

# 万维网 和高级专题

理论课程

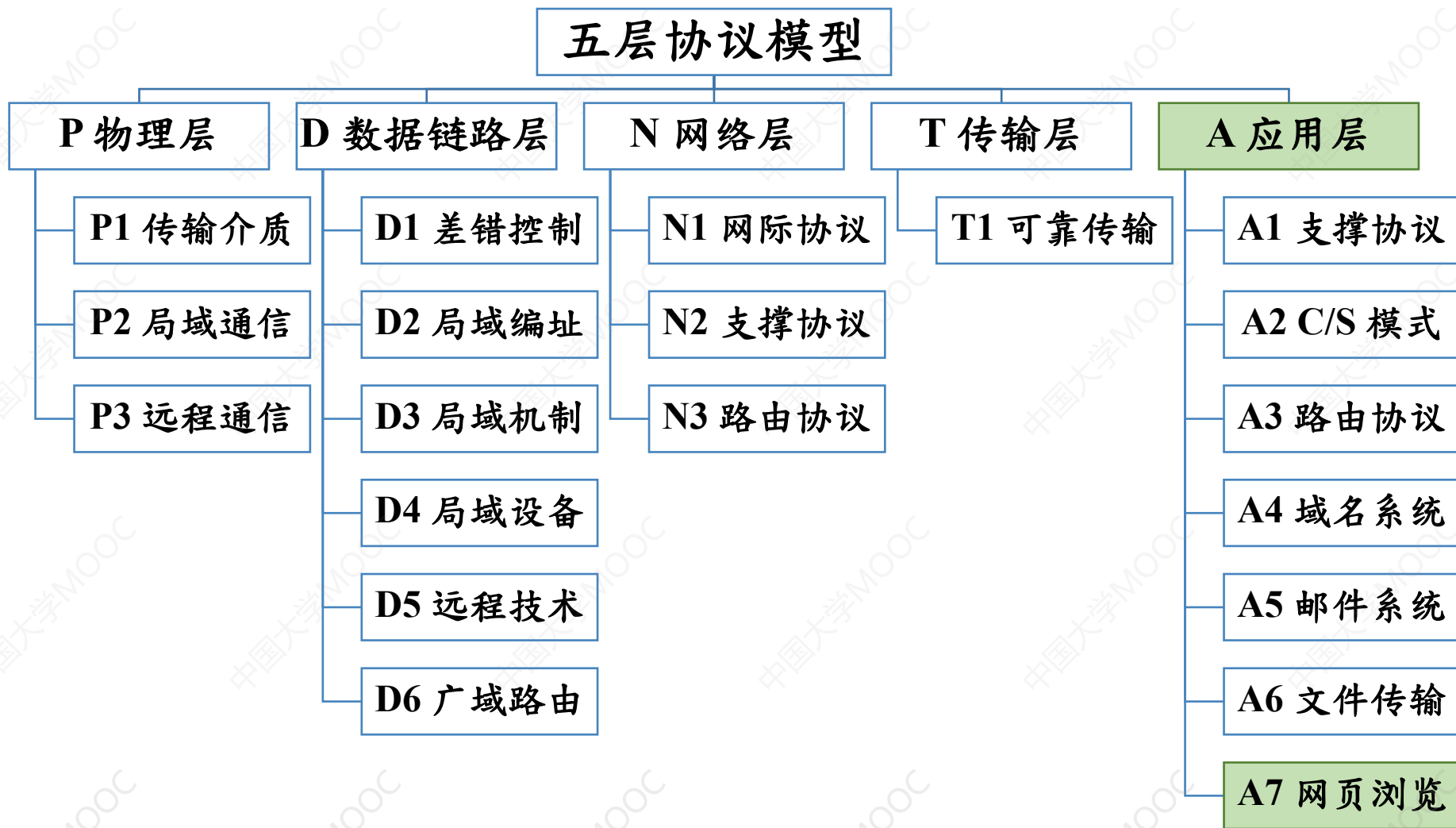


廈門大學  
XIAMEN UNIVERSITY



信息学院 黄 烽  
(特色化示范性软件学院) 博士, 副教授  
School of Informatics Wei Huang

# 知识框架



# 主要内容

- HTTP

- 工作原理与过程
- 错误代码
- URL
- HTML文档

# 对应课本章节

- **PART I Introduction And Internet Applications**

- **Chapter 4 Traditional Internet Applications**

- **4.3~4.9 Representation And Transfer; Web Protocols; Document Representation With HTML; Uniform Resource Locators And Hyperlinks; Web Document Transfer With HTTP; Caching In Browsers; Browser Architecture**

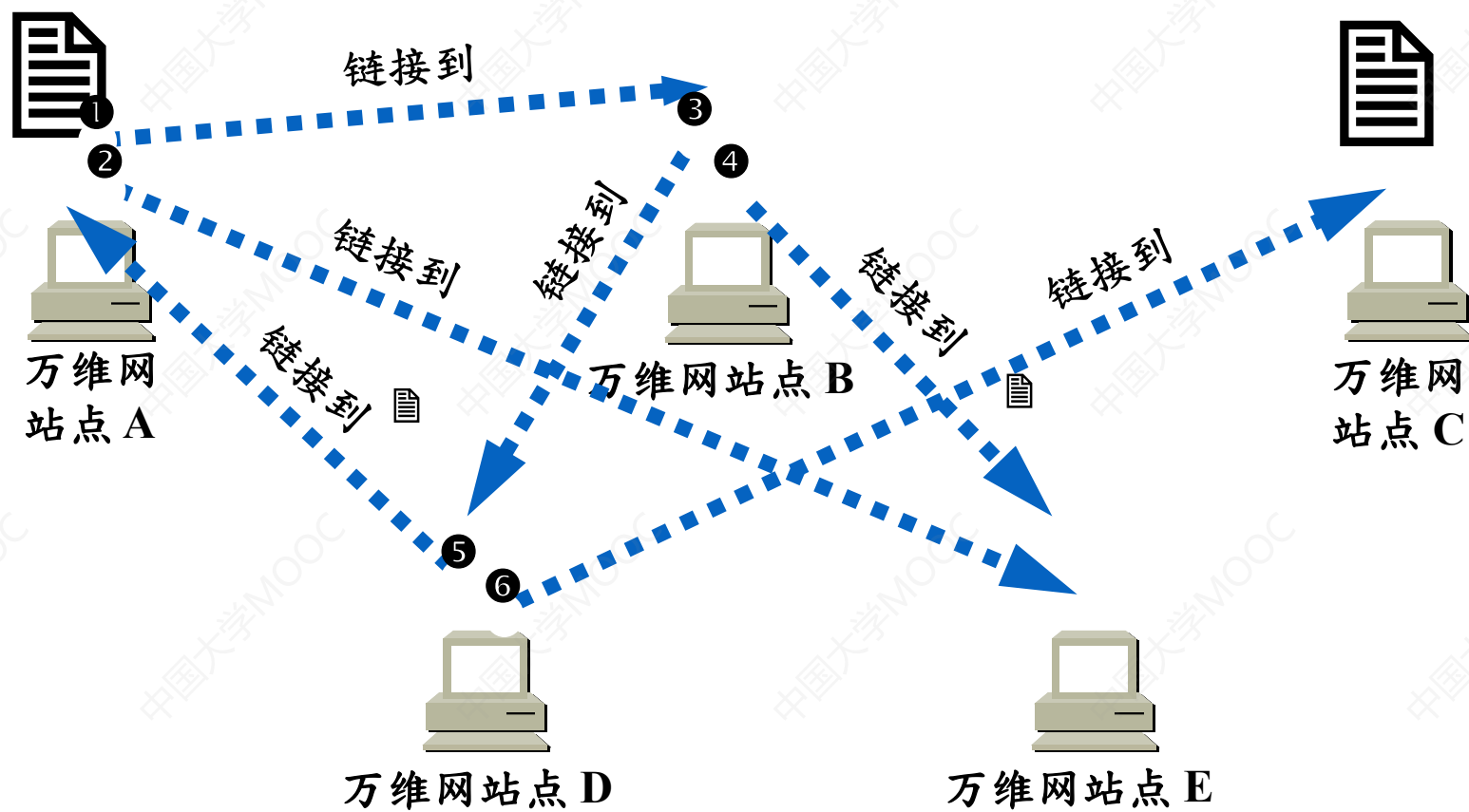
# 内容纲要

1	万维网
2	HTTP协议
3	网络结构设计
4	网络技术展望
5	网络技术和应用趋势

# 万维网概述

- 万维网 WWW (World Wide Web)并非某种特殊的计算机网络。
- 万维网是一个大规模的、联机式的信息储藏所。
- 万维网用链接的方法能非常方便地从因特网上的一个站点访问另一个站点，从而主动地按需获取丰富的信息。
- 这种访问方式称为“链接”。

# 万维网提供分布式服务



# 超媒体与超文本

- 万维网是分布式超媒体(hypermedia)系统，它是超文本(hypertext)系统的扩充。
  - 一个超文本由多个信息源链接成。利用一个链接可使用户找到另一个文档。这些文档可以位于世界上任何一个接在因特网上的超文本系统中。超文本是万维网的基础。
  - 超媒体与超文本的区别是文档内容不同。超文本文档仅包含文本信息，而超媒体文档还包含其他表示方式的信息，如图形、图像、声音、动画，甚至活动视频图像。



# 万维网的工作方式

- 万维网以客户-服务器方式工作。
- 浏览器就是在用户计算机上的万维网客户程序。万维网文档所驻留的计算机则运行服务器程序，因此这个计算机也称为万维网服务器。
- 客户程序向服务器程序发出请求，服务器程序向客户程序送回客户所要的万维网文档。
- 在一个客户程序主窗口上显示出的万维网文档称为页面(page)。

# 万维网必须解决的问题

- 怎样标识分布在整个因特网上的万维网文档？
  - 使用统一资源定位符 URL (Uniform Resource Locator) 来标识万维网上的各种文档。
  - 使每个文档在整个因特网的范围内具有唯一的标识符 URL。
- 用何协议实现万维网上各种超链的链接？
  - 在万维网客户程序与万维网服务器程序之间进行交互所使用的协议，是超文本传送协议 HTTP。
  - HTTP 是一个应用层协议，使用 TCP 连接进行可靠的传送。

# 万维网必须解决的问题

- 怎样使各种文档在各种计算机显示并找到超链位置？
  - 超文本标记语言 HTML (Hyper-Text Markup Language) 使得万维网页面的设计者可以很方便地用一个超链从本页面的某处链接到因特网上的任何一个万维网页面，并且能够在自己的计算机屏幕上将这些页面显示出来。
- 怎样使用户能够很方便地找到所需的信息？
  - 为了在万维网上方便地查找信息，用户可使用各种的搜索工具（即搜索引擎）。

# 统一资源定位符 URL

- URL的格式 `<协议>://<主机>:<端口>/<路径>`
  - 对可从因特网上得到的资源的位置和访问方法的简洁表示。
  - URL 给资源的位置提供一种抽象的识别方法，并用这种方法给资源定位。
  - 只要能够对资源定位，系统就可以对资源进行各种操作，如存取、更新、替换和查找其属性。
  - URL 相当于一个文件名在网络范围的扩展。因此 URL 是与因特网相连的机器上的任何可访问对象的一个指针。

# 内容纲要

1	万维网
2	HTTP协议
3	网络结构设计
4	网络技术展望
5	网络技术和应用趋势

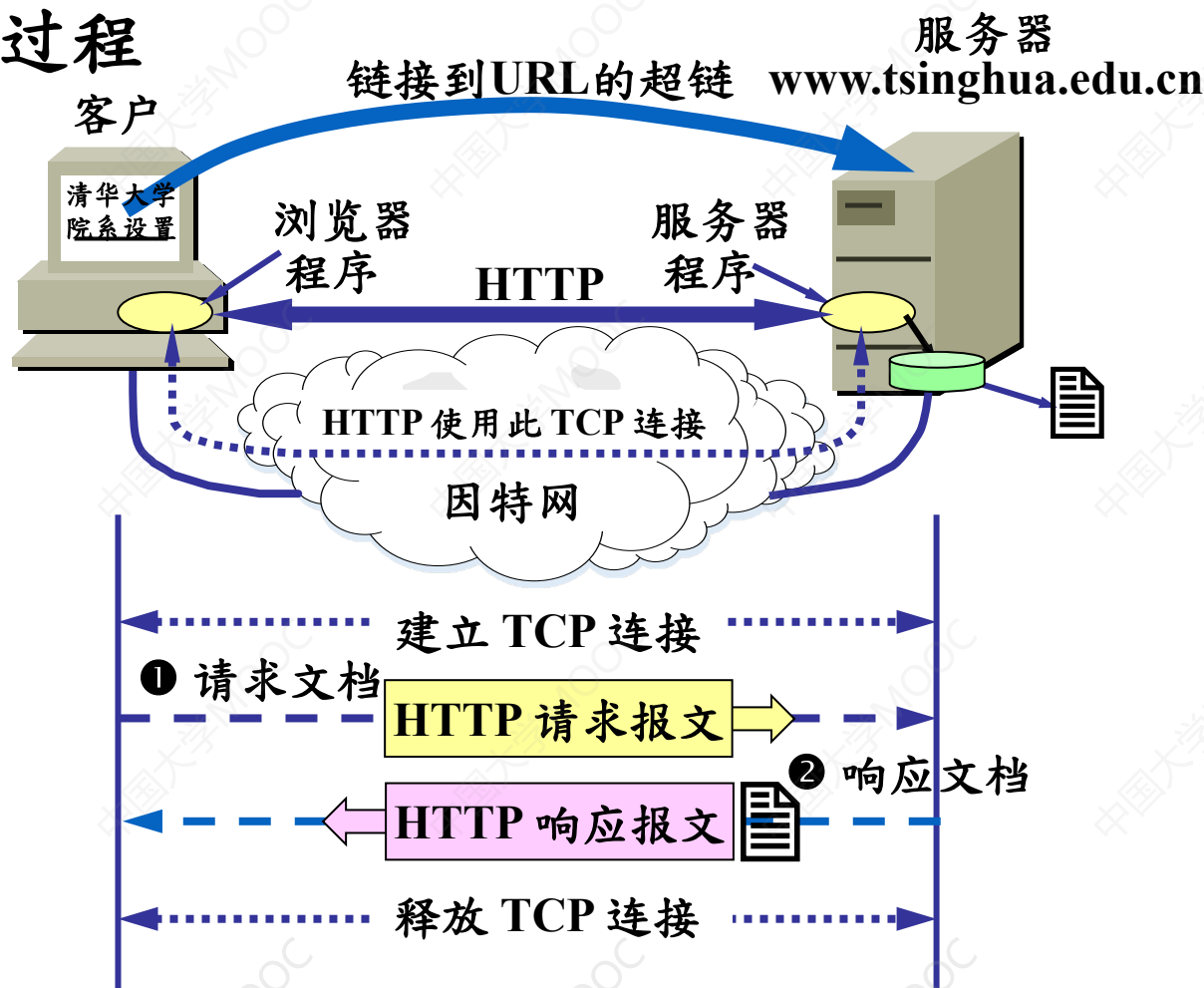
# 超文本传送协议 HTTP

- HTTP 的操作过程

- 为了使超文本的链接能够高效率地完成，需要用 HTTP 协议来传送一切必须的信息。
- 从层次的角度看，HTTP 是面向事务的(transaction-oriented)应用层协议，它是万维网上能够可靠地交换文件（包括文本、声音、图像等各种多媒体文件）的重要基础。

# 超文本传送协议 HTTP

## • HTTP 的操作过程



# 用户点击鼠标后所发生的事件

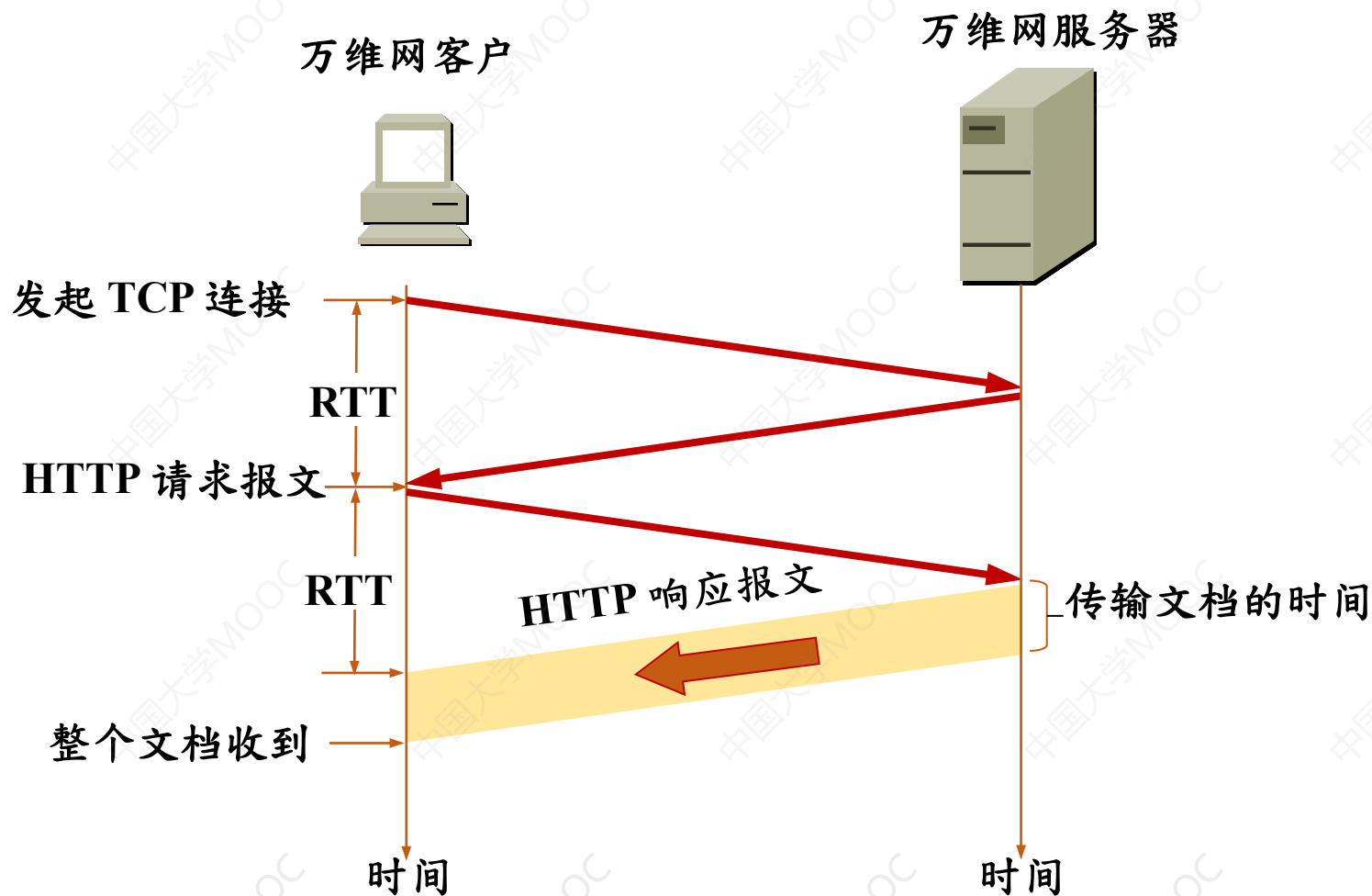
- 浏览器分析超链指向页面的 URL 。
- 浏览器向DNS请求解析 `www.tsinghua.edu.cn` 的 IP 地址 。
- 域名系统DNS解析出清华大学服务器的 IP 地址 。
- 浏览器与服务器建立 TCP 连接
- 浏览器发出取文件命令：`GET /chn/yxsx/index.htm` 。
- 服务器给出响应，把文件 `index.htm` 发给浏览器 。
- TCP 连接释放 。
- 浏览器显示院系设置文件 `index.htm` 中的所有文本 。



# HTTP 的主要特点

- HTTP 是面向事务的客户服务器协议。
- HTTP 1.0 协议是无状态的（stateless）。
  - 无状态的服务器不会保留任何两个请求之间的信息。
- HTTP 协议本身也是无连接的，虽然它使用了面向连接的 TCP 向上提供的服务。

# 请求一个万维网文档所需的时间



# 持续连接 (persistent connection)

- HTTP/1.1 协议使用持续连接。
- 万维网服务器在发送响应后仍然在一段时间内保持这条连接，使同一个客户（浏览器）和该服务器可以继续在这条连接上传送后续 HTTP 请求报文和响应报文。
- 这并不局限于传送同一个页面上链接的文档，而是只要这些文档都在同一个服务器上就行。
- 目前一些流行的浏览器（例如，IE 6.0）的默认设置就是使用 HTTP/1.1。

# 持续连接的两种工作方式

- 非流水线方式

- 客户在收到前一个响应后才能发出下一个请求。这比非持续连接的两倍 RTT 的开销节省了一个 RTT 时间。但服务器在发送完一个对象后，其 TCP 连接就处于空闲状态，浪费了服务器资源。

- 流水线方式

- 客户在收到 HTTP 的响应报文之前就能够接着发送新的请求报文。一个接一个的请求报文到达服务器后，服务器就可连续发回响应报文。使用流水线方式时，客户访问所有的对象只需花费一个 RTT 时间，使 TCP 连接中的空闲时间减少，提高了下载文档效率。

# 代理服务器 ( proxy server )

- 代理服务器 ( proxy server ) 又称为万维网高速缓存 ( Web cache ) ，它代表浏览器发出 HTTP 请求。
- 万维网高速缓存把最近的一些请求和响应暂存在本地磁盘中。
- 当与暂时存放的请求相同的新请求到达时，万维网高速缓存就把暂存的响应发送出去，而不需要按 URL 的地址再去因特网访问该资源。

# 用高速缓存减少访问服务器的时延

- (1) 浏览器访问因特网的服务器时，要先与校园网的高速缓存建立 TCP 连接，并向高速缓存发出 HTTP 请求报文
- (2) 若高速缓存已经存放了所请求的对象，则将此对象放入 HTTP 响应报文中返回给浏览器。
- (3) 否则，高速缓存就代表发出请求的用户浏览器，与因特网上的源点服务器建立 TCP 连接，并发送 HTTP 请求报文

# 用高速缓存减少访问服务器的时延

- (4) 源点服务器将所请求的对象放在 HTTP 响应报文中返回给校园网的高速缓存。
- (5) 高速缓存收到此对象后，先复制在其本地存储器中（为今后使用），然后再将该对象放在 HTTP 响应报文中，通过已建立的 TCP 连接，返回给请求该对象的浏览器。

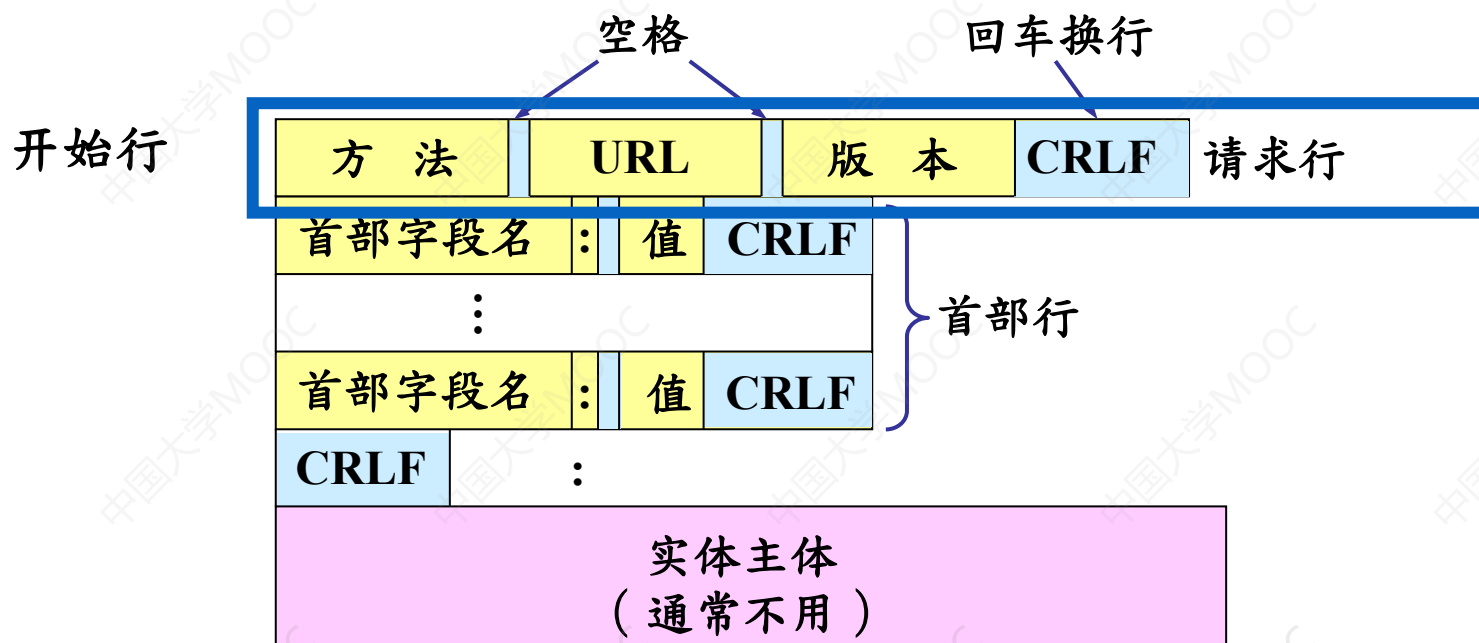
# HTTP 的报文结构

- HTTP 有两类报文：
  - 请求报文——从客户向服务器发送请求。
  - 响应报文——从服务器到客户的回答。
- 由于 HTTP 是面向文本的(text-oriented)，因此在报文中的每一个字段都是一些 ASCII 码串，因而每个字段的长度都是不确定的。



# HTTP 的报文结构

- 报文由三个部分组成，即开始行、首部行和实体主体。
- 在请求报文中，开始行就是请求行。



# HTTP 请求报文的一些方法

## • 方法（操作）

## 意义

- **OPTION** 请求一些选项的信息
- **GET** 请求读取由 URL 所标志的信息
- **HEAD** 请求读取由 URL 所标志的信息的首部
- **POST** 给服务器添加信息（例如，注释）
- **PUT** 在指明的 URL 下存储一个文档
- **DELETE** 删除指明的 URL 所标志的资源
- **TRACE** 用来进行环回测试的请求报文
- **CONNECT** 用于代理服务器

# 实例

- 运行于www.google.com，端口80

- 客户端请求

GET / HTTP/1.1

Host:www.google.com

(紧跟着一个换行，通过敲入回车实现)

- 服务器应答

HTTP/1.1 200 OK

Content-Length: 3059

Server: GWS/2.0

Date: Sat, 11 Jan 2003 02:44:04 GMT

Content-Type: text/html

Cache-control: private

Set-Cookie: PREF=ID=73d4aef52e57bae9:TM=1042253044:LM=1042253044:S=SMCc\_HRPCQiqyX9j; expires=Sun, 17-Jan-2038 19:14:07 GMT; path=/; domain=.google.com

Connection: keep-alive

(紧跟一个空行，并由HTML文本组成Google的主页)

# HTTP状态码

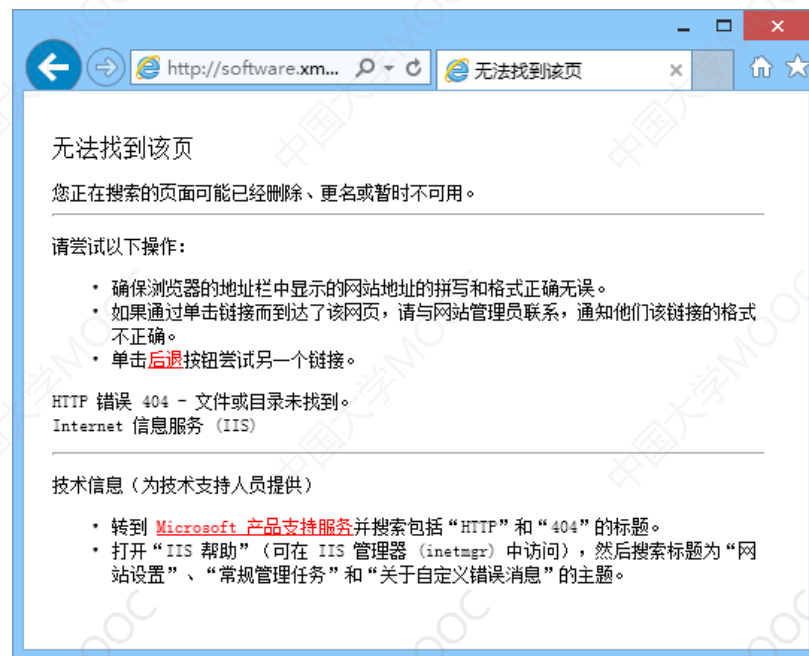
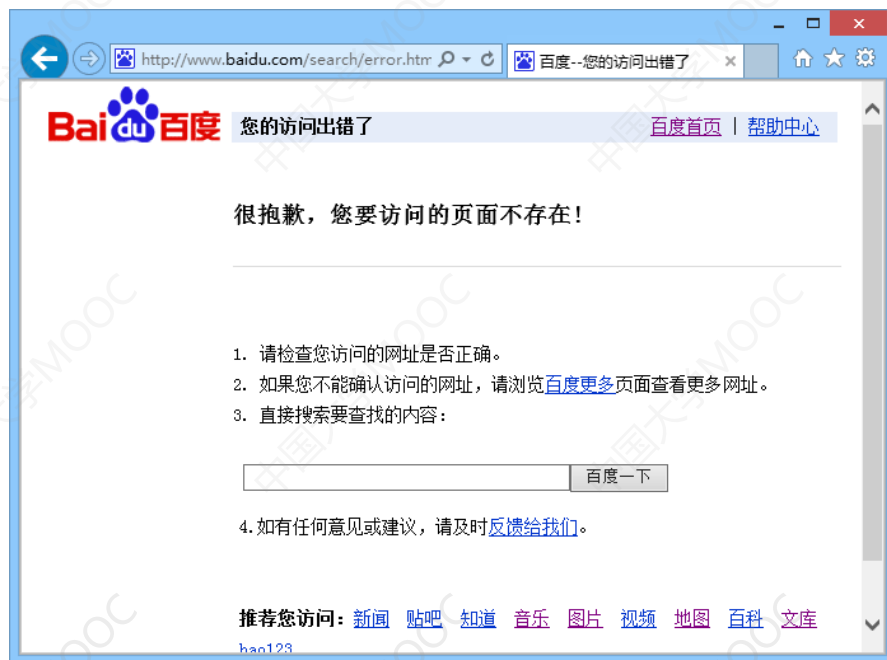
- 1xx 表示通知信息的，如请求收到了或正在进行处理。
- 2xx 表示成功，如接受或知道了。
- 3xx 表示重定向，表示要完成请求还必须采取进一步的行动。
- 4xx 表示客户的差错，如请求中有错误的语法或不能完成。
- 5xx 表示服务器的差错，如服务器失效无法完成请求。

# 状态码

## • 同样的Page Not Found，不一样的代码

— 左图：服务器处理的重定向（HTTP 302）

— 右图：服务器美化的错误页面（HTTP 404）



# 在服务器上存放用户的信息

- 万维网站点使用 Cookie 来跟踪用户。
- Cookie 表示在 HTTP 服务器和客户之间传递的状态信息。
- 使用 Cookie 的网站服务器为用户产生一个唯一的识别码。利用此识别码，网站就能够跟踪该用户在该网站的活动。

# 超文本标记语言 HTML

- 超文本标记语言 HTML 中的 Markup 的意思就是“设置标记”。
- HTML 定义了许多用于排版的命令（即标签）。
- HTML 把各种标签嵌入到万维网的页面中。这样就构成了所谓的 HTML 文档。HTML 文档是一种可以用任何文本编辑器创建的 ASCII 码文件。
- 远程链接：超链的终点是其他网点上的页面。
- 本地链接：超链指向本计算机中的某个文件。

# HTML 程序

<HTML>

<HEAD>

<TITLE>

text that forms the document title

</TITLE>

</HEAD>

<BODY>

body of the document appears here

</BODY>

</HTML>



# 动态万维网文档

- 静态文档是指该文档创作完毕后就存放在万维网服务器中，在被用户浏览的过程中，内容不会改变。
- 动态文档是指文档的内容是在浏览器访问万维网服务器时才由应用程序动态创建。
- 动态文档和静态文档之间的主要差别体现在服务器一端。这主要是文档内容的生成方法不同。而从浏览器的角度看，这两种文档并没有区别。

# 万维网服务器功能的扩充

- 应增加另一个应用程序，用来处理浏览器发来的数据，并创建动态文档。
- 应增加一个机制，用来使万维网服务器把浏览器发来的数据传送给这个应用程序，然后万维网服务器能够解释这个应用程序的输出，并向浏览器返回 HTML 文档。

# 内容纲要

1	万维网
2	HTTP协议
3	网络结构设计
4	网络技术展望
5	网络技术和应用趋势

# 设计理念

## • PDIOO网络生命周期

— 计划-设计-实现-运行-优化

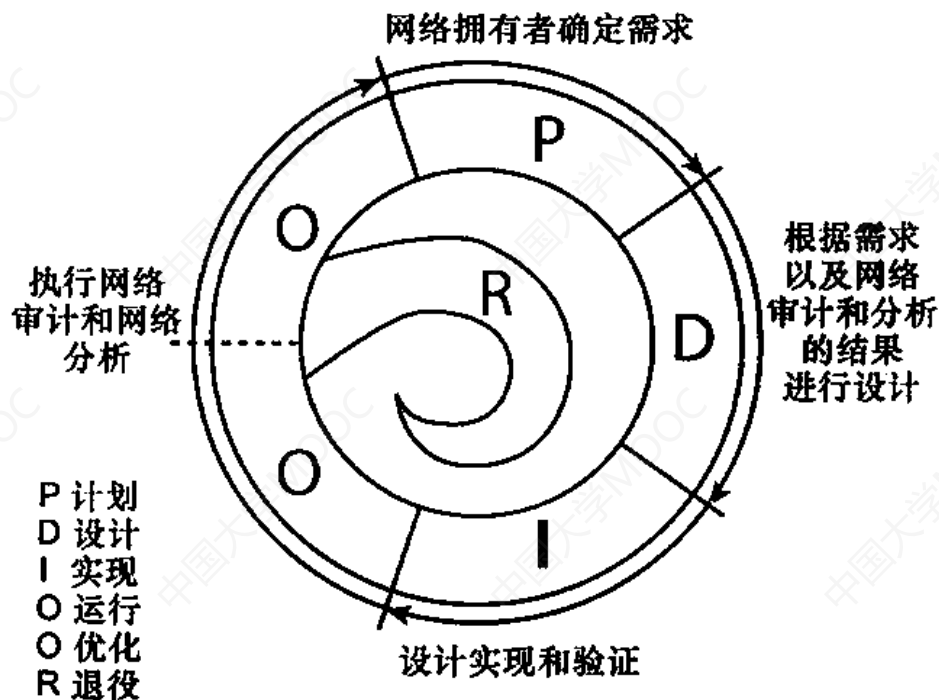


图 1-2 在 PDIOO 网络生命周期中包括设计工作的许多方面<sup>1</sup>

# PDIOO网络生命周期

- 计划阶段：确定详细的网络需求，并检查现有的网络。
- 设计阶段：根据初始的需求和对现有网络分析中所收集到的额外信息设计网络。
- 实现阶段：根据得到认可的设计方案构建网络。
- 运行阶段：网络开始运转，并受到监测，该阶段是对设计方案的最最终测试。
- 优化阶段：发现和纠正一些问题。
- 退役阶段：网络过时了或不用了。

# 网络设计任务

- 确定需求
  - 如果网络已经存在了，需对现有网络进行分析
- 准备概要设计
  - 完成最终的设计开发；部署网络
- 监测网络
  - 如有必要，重新设计网络
- 维护文档

# 网络设计任务

## • 网络设计任务

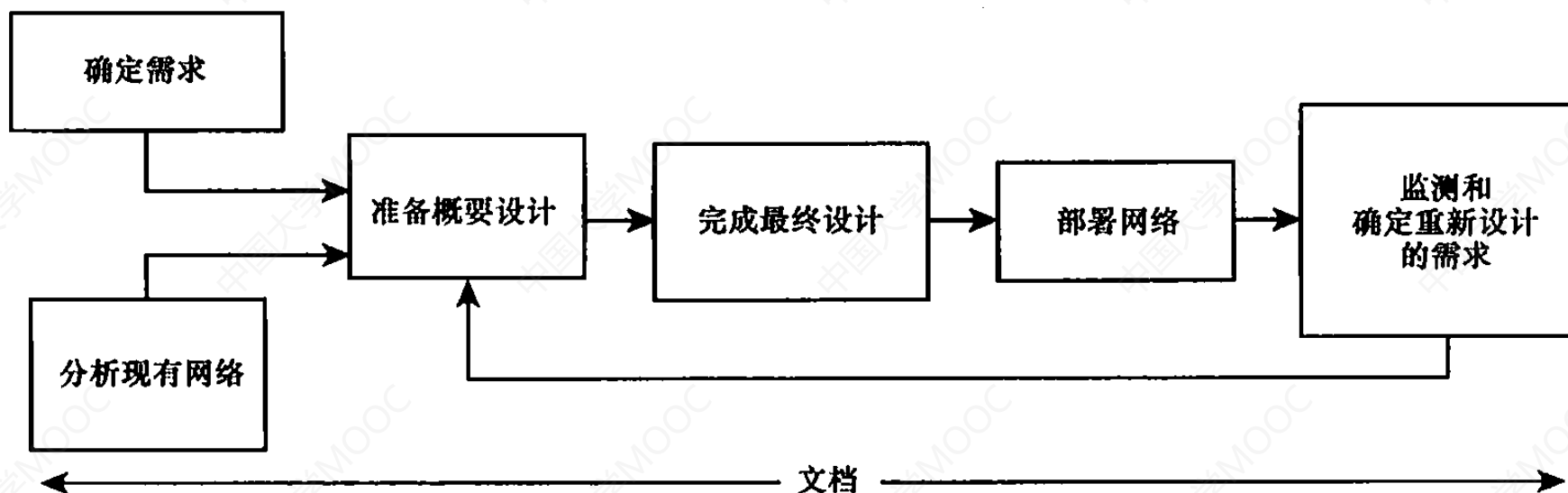


图 1-3 网络设计是一个持续性的过程

# 网络设计任务：确定需求

- 技术需求和限制因素：

- 网络上运行的应用程序；网络上运行的其他协议（如路由协议）
- 互联网连通性要求；
- IPV4地址的限制；支持IPv6地址；
- 布线的要求；备份的要求；私有设备和协议要求；
- 服务质量要求；
- 安全性要求；



# 网络设计任务：确定需求

- 技术需求和限制因素：
  - 要求的网络解决方案（语音服务、内容网络和存储网络）；
  - 网络管理；
  - 可用带宽。
- 涉及的商业问题的需求和限制
  - 预算；进度表；人力；法律法规；历史记录；策略
- 每个需求应该评估出其重要性，分配一个权重系数，以便在需求发生冲突时尽量满足最重要的需求

# 网络设计任务：对现有网络的分析

- 对现有网络的分析

- 现有网络可能会带来一些限制，如：现有的布线需要保留；
- 正在运行的协议；
- 现有网络的拓扑结构；
- 已安装的设备及其配置、设备的利用率以及WAN上关键链路的带宽；

# 网络设计任务：概要设计和方案

- 概要设计的准备

- 考虑所有网络需求和约束条件；确定可选方案；选择设备、功能特性、布线等

- 最终设计方案开发

- 产生详细设计图纸；配置规范；成本计算、地址规划等

- 可以实现一个原型网络来验证设计，或在现有网络中实现一个实验性的网络来验证。

# 网络设计任务：部署和监测

- 网络部署

- 根据计划和进度实施；
- 确定工作内容和实施细节；
- 规划好有关人员的培训等

- 监测和重新设计

- 收集运行统计信息；
- 了解网络运行的各项指标；
- 发现故障和异常情况等

# 模块化网络设计

- 模块化设计

- 模块是复合结构的一个组件。
- 首先建立模块，然后把模块进行组合。
- 常用的技术有：层次化设计模型；复合网络模型

# 层次化网络设计

- 层次化模型由下面三个功能组成：
  - 接入层：使用户和工作组可以访问网络资源；
  - 分布层：提供工作组之间以及工作组到核心层之间的连接；
  - 核心层：提供到核心层资源和分布层设备之间的高速传输服务。

# 层次化网络设计

- 把一个简单的网络映射到层次化模型

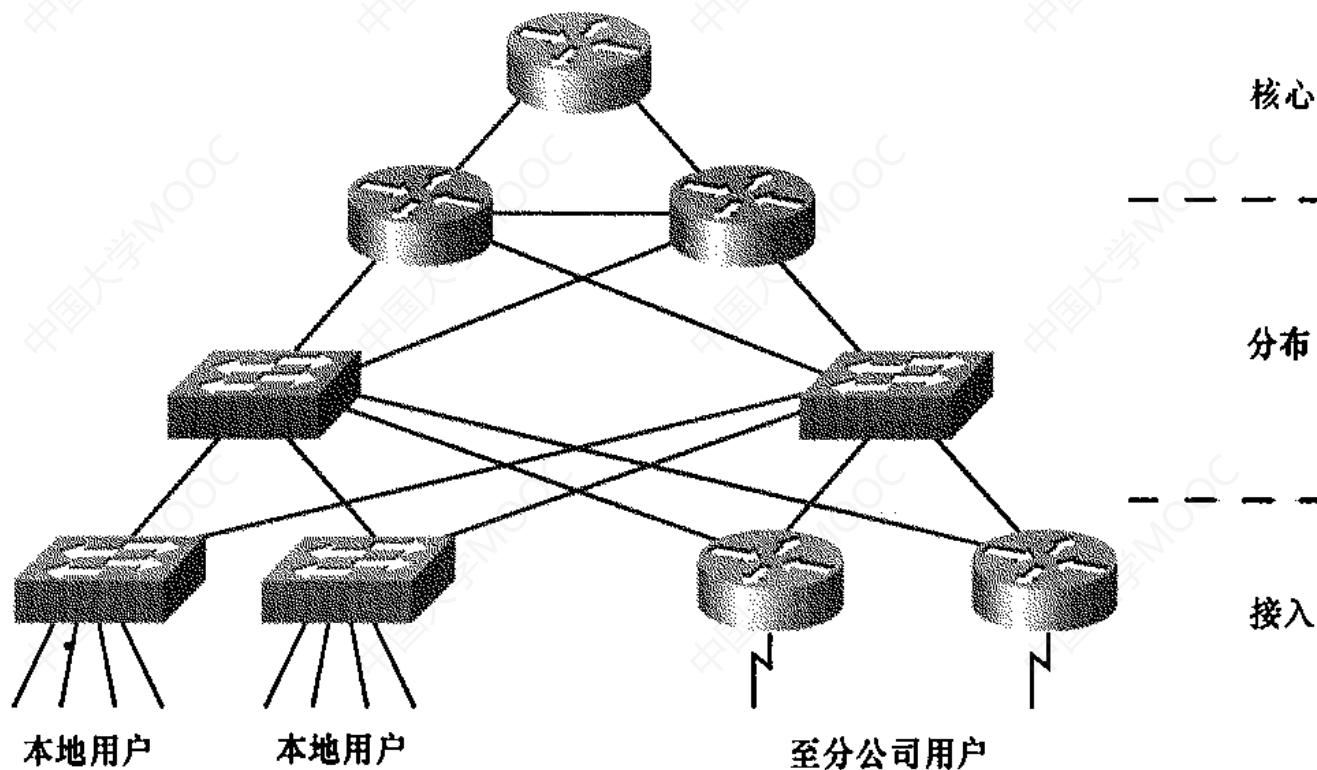


图 1-5 简单网络到层次化模型的映射

# 层次化网络设计：接入层

- 接入层

- 接入层是用户接入网络的地方，可以是本地，也可以是远程。
- 本地用户通过集线器和交换机接入网络。
- 远程用户可以使用VPN连接经互联网接入到内部网络。比如：可以通过PSTN、DSL等方式。其他接入方式可以采用WAN技术，比如：FR、DDN、ISDN



# 层次化网络设计：分布层

## • 分布层

- 分布层是核心层和接入层之间以及接入层工作组之间的接口。
- 在接入层与核心层之间进行路由选择、路由协议处理。
- 执行路由汇总。
- 提供到接入设备和核心设备之间的冗余连接。
- 把多个低速接入的连接汇集到较高速度的核心连接上。

# 层次化网络设计：核心层

- 核心层：核心层提供高速的网络主干。
- 其功能和属性如下：
  - 提供高速、低时延的数据链路和设备。
  - 提供冗余设备和链路使得网络不存在单点故障，实现高可靠、高可用的骨干。
  - 使用快速收敛路由协议可以迅速适应网络变化。

# 路由、交换与远程访问技术

- 一个较为完整的园区网

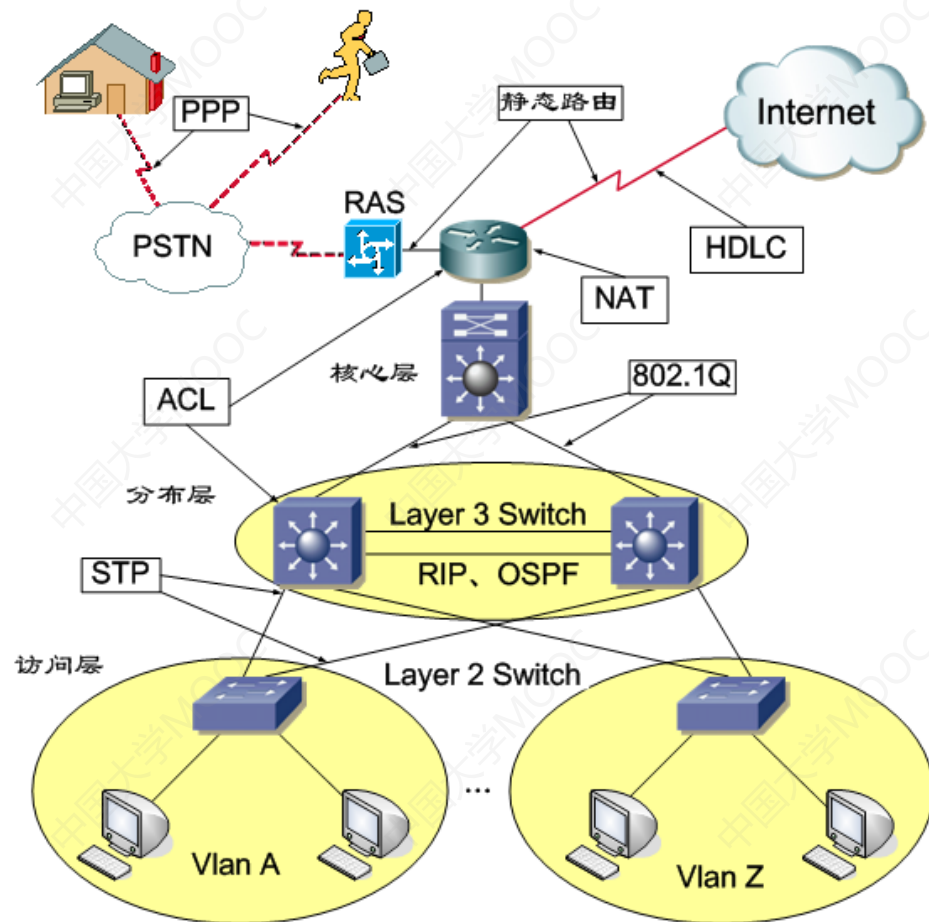
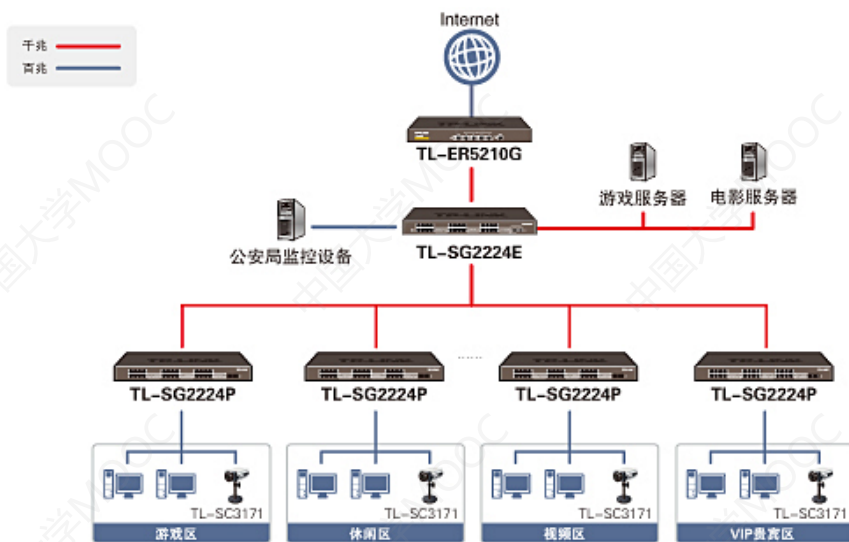
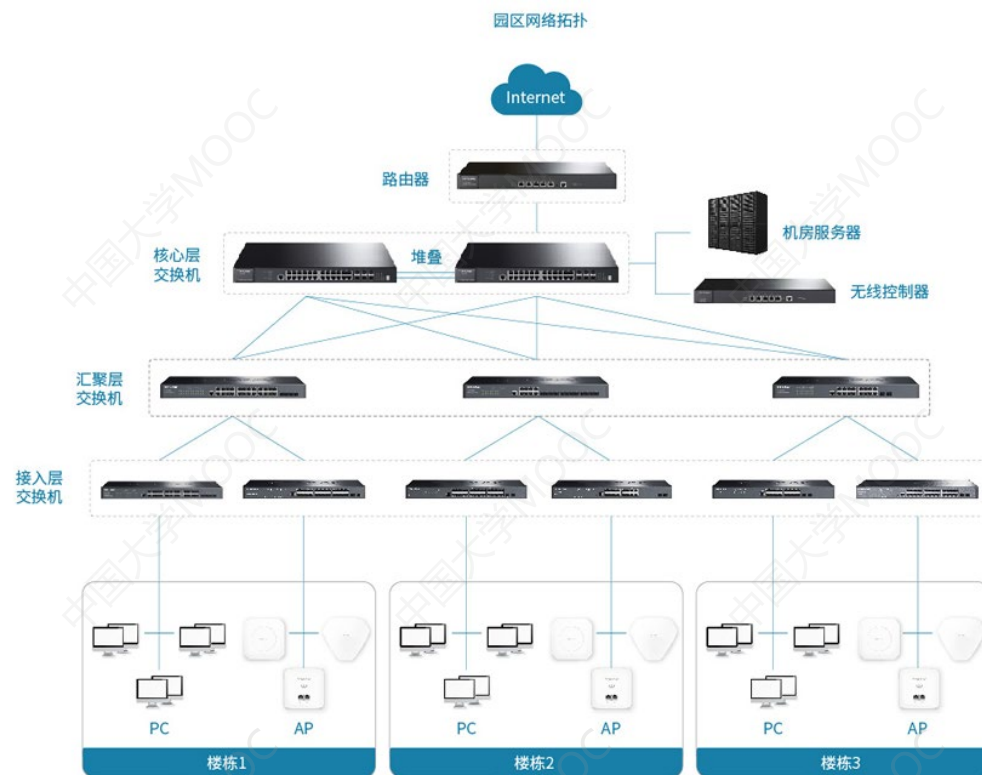
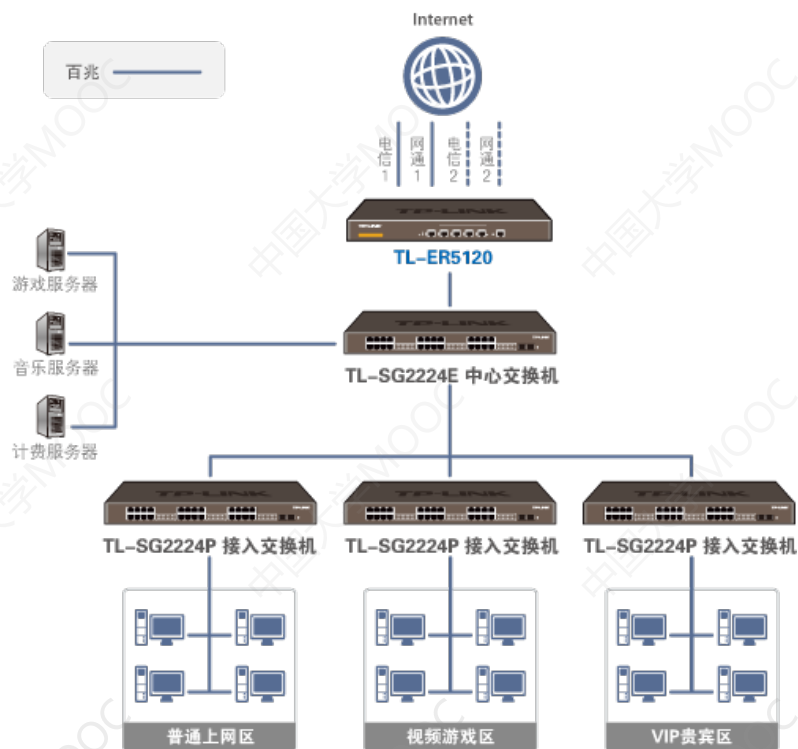


图 1-28 一个较为完整的园区网实现

# 网络实例

- 园区网络实例
- 网吧网络实例



# 内容纲要

1	万维网
2	HTTP协议
3	网络结构设计
4	网络技术展望
5	网络技术和应用趋势

# IP语音

- **Real-time Transport Protocol (RTP)**

- 实时传输协议 ( RTP ) 提供了用于在Internet上传输实时数据的机制
- 基于UDP协议

- **Voice over IP (VoIP)**

- 世界各地的电话公司正用IP路由器取代传统的电话交换机。
- 路由器比传统电话交换机成本低得多。

# 网络安全技术

- 原因

- 每当出现新技术时，罪犯就会考虑他们如何利用这项技术来犯罪。

- 互联网上普遍存在的主要安全问题

- 网络钓鱼；虚假陈述；诈骗
- 拒绝服务；失控；数据丢失

# 网络安全技术

- 安全攻击中使用的技术

- 窃听；重播；缓冲区溢出；地址欺骗；域名欺骗；
- 拒绝服务（DOS）与分布式拒绝服务（DDoS）
- SYN洪流；密码破解；端口扫描；分组拦截

- 信息安全五大要素

- 完整性、保密性、可用性、可控性和不可否认性



# 网络安全技术

- 用于执行安全策略的主要技术。

- 散列；加密；数字签名；数字证书；防火墙

- 入侵检测系统；内容扫描和深度包检查；虚拟专用网VPN

# 网络管理 (SNMP)

- 网络管理员 (网管) 是负责计划、安装、操作、监视和控制构成计算机网络或内部网的硬件和软件系统的人。
- 虽然网络硬件和协议软件包含自动绕过故障或重发丢失分组的机制，但网络管理者需要检测和纠正潜在的问题。
- SNMP协议确切地定义了管理器如何与代理通信。

# 软件定义网络 (SDN)

- 一种新型网络创新架构，是网络虚拟化的一种实现方式
- OpenFlow协议用于控制器与网络元素之间通信
- SDN有潜力引入网络产业的重大变革。
- 一旦网络管理设施被移除，交换硬件将成为商品，网络管理者将不会有动机从单个供应商购买所有设备。

# 物联网（IoT）

- 物联网：使用机器对机器通信的连接的嵌入式系统。
  - 与使用传统计算机的早期应用不同，物联网应用专注于嵌入式系统。
  - 应用包括家庭自动化、智能电网、安全和零售系统。
- 许多技术需要低功耗无线通信。
  - 为了最小化功率利用率，无线网络可以使用网格方法：
  - 低功率节点同意彼此转发数据包，而不是使用能够到达长距离的强大无线电发射器。

# 内容纲要

1	万维网
2	HTTP协议
3	网络结构设计
4	网络技术展望
5	网络技术和应用趋势

# 网络技术和应用趋势

- 对可扩展因特网服务的需求
- 内容缓存 ( Akamai )
- 网络负载均衡器
- 服务器虚拟化
- P2P：对等通信
- 分布式数据中心和复制
- 社交网络

# 网络技术和应用趋势

- 移动性和无线网络
- 数字视频
- 高速接入和交换
- 云计算
- 覆盖网络
- 中间件
- IPv6的广泛部署

谢谢观看



厦門大學  
XIAMEN UNIVERSITY



信息学院 黄 烽  
(特色化示范性软件学院) 博士, 副教授  
School of Informatics Wei Huang