Go-back-N ARQ 协议进行差错控制，采用动态的滑动窗口进行流量控制，提供拥塞控制功能。

（2）UDP

提供无连接、尽力而为的传输服务，采用校验和方式进行差错检测，不提供差错恢复功能。

三、相关协议和设备

1、协议

TCP、UDP

2、设备

无

第七章 应用层

一、解决的主要问题

本章描述应用层通信的客户服务器模型以及常用的应用层协议的要点，包括 DNS、Email 相关协议和 HTTP 等。

二、本章知识点

1、主要概念

（1）客户服务器（C/S）：传统的网络应用采用的通信模型，服务器向客户提供资源和服务，服务器程序先运行，等待客户的请求；通信由客户发起，服务器收到请求之后，将需要的资源返回给客户。

（2）P2P：对等模型，两个网络应用程序（Peer，对等体）在通信时没有严格的客户和服务器的区分，每一个 Peer 既有客户的功能也有服务器的功能。

（3）解析器：DNS 的客户端程序，发送域名请求给本地 DNS 服务器。

（4）递归解析：本地名字服务器向上级名字服务器发送查询请求，上级服务器转发给更上一级服务器，到达根服务器之后再向下转发，直至请求转发给权威名字服务器；查询结果再反向依次转发给本地名字服务器的过程。

（5）迭代解析：本地名字服务器向上级名字服务器发送查询请求，上级名字服务器返回更上一级名字服务器的 IP 地址，以此类推，每一次查询请求均由本地名字服务器发送，直至请求权威服务器的过程。

（6）MIME（多用途因特网邮件扩展）：为邮件头和 HTTP 消息头提供了内容类型扩展，以支持除纯文本之外的其它数据类型，包括图像、音频、视频、压缩文件等。

（7）浏览器：WWW 的客户端程序，除了 HTTP，还支持 FTP、Email 等协议。

（8）Webmail: 用户使用浏览器和 HTTP 协议与邮件服务器通信，收发邮件

（9）统一资源定位符 URL：URL：protocol://domain_name:port/item_name

（10）HTTP 操作过程

（11）持久连接：HTTP1.1 的功能，在一个 TCP 连接上传输一个网页中的多个文件，所有文件传输完之后再关闭连接。

2、主要应用层协议

（1）DNS：域名系统，实现域名与 IP 地址等资源的映射。

（2）SMTP：简单邮件传送协议，实现将电子邮件从用户代理传送到发件人的邮件服务器，以及将邮件从发件人邮件服务器传送到收件人邮件服务器的功能。

（3）POP3：邮局协议，实现将电子邮件从用户的邮件服务器的邮箱下载到用户本地的功能。

（4）IMAP：因特网邮件访问协议，实现将电子邮件从用户的邮件服务器的邮箱下载到用户本地的功能，比 POP3 功能更强。

（5）HTTP：超文本传送协议，浏览器发送 HTTP 请求给 Web 服务器，服务器通过 HTTP 响应将所请求的资源发送给浏览器。默认端口：80，基于 TCP 协议传输。

三、相关协议和设备

1、协议

DNS、SMTP、POP3、IMAP4、MIME、HTTP

2、设备：无

四、重点和难点

（1）DNS 的域名解析过程

（2）Email 发送和接收过程

（3）HTTP1.1 的特点。