1.报文(message)
我们将位于应用层的信息分组称为报文。报文是网络中交换与传输的数据单元，也是网络传输的单元。报文包含了将要发送的完整的数据信息，其长短不需一致。报文在传输过程中会不断地封装成分组、包、帧来传输，封装的方式就是添加一些控制信息组成的首部，那些就是报文头。

2.报文段（segment）
通常是指起始点和目的地都是传输层的信息单元。

3.分组/包(packet)
分组是在网络中传输的二进制格式的单元，为了提供通信性能和可靠性，每个用户发送的数据会被分成多个更小的部分。在每个部分的前面加上一些必要的控制信息组成的首部，有时也会加上尾部，就构成了一个分组。它的起始和目的地是网络层。

4.数据报(datagram)
面向无连接的数据传输，其工作过程类似于报文交换。采用数据报方式传输时，被传输的分组称为数据报。通常是指起始点和目的地都使用无连接网络服务的网络层的信息单元。(指 IP 数据报)

5.帧(frame)
帧是数据链路层的传输单元。它将上层传入的数据添加一个头部和尾部，组成了帧。它的起始点和目的点都是数据链路层。

6.数据单元（data unit）
指许多信息单元。常用的数据单元有服务数据单元（SDU）、协议数据单元（PDU）。
SDU 是在同一机器上的两层之间传送信息。PDU 是发送机器上每层的信息发送到接收机器上的相应层（同等层间交流用的）。

应用层——消息
传输层——报文段(segment)/数据报(datagram) (注：TCP 叫 TCP 报文段，UDP 叫 UDP 数据报，也有人叫 UDP 段)
网络层——分组、数据包（packet）
链路层——帧（frame）
物理层——P-PDU（bit）

第一章 计算机网络与互联网络

1.1 什么是互联网络

从构成的角度来看：

互联网络

点：（端系统，网络应用） + 路由器

边：链路

互联网络 是网络的网络

从服务的角度来看：互联网络=能够为应用提供通信服务的通信架构（有连接可靠的服

务和无连接的不可靠服务）+使用通信服务相互配合工作的应用

协议：对等层实体在通信过程中所遵循的规则的集合　*理解*

语法+语义+时序

(1) 语义。语义是解释控制信息每个部分的意义。它规定了需要发出何种控制信息，以及完成的动作与做出什么样的响应。

(2) 语法。语法是用户数据与控制信息的结构与格式，以及数据出现的顺序。

(3) 时序。时序是对事件发生顺序的详细说明。（也可称为“同步”）。

人们形象地把这三个要素描述为：语义表示要做什么，语法表示要怎么做，时序表示做的顺序。

1.2 网络边缘

网络的结构= 网络边缘（应用，主机）+网络核心（路由器）+接入网络与通信链路

网络边缘：运行应用的端系统　*理解*

（端系统中的应用交互方式）

C/S 模式，特点

服务器-客户机模式，由客户而不是服务提供者发起动作；服务器被动的等待来自客户机的请求；客户机和服务器通过一条通信信道连接起来。两个进程间的通信链路称为连接。连接在内部表现为一些缓冲区和一组协议机制，在外部表现出比无连接高的可靠性。一个完整的网间进程通信需要由两个进程组成，并且只能使用同一种高层协议。因此，一个完整的网间通信需要协议，本机地址、本地端口号、远程端口号、远程地址这五个元素标识。

P2P 模式，特点

P2P 对位于数据中心的专用服务器有着最小的依赖。相反，不同主机上的 P2P 应用程序会相互通信，这样的主机称之为对等方。P2P 是一个去中心化的网络体系结构，每一个主机既是客户机又是服务器，以 P2P 下载为例，每个主机之间都接收着来自不同主机的下载资源，也为不同主机提供者下载资源，类似于“我帮你，你帮我”。P2P 最引人入胜的是他们的自扩展性，比如在一个 P2P 文件共享应用中，尽管每个对等方都由于请求文件产生了工作负载，但是每个对等方通过向其他对等方分发文件也为系统增加了服务能力。

利用网络的服务

面向连接的服务

无连接的服务

1.3 网络核心

网络核心

组成：网络交换设备如：路由器+链路

功能：数据交换

数据交换方式及比较

分组交换：以分组为单位存储转发方式， 统计复用　*理解*

分组交换也称为包交换，它将用户通信的数据划分成多个更小的等长数据段，在每个数据段的前面加上必要的控制信息作为数据段的首部，每个带有首部的数据段就构成了一个分组。首部指明了该分组发送的地址，当交换机收到分组之后，将根据首部中的地址信息将分组转发到目的地，这个过程就是分组交换。能够进行分组交换的通信网被称为分组交换网。

分组交换的本质就是存储转发，它将所接受的分组暂时存储下来，在目的方向路由上排队，当它可以发送信息时，再将信息发送到相应的路由上，完成转发。其存储转发的过程就是分组交换的过程。

为了有效地利用通信线路,希望一个信道同时传输多路信号，这就是所谓的多路复用技术

统计时分复用（动态分配带宽）

统计时分复用的基本原理是把时间划分为不等长的时间片，长短不同的时间片就是传送不同长度分组所需的时间，对每路通信没有固定分配时间片，而是按需使用。这就意味着使用这条复用线传送分组时间的长短，由此可见统计时分复用是动态分配带宽的。

按照实现方式，分组交换可以分为数据报分组交换和虚电路分组交换。

VC 虚电路：在网络层建立起逐渐之间的连接，然后两主机进行通讯

它与数据报方式的区别主要是在信息交换之前，需要在发送端和接收端之间先建立一个逻辑连接，然后才开始传送分组，所

有分组沿相同的路径进行交换转发，通信结束后再拆除该逻辑连接。网络保证所传送的分组按发送的顺序到达接收端。所以网络提供的服务是可靠的，也保证服务质量。这种方式对信息传输频率高、每次传输量小的用户不太适用，但由于每个分组头只需标出虚电路标识符和序号，所以分组头开销小，适用长报文传送。

Datagram 数据报：两主机通信前无需建立连接，分组以数据报的形式来通讯

数据包分组交换要求通信双方之间至少存在一条数据传输通路。发送者需要在通信之前将所要传输的数据包准备好，数据包都包含有发送者和接收者的地址信息。数据包的传输彼此独立，互不影响，可以按照不同的路由机制到达目的地，并重新组合。在这种方式中，每个分组按一定格式附加源与目的地址、分组编号、分组起始、结束标志、差错校验等信息，以分组形式在网络中传输。网络只是尽力地将分组交付给目的主机，但不保证所传送的分组不丢失，也不保证分组能够按发送的顺序到达接收端。所以网络提供的服务是不可靠的，也不保证服务质量。其优点是传输延时小，当某节点发生故障时不会影响后续分组的传输。缺点是每个分组附加的控制信息多，增加了传输信息的长度和处理时间，增大了额外开销。

线路交换

电路交换就是由交换机负责在两部通信站点（如两部电话机）之间建立一条专用的物理线路分配给双方传输数据使用。

FDM 频分多路复用

用户在分配到一个频带后，从始至终都使用这个频带

TDM 时分多路复用

将时间划分为一段段等长 TDM 帧表示一个周期，每个 TDM 帧被分为多个时隙，每个用户占有固定一个时隙

WDM 波分多路复用

光的频分多路复用

1.4 网络接入与物理媒介

将端系统连接到边缘路由器的链路或网络

住宅接入：点到点接入

ADSL

非对称数字用户线路，它利用数字编码技术从现有铜质电话线上获取最大数据传输容量，同时又不干扰在同一条线上进行的常规话音服务。ADSL 的关键概念，也是数字信号与模拟信号能同时在电话线传输的关键，在于其上行与下行的带是不对称的。也就是从 ISP 以客户端（下行通道）传输的带宽比较高，客户端到 ISP（上行通道）的传输带宽比较低。

HFC

混合光纤同轴电缆是一种结合光纤与同轴电缆的宽带接入网，由光纤取代一般电缆线，作为有线电视网络中的主干。由头端（Head End）机房到用户附近的光纤节点（Fiber Node）之间的传输介质为光纤，由光纤节点到用户的终端设备则是 RG-62 等电缆线，因而称之为混合光纤同轴电缆。

Cable Modem

电缆调制解调器（Cable Modem，CM），Cable 是指有线电视网络，Modem 是调制解调器。电缆调制解调器是在有线电视网络上用来上互联网的设备，它是串接在用户家的有线电视电缆插座和上网设备之间的。

Home Networks

家庭网络是融合家庭控制网络和多媒体信息网络于一体的家庭信息化平台，是在家庭范围内实现信息设备、通信设备、娱乐设备、家用电器、自动化设备、照明设备、保安（监控）装置及水电气热表设备、家庭求助报警等设备互连和管理，以及数据和多媒体信息共享的系统。

机构接入：LAN

局域网自然就是局部地区形成的一个区域网络，其特点就是分布地区范围有限，可大可小，大到一栋建筑楼 与相邻建筑之间的连接，小到可以是办公室之间的联系。局域网自身的组成大体由计算机设备、网络连接设备、网络传输介质 3 大部分构成，其中，计算机设备又包括服务器与工作站，网络连接设备则包含了网卡、集线器、交换机，网络传输介质简单来说就是网线，由同轴电缆、双绞线及光缆 3 大原件构成。

以太网络

以太网是一种计算机局域网技术。IEEE 组织的 IEEE 802.3 标准制定了以太网的技术标准，它规定了包括物理层的连线、电子信号和介质访问层协议的内容。

WLAN

无线局域网，指应用无线通信技术将计算机设备互联起来，构成可以互相通信和实现资源共享的网络体系。

物理链路

导向型介质

电磁波 沿着 固体 媒介传播；如 ： 光纤 ， 双绞线 ；

非导向型介质

电磁波 在自由空间中传播 ；如 ： 空气 ， 真空 ， 水中 ；

常用介质

TP 双绞线

同轴电缆

光纤

Radio

1.5 互联网结构与 ISP

ISP—互联网业务提供商，比如家庭 ISP，机构 ISP，大学 ISP 之类的，移动接入是移动的，在有基站的地方都可以，端系统通过接入 ISP 访问互联网，整个 Internet 就是将所有运营商 ISP 的网络给连接到一起组成的网络的网络

近似层次型结构

T-1 ISP（global ISP）

T-2 ISP（Redional ISP）区域性 ISP

Local ISP 更小规模的 ISP

ISP 之间的连接

对等连接：2 个 ISP 对等互接

LXP：多个对等 ISP 互联互通之处

内容提供商网络

ICP （internet content provider）互联网内容提供商，向广大用户综合提供互联网信息业务和增值业务的电信运营商。比如百度，是提供上层的业务，而 ISP 是提供网络的，提供连接。

在全球部署数据中心机房 DC

内容提供商网络在多处与各个 ISP 相联

内容提供商自己部署网络将全球的 DC 相联

内容提供商 DC 自己之间的访问，通过自己部署的专网

用户接入后通过离用户最近的 DC 为之服务

1.6 分组交换网络中的延迟与丢失

延迟的 4 个原因 （计算） *掌握*

处理延迟

分组到达路由器，路由器要对这个分组进行处理，处理包括检查分组有没有错误（差错检测），根据分组携带的地址信息决定将分组在哪个链路发出去。

排队延迟

当分组确定从某一链路传出，但这一链路刚好在传输其他的分组，就要在结点排队等待。

传输延迟

当某一链路开始传输分组时，从分组的第一个 bit 开始到最后一个 bit 发完，需要的时间称为传输延迟。排队延迟取决于分组长度和链路带宽，分组越长所需时间也越长，链路带宽表示能以怎样的速率发送。

传播延迟

当分组从某一结点发出后，到达下一个结点所花的时间，传播延迟取决于物理链路的长度和信号的传播速度。

eg：n 段，分组 L，R，传播延迟 d/每段，如何计算总体延迟；

流量强度：La/R *掌握*

a：当前时间内到达的分组个数；L：分组的长度；R：链路的输出能力；结果越接近 1，排队延迟越大

排队延迟 依赖流量强度公式；

丢失原因：缓冲区溢出+出错没通过校验

吞吐量：了解

瞬间吞吐量

平均吞吐量

瓶颈链路决定了主机之间的吞吐量（从每段链路获得的大致带宽是 1/N，瓶颈链路是所有链路段中获得带宽最小的）

1.7 协议层次与服务模型

为什么要分层：网络比较复杂，分层实现比较容易设计，调试，实现；

分层：将复杂的网络功能划分成功能明确的层次，上层利用下层提供的服务来实现本层的协议，从而为上层提供更复杂的功能；

一些术语和概念： *理解*

服务、服务访问原语、服务访问点

并非在一个层内完成的全部功能都称为服务，只有那些能够被高一层实体“看得见”的功能才能称为服务。

上层使用下层所提供的服务必须通过与下层交换一些命令，这些命令在 OSI 中称为服务原语。

在同一系统中相邻两层的实体进行交互(即交换信息)的地方，通常称为服务访问点 SAP(Service Access Point)。

面向连接的服务、无连接的服务

协议、协议数据单元 PDU

协议是控制两个对等实体(或多个实体)进行通信的规则的集合。

协议的语法方面规则定义了所交换的信息的格式。

协议的语义方面规则定义了发送者或接收者所要完成的操作，例如，在何种条件下数据必须重传或丢弃。

服务和协议之间的关系（区别与联系）

协议是水平的，即协议是控制对等实体之间通信的规则。

服务是垂直的，即服务是由下层向上层通过层间接口提供的。

在协议的控制下，两个对等实体间的通信使得本层能够向上一层提供服务。要实现本层协议，还需要使用下面一层所提供的服务。

使用本层服务的实体只能看见服务而无法看见下面的协议。

互联网络分层模型及每一层的功能 *理解*

应用层

两个分布式应用交换报文，实现网络应用

传输层

将网络层提供的主机到主机的服务区分为进程到进程的服务，同时加强网络层的服务（网络层的服务可能不可靠）

拆包组包 TCP：保证数据包的完整性 和 以及处理传输过程中可能发生的危险 UDP：发完就完了

报文----拆分成---->包

网络层

在链路层提供的点到点的服务的基础之上，扩展为端到端的服务

路由和地址解析。选择适当的网络节点进行路由。

包 ---->数据帧

链路层

提供两点之间，以帧为单位的传输

控制对物理设备的访问 规定数据如何在不同物理设备上进行传出 并提供数据纠错功能。在不可靠的物理设备上提供可靠传输 数据

数据帧

物理层

将链路层的数据流变为物理信号进行发送；接收到物理信号后将之翻转为数据流

定义通信的物理设备的规格。网线接口类型，光纤接口类型，传输速率等

封装和解封装

从发送端进行五层的封装，到接受的进行五层的解封装，交换机进行二层解封装再封装，路由器进行三层的解封装再封装（解封装到三层的时候查路由表，匹配，找到下一跳，根据吓一跳的 IP 地址通过 ARP 协议找到对应的 MAC 地址，通过这一跳对应的网卡，把它放出去）

地址解析协议，即 ARP（Address Resolution Protocol），是根据 IP 地址获取物理地址的一个 TCP/IP 协议。主机发送信息时将包含目标 IP 地址的 ARP 请求广播到局域网络上的所有主机，并接收返回消息，以此确定目标的物理地址；收到返回消息后将该 IP 地址和物理地址存入本机 ARP 缓存中并保留一定时间，下次请求时直接查询 ARP 缓存以节约资源。

1.8 历史 了解

ARPANET：美国军方资助的分组交换实验网，由于 TCP/IP 架构的包容性、免费使用、架构便于应用创新吸引更多的用户等原因，用户数量、节点数量和应用数量越来越多

NSF 美国自然科学基金会建立 ARPANET 的访问网（变成双主板，军方、民用）

民用网络从军用网络脱开，成为现在的互联网络

术语：IETE（ITU，IEEE）、RFC

IETE 国际互联网工程任务组（The Internet Engineering Task Force，简称 IETF）是一个公开性质的大型民间国际团体，汇集了与互联网架构和互联网顺利运作相关的网络设计者、运营者、投资人和研究人员，并欢迎所有对此行业感兴趣的人士参与。IETF 的主要任务是负责互联网相关技术标准的研发和制定，是国际互联网业界具有一定权威的网络相关技术研究团体。

ITU 国际电联是主管信息通信技术事务的联合国机构，负责分配和管理全球无线电频谱与卫星轨道资源，制定全球电信标准，向发展中国家提供电信援助，促进全球电信发展。

IEEE 电气与电子工程师协会，总部位于美国纽约，是一个国际性的电子技术与信息科学工程师的协会，也是全球最大的非营利性专业技术学会。

RFC(Request For Comments)请求评议，是因特网上一类文件的统称：通俗的说，某家机构或团体，开发出一套网络标准或对标准的构想，想征询因特网上的意见，就会在网上发一份 RFC，对这一问题感兴趣的人可以阅读该 RFC 并提出自己的意见；绝大部分网络标准的指定都是以 RFC 的形式开始，经过大量的论证和修改过程，由主要的标准化组织所指定的。”

第二章 应用层

原理+应用实例+SOCKET 编程

应用的开发只集中在端系统上，对路由器没有任何改变=》互联网架构鼓励应用创新

2.1 网络应用原理

应用架构

C/S

P2P

混合

进程间通信

同主机：操作系统定义的通信方法，eg：管道、共享缓存、信号量

不同主机：利用网络提供的架构交换报文

SOCKET

活在应用层和传输层中间的，传输层把接收到的报文段中东西取出来个通过 socket 给对应的应用

一个整数，OS 用于表示应用通信关系所采用的本地标示

TCP：连接的本地标示

UDP：端节点的本地标示

进程编址：IP+PORT（本质上在传输层上应用了端口号，用于区分应用，TCP 和 UDP 使用端口号的方式不同）

应用所需要的服务需要考虑的因素（网络所提供服务的主要指标）

丢失率可靠性

延迟、延迟差（抖动）

网络抖动是指网络发生拥塞的情况下，排队产生的延迟会影响端到端的延迟，并导致通过同一连接传输的分组延迟各不相同，而抖动就是用来描述这样一延迟变化的程度，一般网络抖动值指的是网络通信中延迟最大值与最小值之差，网络抖动值越小说明网络质量越稳定。

带宽

安全性

传输层协议

TCP 提供的服务特性：可靠字节流服务，面向连接，流量控制，拥塞控制

UDP 提供的服务特性：无连接，不可靠的数据报服务

都能提供进程的标示，区分不同的进程

2.2 WEB 和 HTTP

Web 是 World Wide Web 的简称，Web 提供了全新的信息发布和浏览模式，实际上 Web 是运行在 Internet 之上的所有 Web 服务器和所管理对象的集合，对象主要包括网页和程序。Web 是基于浏览器/服务器（B/S）的一种体系结构，客户在计算机上使用浏览器向 Web 服务器发送请求，服务器响应客户请求，向客户回送所请求的网页，客户在浏览器窗口上显示网页的内容。

Web 体系结构主要由三部分组成：

Web 服务器：必须有一个服务器来提供用户要访问的 Web 页面或这些资源，这种服务器就是 Web 服务器，也称为网站。

客户端：用户一般是通过浏览器访问 Web 资源的，它是运行在客户端的一种软件。

通信协议：客户端和服务器之间采用 HTTP（超文本传输协议）进行通信，HTTP 是客户浏览器和 Web 服务器通信的基础。

WEB 应用包括：

HTTP 协议

HTML

HTML 的全称为超文本标记语言，是一种标记语言。它包括一系列标签. 通过这些标签可以将网络上的文档格式统一，使分散的 Internet 资源连接为一个逻辑整体。HTML 文本是由 HTML 命令组成的描述性文本，HTML 命令可以说明文字，图形、动画、声音、表格、链接等。

CLIENT，SERVER

术语：网页，对象，引用 URL

Web 页：由一些对象组成，对象可以是 HTML 文件、JPEG 图像、Java 小程序、声音剪辑文件等 。Web 页含有一个基本的 HTML 文件，该基本 HTML 文件又包含若干对象的引用（链接）。通过 URL 对每个对象进行引用 （URL 包含访问协议，用户名，口令字，端口等）

统一资源定位系统（URL）是因特网的万维网服务程序上用于指定信息位置的表示方法。

HTTP 协议

定义了 C 和 S 之间通信的报文格式，解释和时序

HTTP 连接

持续性连接

在一个（在客户端和服务器之间的）TCP 连接上可以传输多个对象

非持续性连接

一个 TCP 连接上只能发送一个对象，下载多个对象需要多个 TCP 连接

往返延迟 RTT 和对象的抓取时间

从发送端发送数据开始，到发送端收到来自接收端的确认总共经历的时间（传输时间忽略）

响应时间：

一个 RTT 用来发起 TCP 连接

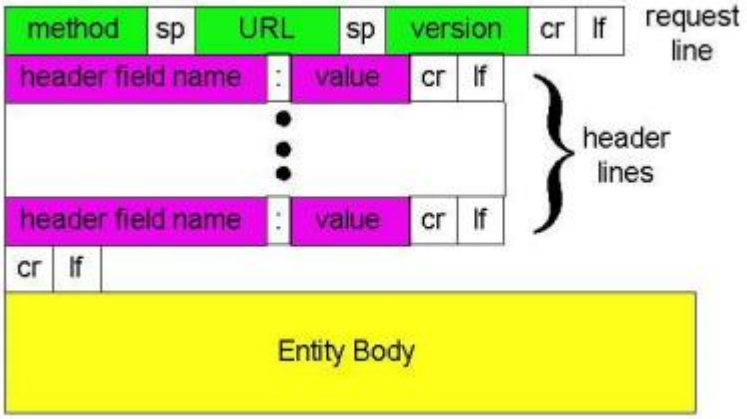
一个 RTT 用来 HTTP 请求并等待 HTTP 响应

文件传输时间

共：2RTT+传输时间

报文格式

请求报文



方法

GET 方法用来请求访问服务器上的资源，该资源由请求 URI 指定

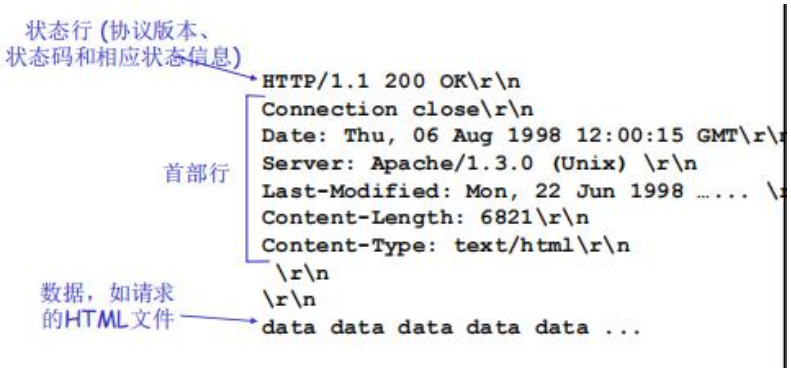
POST 方法一般用来传输实体主体，所谓的实体就是要传输的有效载荷数据

PUT 方法用于上传文件，需要在请求报文的主体中包含文件内容，然后保存到请求 URI 指定的位置

HEAD 方法和 GET 方法的区别只是 HEAD 方法不返回报文主体部分，它用于获取与指定 URI 有关的首部信息

DELETE 方法和 PUT 方法相反，它用于删除请求 URI 指定的资源

响应报文



状态码

2xx:成功

200 OK 请求成功，请求对象包含在响应报文的后续部分

3xx:重定向

301 Moved Permanently 请求的对象已经被永久转移了：新的 URL 在响应报文的 Location:首部行中指定客户端软件自动用新的 URL 去获取对象

4xx:客户端错误

400 Bad Request 一个通用的差错代码，表示该请求不能被服务器解读

404 Not Found 请求的文档在该服务上没有找到

5xx:服务器错误

505 HTTP Version Not Supported 服务器不支持请求中指明的 HTTP 协议版本

COOKIES

HTTP 是一种无状态的协议，但是服务器需要维护客户端的状态，因此大多数主要的门户网站使用 cookies 来改善

1.在 HTTP 响应报文中有一个 cookie 的首部行

2.在 HTTP 请求报文含有一个 cookie 的首部行

3.在用户端系统中保留有一个 cookie 文件，由用户的浏览器管理

4.在 Web 站点有一个后端数据库存储各用户端的 cookie

WEB 缓存

作用：通过本地命中，减少这些对象的访问延迟；进一步减少接入链路的流量强度，从而降低派对延迟带来总体平均延迟的减少；减轻服务器的负担。

2.4 EMAIL

电子邮件应用的构成

用户代理

用户代理就是用户与电子邮件系统的接口。其必须具有撰写、显示、处理（例如阅读后删除、存盘、打印、转发）、通信这四项功能。

邮件服务器

它的功能是发送和接收邮件，同时还要向发件人报告邮件传送的结果。

SMTP 协议

简单邮件传输协议。是一组用于从源地址到目的地址传输邮件的规范，通过它来控制邮件的中转方式。用于用户代理向邮件服务器发送邮件或在邮件服务器之间发送邮件。

邮件报文格式解析

报文头

报文体

MIME：邮件多媒体扩展，可以在邮件中编码多媒体内容

邮件存取协议

作用

常用

IMAP

交互式邮件存取协议，它是跟 POP3 类似邮件访问标准协议之一。不同的是，开启了 IMAP 后，您在电子邮件客户端收取的邮件仍然保留在服务器上，同时在客户端上的操作都会反馈到服务器上，如：删除邮件，标记已读等，服务器上的邮件也会做相应的

动作。所以无论从浏览器登录邮箱或者客户端软件登录邮箱，看到的邮件以及状态都是一致的。

POP3

邮局协议的第 3 个版本，是规定怎样将个人计算机连接到 Internet 的邮件服务器和下载电子邮件的电子协议。

2.5 DNS

域名系统，是互联网的一项服务。它作为将域名和 IP 地址相互映射的一个分布式数据库，能够使人更方便地访问互联网。

DNS 作用：完成域名到 IP 地址的转换（还包括，别名->正规名字；邮件服务器名字->正规名字转换等），应用层面的互联网基础设施。其他应用使用的应用。

DNS 的概念：分布式、层次数据库

命名是分层的：

域名信息存储和服务是分布的，每个域名服务器担任一个区域 ZONE 的名字到 IP 地址的权威转换，也缓存名字-ip 信息的转换

DNS 的构成

解析器：本地应用

域名服务器

DNS 协议

报文：请求和应答格式相同

RR：资源记录

域名解析过程（解析器->本地 DNS 服务器->上层域名服务器->...->权威名字服务器，返回）

递归解析

迭代解析

DNS 缓存

DNS 缓存是由操作系统维护的临时数据库。它存储了以前的 DNS 信息。有助于减少 DNS 服务器的网络流量和 DNS 的平均时延。

DNS 缓存的原理非常简单。 当 DNS 服务器收到 DNS 的应答时，它能将包含某域名到 IP 地址的映射缓存在本地。

作用

本地缓存+服务器缓存

2.6 P2P（了解 P2P 的概念和优势即可）

P2P 概念：每个对等体既是客户端又是服务器

P2P 网络是这些 peer 构成的应用层面的逻辑网络

P2P 网络比 C/S 方式分发内容快的原因：这些 peer 节点参与到内容的上载，流量和服务都是分布的，具可扩展性；

典型 P2P 应用及其原理（不要求）

Napster（集中式目录）

Gnutella（查询泛洪）

KaZaA（半集中+泛洪）

BT

DHT

2.7 视频流化服务和 CDN 了解

视频流化：视频一边下一边看

服务器向客户端进行视频流化的方式：UDP 流化，http 流化，DASH（Dynamic，Adaptive Streaming over HTTP）

DASH 流化的过程

服务器将一个视频分隔为若干个块，每个块都独立存储，有着不一样的编码来决定不一样的速率；客户端在获取的时候，需要获取告示文件，告示文件提供了每个块的信息、url 等，若客户端当前缓存区足够大，通往服务器的带宽足够大，那就会选择相对而言更加清晰的块，所以会话中不同的时刻，可以切换不同的编码块，这取决于当时的可用带宽。

客户端获得告示文件

客户端按照情况，向（可能是不同的）服务器请求不同视频质量的内容块，客户端智能；

CDN：

其实就是存储一些资源，但用户向服务器发送请求资源的时候，如果每一次都需要向服务器发送请求，那在这个过程就会浪费掉许多资源，所以有了 CDN，CDN 供应商会向这些服务器提供服务，用来存储他们需要面向用户的资源：静态资源或者动态资源，当用户请求的时候就可以向这个 CDN 请求资源

CDN+视频流化服务工作流程：内容上载到 CDN 节点->用户认证->域名解析重定向->用户得到最合适的 CDN 节点->节点提供 DASH 服务

单个服务器，或者服务器群向客户端提供海量内容并发服务的问题：扩展性差

CDN：原理

应用层面的协作服务网络

在全网部署缓存节点，内容预先部署到 CDN 缓存节点上；

用户请求通过域名解析重定向向离自己最近的节点请求内容

缓存节点放置的方式

Enter Deep

将 CDN 服务器深入到许多接入网，更接近用户，数量多，离用户近，但是管理困难

Bring Home

部署在少数的（十个左右）关键位置，在服务器周边的关键位置

2.9 TCP 的 SOCKET 编程 *理解*

SOCKET 概念：TCP 连接的本地标示，向这个 SOCKET 写就是发送给对方的进程：从 SOCKET 中读，就是读取对方发送过来的数据

SOCKET API：创建，使用（读和写），关闭；

TCP SOCKET 数据传输的特点：面向连接，可靠字节流服务

TCP SOCKET 编程

建立 SOCKET

客户端 TCP 实体动作：和服务器端的 TCP 实体握手沟通

服务器端的动作

三次握手

使用 SOCKET

关闭 SOCKET

2.10 UDP 的 SOCKET 编程

UDP SOCKET 数据传输的特点

编程 不要求

建立 SOCKET（此前客户端 UDP 实体和服务器不用握手，不为之后的通信做准备）

使用

关闭

第三章 传输层

3.1 传输层服务

传输服务：能够使端系统应用之间进行逻辑通信

传输协议：运行于端系统的 2 个对等传输层实体相互通信应该遵守的规则集合

传输服务和网络服务的区别

网络服务：主机到主机的通信

传输服务：进程到进程的通信

互联网络传输层协议

TCP：有连接，可靠保序数据传输服务

UDP：无连接，不可靠，不保序的数据传输服务

3.2 复用与解复用

复用：源端多个上层应用收集数据：应用报文，封装报文

解复用：接收端将数据按照端口号（结合 IP 地址）给相应的 SOCKET 对应的应用

复用和解复用的工作原理：IP PORT

TCP 有连接情况：SOCKETS 为 4 元组

UDP 无连接情况：SOCKETS 为 2 元组

3.3 无连接传输层协议 UDP

UDP 的必要性：有些应用对实时性比较在乎，对可靠性要求不高

UDP 报文（无连接的，因此叫做 UDP 数据报）格式

UDP 报文校验和的计算 *理解*

3.4 可靠数据传输原理

协议演进的方式讲解如何进行 RDT

加入一些假设，简单的协议可以提供 RDT 服务

去掉一些假设，需要协议实体做相应的变化从而能够进行 RDT

技术机制 *理解*

校验和，正向确认，反向确认

序号：检查重复

只有正向确认的机制

检错重发和超时重发：处理丢失

滑动窗口

利用率计算 *了解*

停止等待技术：链路带宽延迟积（容量） 效率低

管道技术：在未经对方确认的情况下，可以连续发送多个 PDU（协议数据单元）

GBN 回退 N 步协议：发送窗口大于 1，接收窗口=1（只能顺序接收；发送方只设置一个超时定时器，一旦出错，返回到出错的那一个 PDU 重发）

容许发送方发送多个分组，而不需要等待确认，但他也受限与在流水线中未确认的分组数不能超过某个最大数 N。

SR 选择重传协议：发送窗口大于 1，接收窗口大于 1（能够乱序接收；发送方为每个发送出去的 PDU 设置超时定时器，哪个超时重发哪个）

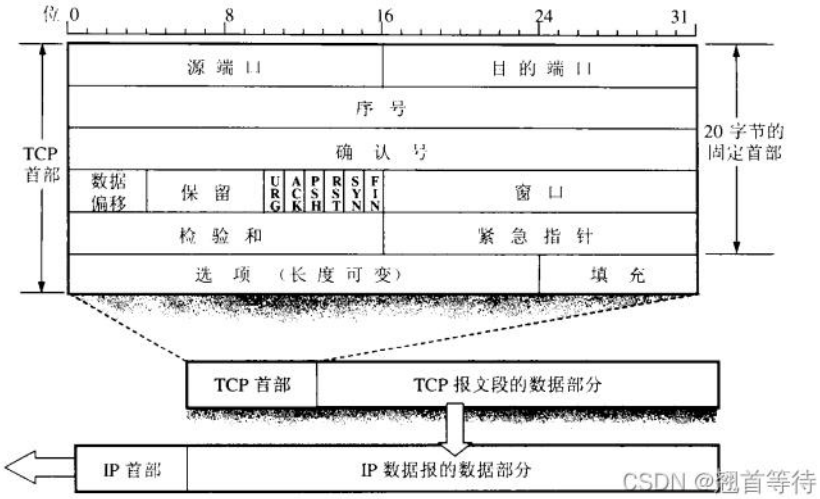
3.5 有连接传输层协议 TCP

TCP 服务特性

点-点；可靠保序；字节流；管道（在未加确认情况下一次传多个未经确认的段）；缓冲；全双工；面向连接；流控制；

TCP 段结构

各个字段的作用



源端口和目的端口：各占 2 个字节，分别写入源端口号和目的端口号。

序号：占 4 字节。本报文段所发送的数据的第一个字节的序号。序号范围是 [0, 2^32 -1]，共 2^32 个序号，序号增加到 2^32 -1 后，下一个序号就又回到 0。TCP 是面向字节流的。在一个 TCP 连接中传送的字节流中的每一个字节都按顺序编号。整个要传送的字节流的起始序号必须在连接建立时设置。首部中的序号字段值则指的是本报文段所发送的数据的第一个字节的序号。

确认号：占 4 字节，是期望收到对方下一个报文段的第一个数据字节的序号。只有在 ACK = 1 时确认号字段才有意义。

数据偏移：占 4 位，它指出 TCP 报文段的数据起始处距离 TCP 报文段的起始处有多远。这个字段实际上是指出 TCP 报文段的首部字段长度。由于首部中还有长度不确定的选项字段，因此数据偏移字段是必要的。

保留：占 6 位，保留位今后使用，但目前应置为 0 。

6 个控制位：包括 URG、ACK、PSH、RST、SYN、FIN，用来说明本报文段的性质。

同步 SYN

在连接建立时用来同步序号。当 SYN = 1 而 ACK = 0 时，表明这是一个连接请求报文段。对方若同意建立连接，则应在响应的报文段中是 SYN = 1 和 ACK = 1。因此 SYN 置为 1 就表示这是一个连接请求或连接接受报文。

确认 ACK

仅当 ACK = 1 时确认号字段才有效。当 ACK = 0 时，确认号无效。TCP 规定，在连接建立后所有传送的报文段都必须把 ACK 置 1。

推送 PSH

当两个应用进程进行交互式的通信时，有时在一端的应用进程希望在键入一个命令后立即就能够收到对方的响应。在这种情况下，TCP 就可以使用推送操作。这时，发送方 TCP 把 PSH 置 1，并立即创建一个报文段发送出去。接收方 TCP 收到 PSH = 1 的报文段，就尽快地（即 “推送” 向前）交付接收应用进程，而不再等到整个缓存都填满了后再向上交付。

紧急 URG

当 URG = 1 时，表明紧急指针字段有效。它告诉系统此报文段中有紧急数据，应尽快传送（相当于高优先级的数据），而不要按原来的排队顺序来传送。

当 URG 置 1 时，发送应用进程就告诉发送方的 TCP 有紧急数据要传送。于是发送方 TCP 就把紧急数据插入到本报文段数据的最前面，而在紧急数据后面仍是普通数据。这时要与首部中紧急指针字段配合使用。

紧急指针指出本报文字段中的紧急数据的字节数（紧急数据之后就是普通数据）。因此，紧急指针指出了紧急数据的末尾在报文段中的位置。

终止 FIN

用来释放一个连接。当 FIN = 1 时，表明此报文段的发送方的数据已发送完毕，并要求释放运输连接。

复位 RST

当 RST = 1 时，表明 TCP 连接中出现严重差错（如由于主机奔溃或其他原因），必须释放连接，然后再重新建立运输连接。RST 置 1 还用来拒绝一个非法的报文段或拒绝打开一个连接。RST 也可称为 重建位 或 重置位。

窗口：占 2 字节。窗口指的是发送本报文段的一方的接收窗口。窗口值告诉对方：从本报文段首部中的确认号算起，接收方目前允许对方发送的数据量（以字节为单位）。

检验和：占 2 字节。检验和字段检验的范围包括首部和数据这两部分。和 UDP 数据报一样，在计算检验和时，要在 TCP 报文段的前面加上 12 字节的伪首部。伪首部的格式与 UDP 用户数据报的伪首部一样。但应把伪首部第 4 个字段中的 17 改为 6（TCP 的协议号是 6），把第 5 个字段中的 UDP 长度改为 TCP 长度。接收方收到此报文段后，仍要加上这个伪首部来计算校验和。若使用 IPv6，则相应的伪首部也要改变。

紧急指针：占 2 字节。紧急指针仅在 URG = 1 时才有意义，它指出本报文字段中的紧急数据的字节数（紧急数据之后就是普通数据）。因此，紧急指针指出了紧急数据的末尾在报文段中的位置。当所有紧急数据都处理完时，TCP 就告诉应用程序恢复到正常操作。值得注意的是，即使窗口为零时也可发送紧急数据。

选项：长度可变，最长可达 40 字节。当没有使用 “选项” 时，TCP 的首部长度是 20 字节。

选项

最大报文段长度 MSS；窗口扩大选项；时间戳选项；选择确认（SACK）

连接建立时协商好双方的起始序号；

序号是首字节在字节流的偏移量；

确认：是对顺序收到的最后一个字节+1

RTT 时间估计和重发超时时间估计

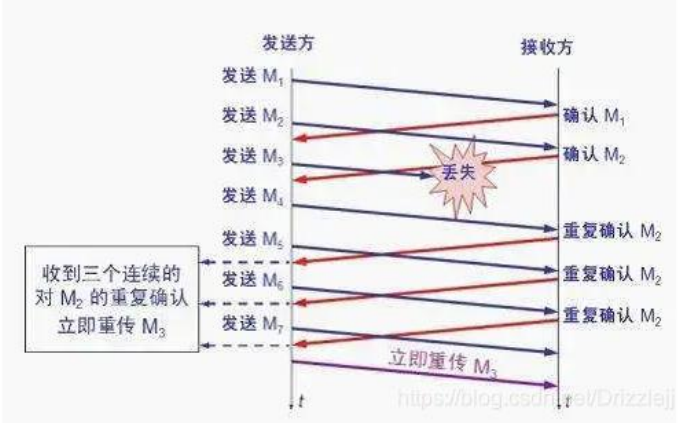
移动平均计算：平均往返延迟

移动平均计算：当前往返延迟采样值 与 平均值的 偏差

重发超时时间=平均值+4*偏差

TCP 的可靠数据传输原理

快速重传：在没有超时情况下，收到对方对于某一个段的重复三次（一共 4 个）ACK



流量控制

流控目的：防止淹没接收方

流控手段：将接收窗口大小捎带方式传递给发送端

TCP 连接管理 *理解*

连接建立：3 次握手技术 对双方选择的初始序号给予确认，准备好缓冲区

第一次握手：SYN=1，ACK=0；发起端的序号

第二次：SYN=1，ACK=1；被呼叫方的序号

第三次：（SYN=0）ACK=1

连接拆除：对称，存在 2 军问题不完美（也不存在完美释放连接的方案，用定时器凑活解决）

连接状态及其变迁

3.6 拥塞控制原理

拥塞的概念，什么是拥塞，为什么会发生拥塞

拥塞是指到达通信子网中某一部分的分组数量过多，使得该部分网络来不及处理，以致引起这部分乃至整个网络性能下降的现象，严重时甚至会导致网络通信业务陷入停顿。拥塞表现为分组丢失 (路由器缓冲区溢出)或分组经历比较长的延迟(在路由器的队列中排队)。

原因：（1）多条流入线路有分组到达，并需要同一输出线路，此时，如果路由器没有足够的内存来存放所有这些分组，那么有的分组就会丢失。（2）路由器的慢带处理器的缘故，以至于难以完成必要的处理工作，如缓冲区排队、更新路由表等。

拥塞控制目的

拥塞控制手段 *理解*

端到端的拥塞控制：TCP 采用这种方式

网络辅助的拥塞控制：ATM 网络标志和携带拥塞信息，反馈给主机（不要求）

3.7 TCP 拥塞控制原理

不依赖与网络中心提供的信息和服务，只靠延迟和丢失事件来推断网络中是否存在拥塞问题，发生在网络的边缘而非网络中心。

TCP 拥塞控制原理 *掌握*

检测拥塞：超时（拥塞，存在误判的可能性，但是概率比较低）、三个冗余 ACK（轻微拥塞）

拥塞控制机制：AIMD 慢启动 超时之后的保守策略

AIMD 为 TCP/IP 模型中，属于运输层，为了解决拥塞控制的一个方法，即：加性增，乘性减，或者叫做“和式增加，积式减少”。当 TCP 发送方感受到端到端路径无拥塞时就线性的增加其发送速度，当察觉到路径拥塞时就乘性减小其发送速度。

拥塞窗口（cwnd）:congestion window，当前端在一个 RTT 内能发送的窗口大小

ssthresh：slow start thresh，慢启动门限值

TCP 拥塞控制的 2 种算法：

Tahoe：超时事件和 3 个冗余 ACK 处理一样的 不要求

reno 算法（超时事件发生和 3 个冗余 ACK 处理不一样 *掌握*）

如果发生了超时（重度拥塞），那么下一个 RTT 将会重新进入慢启动，并且到达警戒值后采取 MSS 线性增长的方式增加发送速率

如果发生了 3 次冗余 ACK（轻度拥塞），那么下一个 RTT 将会直接从发生拥塞的 MSS 的一般速率开始线性增长，省去了一个慢启动阶段

平均延迟和超时定时器时间的设置 *不要求*

JACSON 算法（具体初始化和迭代算法 不要求）：

平均往返延迟公式

Dev 算法：第一个 超时时间=延迟的 1/2，初始设置；后面按公式

超时时间设置：es+4*dev

TCP 公平性： *不要求*

TCP 的吞吐量计算：不要求

第四章 网络层之数据平面

4.1 简介

网络层的主要服务和功能

服务：向传输层提供主机到主机的段传输服务

功能 1--转发，数据平面功能：从路由器的一个端口流入，从另外一个端口流出

功能 2--路由，控制平面功能：决定从源到目的地的路径

两个功能相互配合将数据报从源传送到目标主机；关联是转发表、流表

实现网络层功能的两种方式

传统方式

控制平面和数据平面功能垂直集成在每个设备上（路由器）；

控制平面功能：路由协议实体分布式地计算路由表；

数据平面功能：IP 协议按照路由表进行分组的转发；

SDN 通用转发方式

控制平面和数据平面分离，在不同设备上实现

sdn 控制器集中式计算、下发流表实现控制平面功能

sdn 分组交换机按照流表对到来的分组进行转发，实现数据平面的功能

网络层提供服务的一些重要指标

带宽

延迟，延迟差

丢包与否，丢包率

4.2 路由结构和工作原理

路由器的 2 大功能 *理解*

路由协议：结果形成路由表（转发表）

转发分组：使用转发表转发分组，交换

构成

输入端口：线路终端实现物理层功能、链路协议实体实现链路层功能，网络层功能实现分布式分组转发；

最长前缀匹配

最长前缀匹配是指在 IP 协议中，被路由器用于在路由表中进行选择的一个算法。因为路由表中的每个表项都指定了一个网络，所以一个目的地址可能与多个表项匹配。最明确的一个表项——即子网掩码最长的一个——就叫做最长前缀匹配。之所以这样称呼它，是因为这个表项也是路由表中，与目的地址的高位匹配得最多的表项。

交换结构：基于内存的，基于 bus 的，基于 CROSSBAR 的

memory：在 CPU 直接控制下的交换，采用传统的计算机；分组被拷贝到系统内存，CPU 从分组的头部提取出目标地址，查找转发表，找到对应的输出端口，拷贝到输出端口。

bus：通过总线交换（bus），数据报通过共享总线，从输入端口转发到输出端口

crossbar：通过互联网络的交换（crossbar 等），同时并发转发多个分组，克服总线带宽限制

输出端口三个层面的功能

网络层可以实现分组的调度：FIFO，RR，WFQ

FIFO 不对报文进行分类。FIFO 按报文到达接口的先后顺序让报文进入队列，先进先出。

轮询 RR 调度采用轮询的方式，对多个队列进行调度。

加权公平队列 WFQ（Weighted Fair Queuing）调度是按队列权重来分配每个流应占有出口的带宽。

调度支持对多媒体分组等优先级分组的传输支持

路由处理器：控制各部分协调工作

4.3 互联网网络层协议

IP 网络提供的服务模型：尽力而为

包括含义：丢包、乱序、不可靠

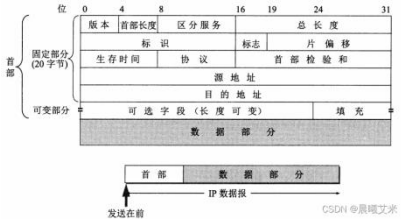
网络层构成

IP 协议　路由选择协议　ICMP 协议

ICMP 协议的全称是“Internet 控制消息协议”，主要用于在 IP 网络中发送控制消息，提供在通信环境中可能发生的各种问题的反馈。ICMP 采取“错误侦测与回馈机制”，通过 IP 数据包封装，用来发送错误和控制消息。

转发表

IP 数据报格式



各个字段的作用

1、版本占 4 位，指 IP 协议的版本。通信双方使用的 IP 协议的版本必须一致。目前广泛使用的 IP 协议版本号为 4（即 IPv4）。版本号为 6（即 IPv6）

2、首部长度占 4 位，可表示的最大十进制数值是 15。首部长度字段所表示的单位是 32 位（4 字节，与 TCP 首部中长度字端单位一致）。因为 IP 首部的固定长度是 20 字节，因此首部长度字段的最小值为 5（0101）。当首部长度为 15（1111）时，表示的长度为 60 字节；当 IP 分组的首部长度不是 4 的整数倍时，必须利用最后的填充字段加以填充达到 4 的整数倍。

3、区分服务占 1 字节，用来获得更好的服务。这个字段在旧标准中叫做服务类型，但实际上一直没有被使用过。只有在区分服务时，这个字段才起作用。在一般情况下都不使用这个字段。

4、总长度占 2 字节，指首部和数据之和的长度，单位为字节。能表示的最大长度为 65535 字节。在 IP 层下面的链路层协议规定了一个数据帧的数据字段的最大长度，这称为最大传输单元 MTU（maximum transfer unit）。当一个 IP 数据报封装成链路层的帧时，此数据报的总长度（即首部加上数据部分）一定不能超过下面的链路层所规定的的 MTU 值。

5、标识（identification）占 2 字节。网络层软件在存储器中维持一个计数器，每产生一个数据报，计数器就加 1，并将此值赋给标识字段。但这个“标识”并不同于 TCP 首部中的序号，因为 IP 是无连接的服务，数据报不存在按序接收的问题。当数据报长度超过网络的 MTU 而必须分片时，这个标识字段的值就被复制到所有被分片报文片的标识字段中。相同的标识字段的值使分片后的各数据报片最后能正确地重装层原来的数据。

6、标志占 3 位，目前只有两位有意义。标志字段中间的一位记为 DF（dongt fragment），意思是“不能分片”。当 DF=0 时才允许分片。标志字段最低位 MF（more fragment）。MF=1 即表示后面“还有分片”的数据报。MF=0 表示这已是若干数据报片中的最后一个。

7、片偏移占 13 位。片偏移指出：较长的 IP 报文在分片后，某片在原分组中的相对位置。也就是说，相对于用户数据字段的起点，该片从何处开始。片偏移以 8 个字节为偏移单位。没片的长度一定是 8 字节的整数倍。

分片和重组　*掌握*

一个分组的总体大小超过了转发链路的 MTU，因此要切片

到目标主机重组

IP 编址

IP 地址：主机或路由器和网络接口的标识

子网

在一个子网内的设备之间的通信有 2 个特点：1）通信无需借助路由器；2）子网前缀一样；

IP 地址分类：ABCDE

特殊 IP 地址

子网部分：全为 0---本网络

主机部分：全为 0---本主机

主机部分：全为 1--广播地址，这个网络的所有主机

子网掩码和 CIDR

CIDR（无类域间路由），子网部分可以在任意的位置，地址格式: a.b.c.d/x，其中 x 是地址中子网号的长度

NAT 不要求

DHCP 协议：上网主机获得 IP、掩码、默认网关和 loacl name server

DHCP（动态主机配置协议）是一个局域网的网络协议。指的是由服务器控制一段 IP 地址范围，客户机登录服务器时就可以自动获得服务器分配的 IP 地址和子网掩码。

路由聚集：连续的子网前缀的子网可达信息可以做聚集，减少向外部传输路由的数量，减少路由计算的负担。支持大概的路由聚集，与此对应的是最长前缀匹配的措施

IPv6

IPv6 格式（固定头部长度 40B），地址：128bits

IPv6 的变化

Checksum：被移除掉，降低在每一段中的处理速度；Options：允许，但是在头部之外，被 “Next Header” 字段标示

IPv4 到 IPv6 的迁移

隧道 *理解*

在 IPv4 路由器之间传输的 IPv4 数据报中携带 IPv6 数据报

4.4 通用转发和 SDN

SDN 方式控制平面和数据平面分离的优点 *理解*

集中在控制器上实现控制逻辑，网络可编程，可以实现各种复杂的网络功能、新功能（一次部署，持续升级）、方便管理

形成开发生态（控制器，分组交换机，网络应用，在一个开放的框架下协作）

SDN 分组交换机按照计算出的流表进行分组转发、通用、便于升级

SDN 交换机替代了传统的路由器和交换机。

分组交换机的工作原理 *理解*

模式匹配+行动（不仅仅是转发，还可以组播，泛洪，修改字段和阻塞等）

进来分组，按照各级字段匹配流表，按照相应的行动动作分组

按照优先权进行判断；之后，统计计数

第五章 网络层之控制平面

5.1 概述

两种方式实现控制平面功能：传统方式和 SDN 方式

传统方式：在每个路由器上分布式实现路由功能

SDN 方式：在 SDN 控制器上由网络应用集中式计算、生成流表

5.2 路由选择算法

路由目标：根据收集到的路由信息（拓扑，链路代价等）计算出源到目标较好的路径，

代价比较低的路径

主机-主机路径==路由器--路由器的路径；

路由目标实际上是计算出节点的汇集树；

路由原则：完整正确，简单，健壮，稳定公平，最优（次优）

路由分类

静态和动态（自适应）

静态：路由随时间变化缓慢；动态：路由变化很快

局部和全局的

全局：所有的路由器拥有完整的拓扑和边的代价的信息；分布式：路由器只知道与它有物理连接关系的邻居路由器，和到相应邻居路由器的代价值

LS 算法：全局的路由选择算法，工作原理 *掌握*

每个节点向网络中所有广播链路状态，即该节点和它所连接的链路的特征和费用，这样，网络中的每个节点都具有了该网络的等同的、完整的视图。接下来，每个节点各自计算都其他节点的路径。路由选择算法使用的是 dijkstra 算法

每个节点收集邻居信息，生成 LS；LS 全网泛洪

节点收集 LS 状态分组，形成网络拓扑

按照最短路径算法算出到其他节点的最优路径

DV 算法：局部的路由选择算法，工作原理 *掌握*

每个节点刚开始只有它到直接相连邻居的链路费用，每个节点等待来自邻居的更新，当收到一个更新时，重新计算新的邻居向量。

每个节点维护到所有其他节点的下一跳和代价值

邻居节点之间定期交换 DV

按照 Bellman-Ford 不断迭代生成到所有目标的代价和相应的下一跳

层次路由 *理解*

一个平面解决路由的问题：计算、传输和存储路由信息的量太大，不具备可扩展性，也不满足不同网络运营方不同的管理需求

分成 AS，AS 内部之间的节点路由有内部网关协议解决：AS 之间的路由，分层解决（路由到网关，由网关路由到目标网关，到了目标 AS 内部，采用 AS 内部的路由解决）

自治系统：autonomous system。在互联网中，一个自治系统(AS)是一个有权自主地决定在本系统中应采用各种路由协议的小型单位。这个网络单位可以是一个简单的网络也可以是一个由一个或多个普通的网络管理员来控制的网络群体，它是一个单独的可管理的网络单元（例如一所大学，一个企业或者一个公司个体）。

优势：分层路由，解决了规模性问题，管理性问题

5.3 互联网络的路由协议

路由协议分类

内部网关协议 IGP

IGP（内部网关协议）是在一个自治网络内网关（主机和路由器）间交换路由信息的协议。

RIP

路由信息协议 RIP 是基于距离矢量算法的路由协议，利用跳数来作为计量标准。

OSPF：AS 内部支持分层路由，同时支持多种代价

OSPF 协议是一种链路状态协议。每个路由器负责发现、维护与邻居的关系，并将已知的邻居列表和链路费用 LSU 报文描述，通过可靠的泛洪与自治系统 AS 内的其他路由器周期性交互，学习到整个自治系统的网络拓扑结构；并通过自治系统边界的路由器注入其他 AS 的路由信息，从而得到整个 Internet 的路由信息。

IGRP

内部网关路由协议，是一种内部网关协议，采用距离向量算法。以自治系统的方式提供路由选择路由协议。其算法与路由信息协议（RIP）类似，透过用户配置，如延迟、带宽、可靠性及负载量等于各路由器进行的路由管理。

外部网关协议 EGP

BGP

网关路由器参与 AS 内部路由计算，收集 AS 内部子网可达信息

网关路由器通过 AS 间路由向其他 AS 网关通告子网可达信息

网关路由器还转发“过手”子网可达信息，但是 AS 路径要加上它自己 AS 编号（防止形成环路）

网关路由器 i-BGP 向 AS 内部所有路由节点通告收集到的子网可达信息

内部路由器，通过 AS 内路由和 AS 间路由共同决定向 AS 外部子网的下一跳（内部网关协议决定如何往网关，外部网关协议决定通过那个子网可到达 AS 子网外部）

内部网关协议和外部网关协议的对比　了解

内部网关协议重视效率，性能

外部网关协议重视策略：经济策略和政治策略

5.4　SDN 控制平面

在控制器上集中实现控制功能

控制器和 SDN 交换机按照 OPENFLOW 等南向接口协议等下发流表，上报设备状态

SDN 控制器按照北向接口和网络应用打交道

南向接口：管理其他厂家网管或设备的接口，即向下提供的接口。

北向接口（Northbound Interface）是为厂家或运营商进行接入和管理网络的接口，即向上提供的接口。

5.5　ICMP 协议　了解

ICMP 互联网控制报文协议。它是 TCP/IP 协议簇的一个子协议，用于在 IP 主机、路由器之间传递控制消息。控制消息是指网络通不通、主机是否可达、路由是否可用等网络本身的消息。这些控制消息虽然并不传输用户数据，但是对于用户数据的传递起着重要的作用。

作用：包括错误，echo 请求和应答

报文类型

第六章　数据链路层与局域网

6.1　引论

链路层提供的服务

成帧、链路存取控制（链路访问控制）

在相邻节点间进行可靠数据传递

流量控制

检错

纠错

全双工和半双工服务

链路层网络节点的连接方式

点到点方式：比较适合广域

多点连接方式：比较适合局域、联网方便，但是需要解决 MAC 问题

6.2　检错与纠错

检错原理

奇偶校验

所谓奇偶校验就是在发送的每一个字节后都加上一位，使得每个字节中 1 的个数为奇数个或偶数个。比如我们要发送的字节是 0x1a，

二进制表示为 0001 1010。

采用奇校验，则在数据后补上个 0，数据变为 0001 1010 0，数据中 1 的个数为奇数个（3 个）

采用偶校验，则在数据后补上个 1，数据变为 0001 1010 1，数据中 1 的个数为偶数个（4 个）

接收方通过计算数据中 1 个数是否满足奇偶性来确定数据是否有错。

CRC　*掌握*

原理

CRC 算法的基本思想是将传输的数据当做一个位数很长的数。将这个数除以另一个数。得到的余数作为校验数据附加到原数据后面。

生成多项式

冗余位计算方法以及验证方法

6.3　多路访问协议

多路访问协议大致分为三种：信道划分协议（把信道划分成小片（时间、频率、编码），分配片给每个节点专用）、随机接入协议（信道不划分，允许冲突，冲突后恢复）和轮流协议（节点依次轮流，但是有很多数据传输的节点可以获得较长的信道使用权）。理想的多路访问协议应具有以下两种特性：1. 只有一个结点活跃时，该活跃节点具有 R 的吞吐量。2. 当有 M 个结点活跃时，每个活跃节点吞吐量接近 R/M。

MAC 的必要性

局域网的数据链路层分为逻辑链路层 LLC 和介质访问控制 MAC 两个子层。

逻辑链路控制（简称 LLC）是局域网中数据链路层的上层部分，IEEE 802.2 中定义了逻辑链路控制协议。用户的数据链路服务通过 LLC 子层为网络层提供统一的接口。在 LLC 子层下面是 MAC 子层。

MAC 介质访问控制属于 LLC 下的一个子层。 是解决当局域网中共用信道的使用产生竞争时，如何分配信道的使用权问题。

MAP

信道划分

TDMA

FDMA

CDMA：删掉

码分多址是指利用码序列相关性实现的多址通信。码分多址的基本思想是靠不同的地址码来区分的地址。

RAP：随机访问协议

slotted ALOHA

时隙 ALONA，当节点获取新的帧，在下一个时隙传输。传输时没有检测到冲突，成功，节点能够在下一时隙发送新帧；检测时如果检测到冲突，失败，节点在每一个随后的时隙以概率 p 重传帧直到成功。

ALOHA

纯 ALOHA，当有帧需要传输，则马上传输。

冲突的概率增加：帧在 t0 发送，和其它在[t0-1, t0+1]区间内开始发送的帧冲突，和当前帧冲突的区间（其他帧在此区间开始传输）增大了一倍

CSMA，CSMA/CD（至少 2t 长度帧），CSMA/CA

CSMA（载波侦听多路访问）：在传输前先侦听信道

CSMA/CD（CD 为冲突检测，以太网）：没有传完一个帧就可以在短时间内检测到冲突；冲突发生时则传输终止，减少对信道的浪费

CSMA/CA（无线局域网，无 CD）：如果站点侦测到信道空闲持续 DIFS 长，则传输整个帧 (no CD)；如果侦测到信道忙碌，那么选择一个随机回退值，并在信道空闲时递减该值；如果信道忙碌，回退值不会变化；到数到 0 时（只生在信道闲时）发送整个帧。如果没有收到 ACK，增加回退值，重新侦测

轮转协议：不要求

令牌协议

6.4 链路层编地址

MAC 地址

格式

分配

MAC 地址和网络层 IP 地址的区别

1、MAC 地址应用在数据链路层。数据链路层协议可以使数据从一个节点传递到相同链路的另一个节点上（通过 MAC 地址）。

2、IP 地址应用在网络层。网络层协议使数据可以从一个网络传递到另一个网络上（ARP 根据目的 IP 地址，找到中间节点的 MAC 地址，通过中间节点传送，从而最终到达目的网络）。

1、MAC 地址的分配是基于制造商。

MAC 地址由网络设备制造商生产时写在硬件内部。这个地址与网络无关，也即无论将带有这个地址的硬件（如集线器、网卡、路由器等）接入到网络的何处，它都有相同的 MAC 地址，是不可变的。

2、IP 地址的分配是基于网络拓朴。

mac 地址 48bit，ip 地址 32bit

IP 地址由网络地址和主机地址两部分组成，分配给这两部分的位数随地址类（A 类、B 类、C 类等）的不同而不同。

层次不同

MAC 地址平面的，用于标示一个物理网络的不同站点；IP 是可以聚集的，便于计算路由；

ARP 协议

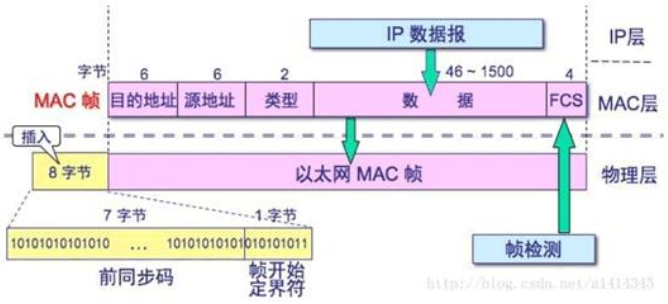
目的：物理网络范围内 IP 地址到 MAC 地址的转换

工作原理：广播查询，单播应答

6.5 以太网络

IEEE802.3 标准，链路层和相应的物理层

以太网络的帧结构



向上提供服务的特点

无连接

不可靠（接收方适配器不发送 ACKs 或 NAKs 给发送方）

访问控制技术

CSMA/CD（jam+指数后退） 掌握

指数后退：在第 m 次失败后，适配器随机选择一个{0，1，2， ，2^m-1}中 K，等待 K*512 位时，然后重新 csma

CSMA/CA 理解

编码

Manchester 编码

6.6 HUB 和交换机

HUB（集线器）连接方式的问题：无法隔离冲突，在一个冲突域之中

集线器会把接收到的数据包每次都广播给局域网局域网的所有计算机，集线器的数据传输方式是广播方式，而交换机的数据传输是有目的的，数据只对目的节点发送，只是在自己的 MAC 地址表中找不到的情况下第一次使用广播方式发送

交换机的工作原理

选择性转发

自学习

流量隔离

专用接入

路由器和交换机的区别　理解

交换机是用来连接局域网的，路由器是用来连接互联网的（也可以连接多个局域网）；路由器：寻址，转发（依靠 IP 地址），交换机：过滤，转发（依靠 MAC 地址）。交换机用于连接局域网，数据包在局域网内网的数据转发，路由器用于连接局域网和外网，数据包可以在不同局域网转发。