

Lecture 1-1

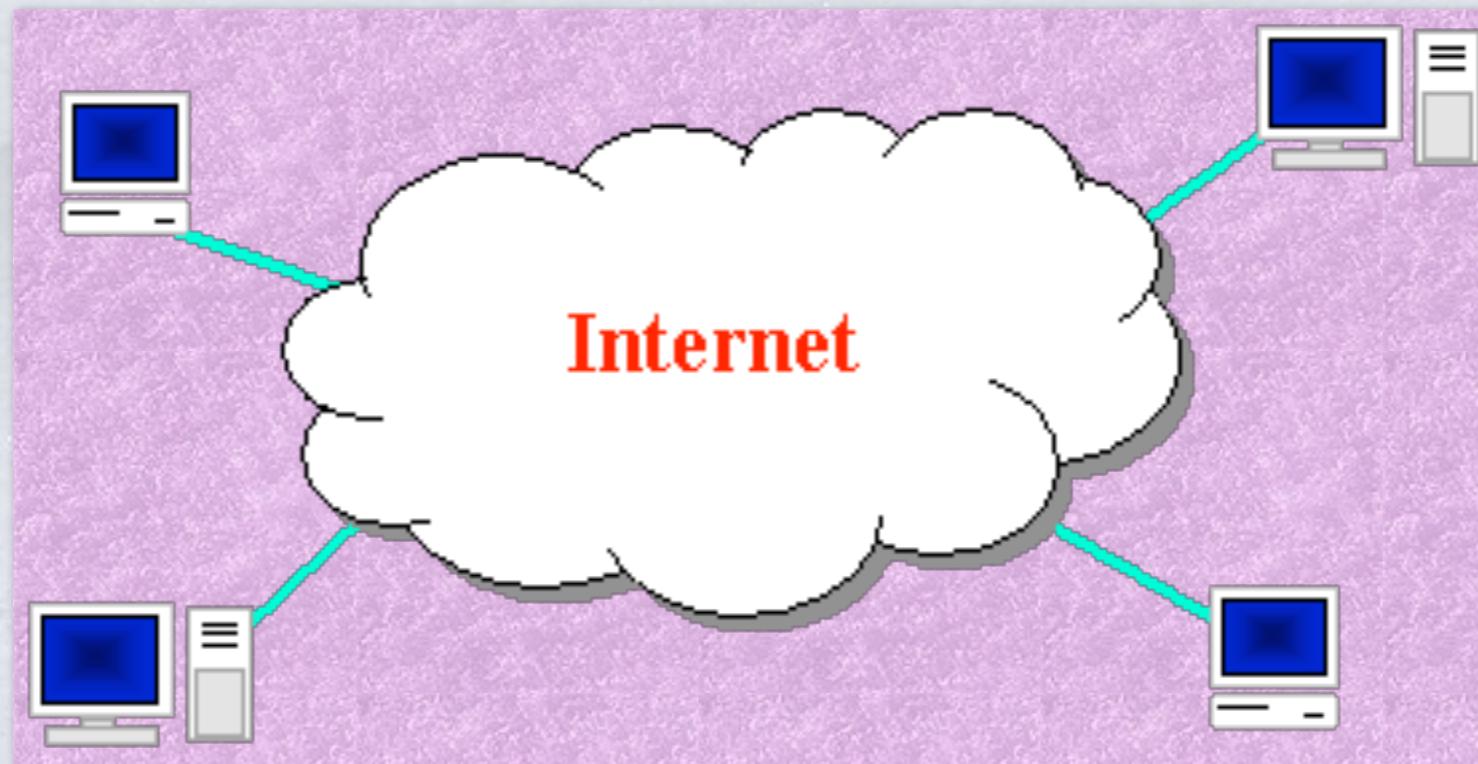
互联网



Play



互联网



❄ Wikipedia: <http://en.wikipedia.org/wiki/Internet>

- * a connection of computer networks using the Internet Protocol Suite (TCP/IP)

❄ 互联网与万维网有何不同?

- * WWW = HTML* + HTTP(S)
- * 包含 CSS, JavaScript, 以及其他浏览器内容

简要发展史

- ❄ 万维网，1989-91由Tim Berners-Lee创造
 - * This is for Everyone
- ❄ 发布流行的浏览器:
 - * Netscape 1994, IE 1995
- ❄ Amazon.com 1995
- ❄ Google 1996
- ❄ 2000-2002 .COM
- ❄ 2004 web2.0
- ❄ 2016 web3.0



相关组织机构

Internet Engineering Task Force (IETF)

- * 互联网协议标准

Internet Corporation for Assigned Names and Numbers (ICANN)

- * 顶级域名

World Wide Web Consortium (W3C)

- * Web标准



网际互连协议(IP)

IPv4

IPv6

传输控制协议(TCP)

✿ 是一种面向连接的、可靠的、基于字节流的传输层通信协议，由IETF的RFC 793定义。

✿ 命令：

- * netstat (Windows)
- * CurrPorts

用户数据报协议 (UDP)

- ❄ User Data Protocol, 是传输层协议
- ❄ 一个简单的面向数据报的传输层协议，提供的是非面向连接的、不可靠的数据流传输
- ❄ 没有超时重发等机制，故而传输速度很快

域名系统

- ❄ DNS是域名系统,Domain Name System的缩写
- ❄ 因特网上作为域名和IP地址互相映射的一个分布式数据库

❄ DNS系统的作用

- * 正向解析：根据主机名称（域名）查找对应的IP地址
- * 反向解析：根据IP地址查找对应的主机域名

DNS之前

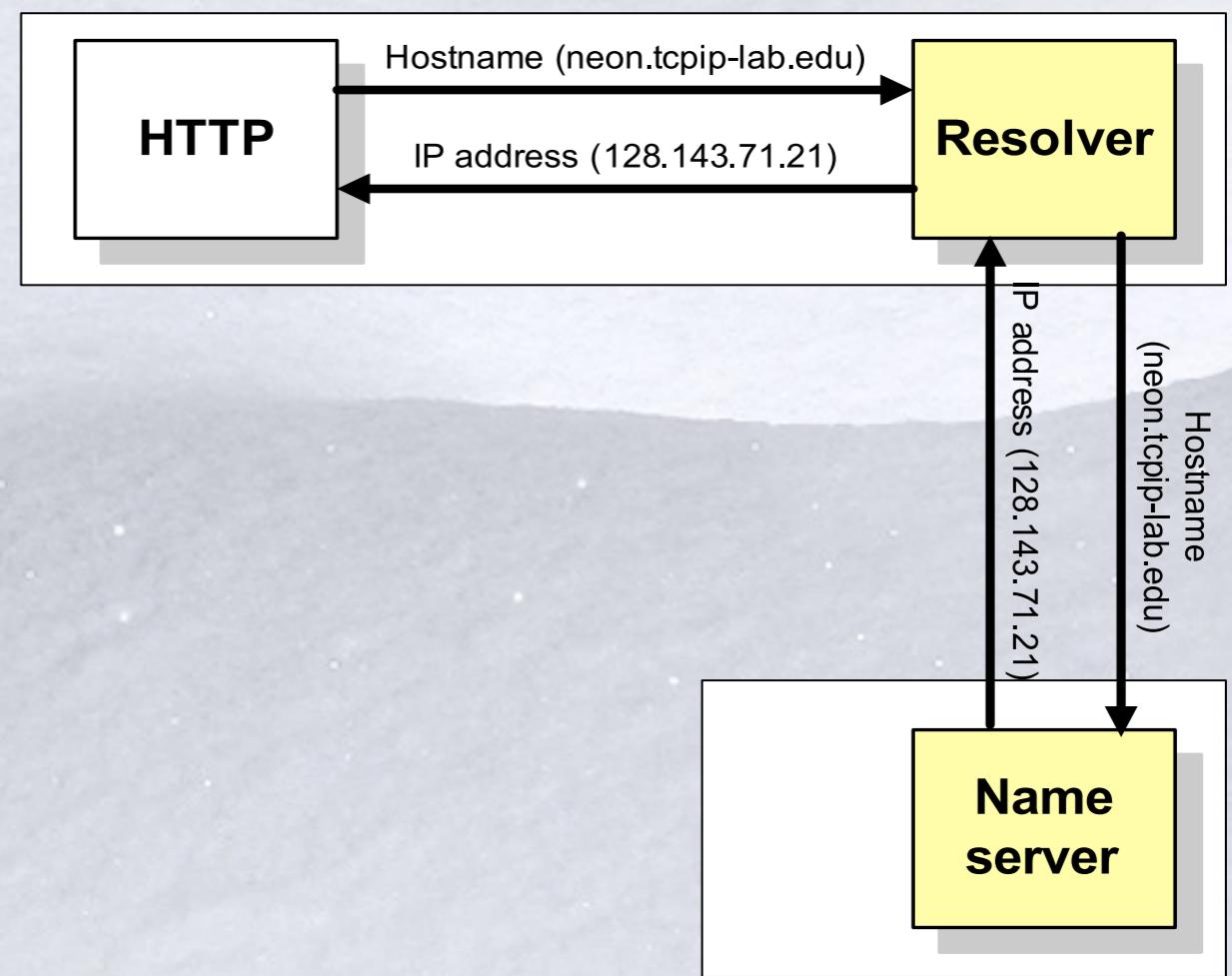
.... hosts.txt文件

❄ DNS 之前(直到1985), name-to-IP地址是通过使用FTP从中央服务器下载单个文件(hosts.txt)完成的

- * hosts.txt中的名称不是结构化的
- * hosts.txt文件在大多数操作系统上仍然可以工作，可用于定义本地名称
 - * Windows: C:\Windows\system32\drivers\etc\hosts
 - * Mac: /private/etc/hosts
 - * Linux: /etc/hosts

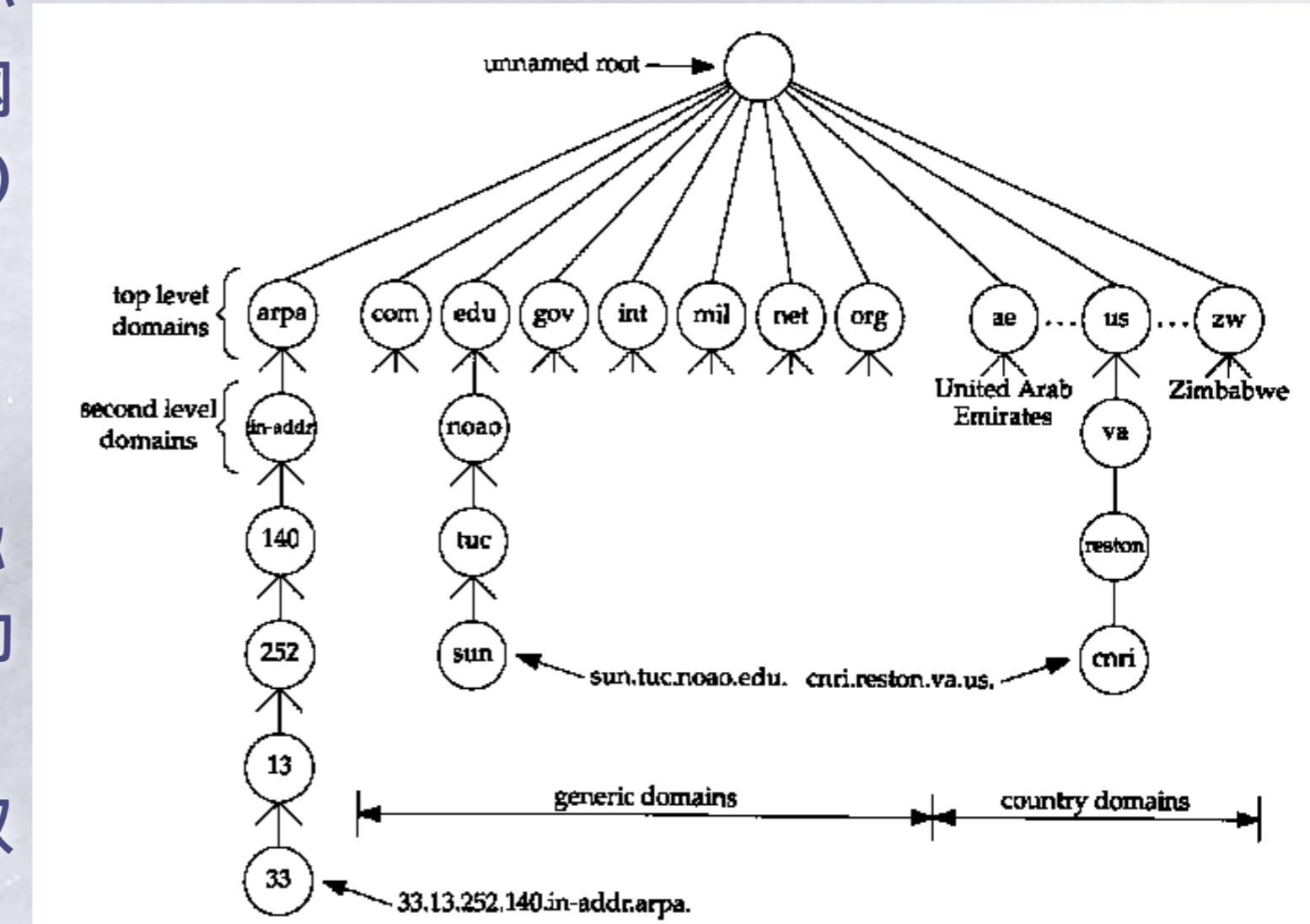
解析器与名称服务器

- 主机上的应用程序通过 DNS客户端访问域名系统，称为解析器
- 解析器联系DNS服务器，称为名称服务器
- DNS服务器返回IP地址给解析器，解析器将IP地址传递给应用程序
- 反向查找也可以，即查找给定IP地址的主机名



DNS名称层次结构

- ❄ DNS层次结构可以用树表示
- ❄ 根域名和顶级域名由互联网中央名称注册机构(ICANN)管理。
- ❄ 在顶级域名(TLD, Top-Level Domain)之下，名称空间的管理被委派给不同的组织
- ❄ 每个组织都可以进一步授权



域名

- ✿ 主机和DNS域名根据其在域名树中的位置命名
- ✿ DNS域名树中的每个节点都可以通过唯一的全限定域名(FQDN)来标识。FQDN给出了在DNS树中的位置。

software.nju.edu.cn or software.nju.edu.cn.
- ✿ 一个FQDN由标签(“software”, “nju”, “edu”, “cn”)组成，标签之间用句号(“.”)隔开。
 - * 结尾可以有一个句号(“.”)。
 - * 每个标签最多可以有63个字符
 - * FQDN包含字符、数字和破折号(" - ")
 - * fFQDN不区分大小写

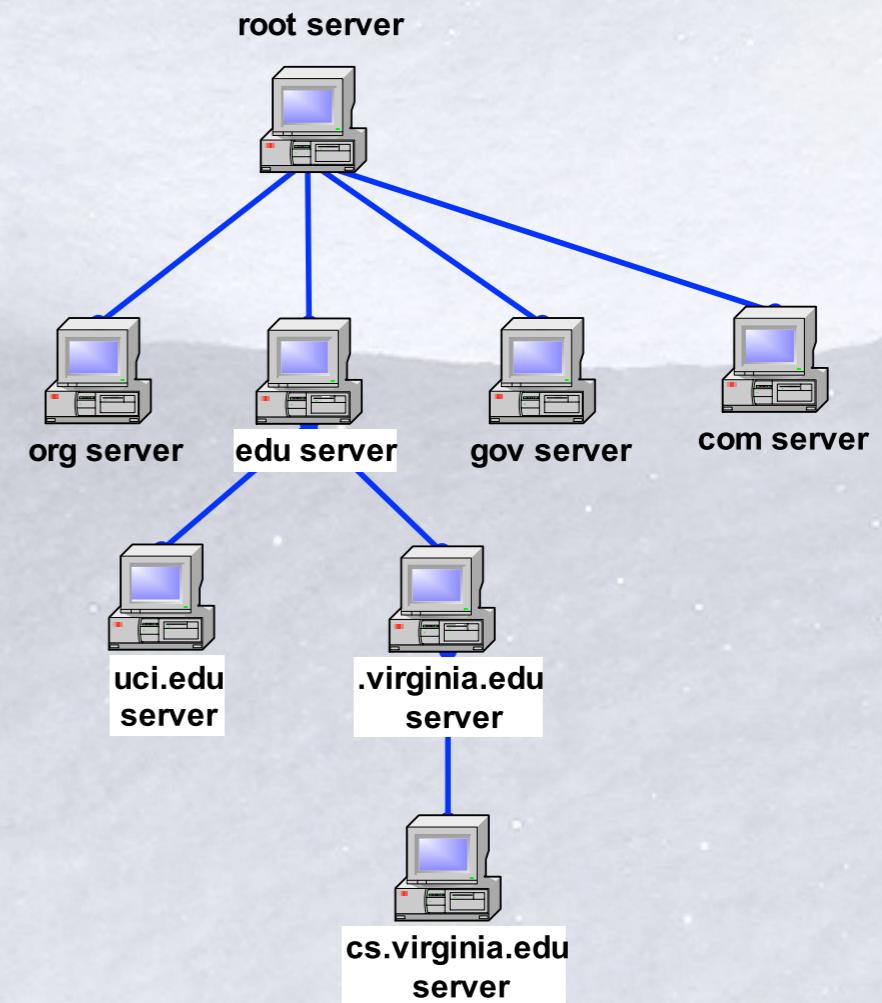
名称服务器的层次结构

❄ 层次名称空间的解析是由层次化的名称服务器完成的

❄ 每个服务器负责(授权)DNS名称空间的一个连续部分，称为区域。

❄ 区域是子树的一部分

❄ DNS服务器对所属区域内的主机进行应答



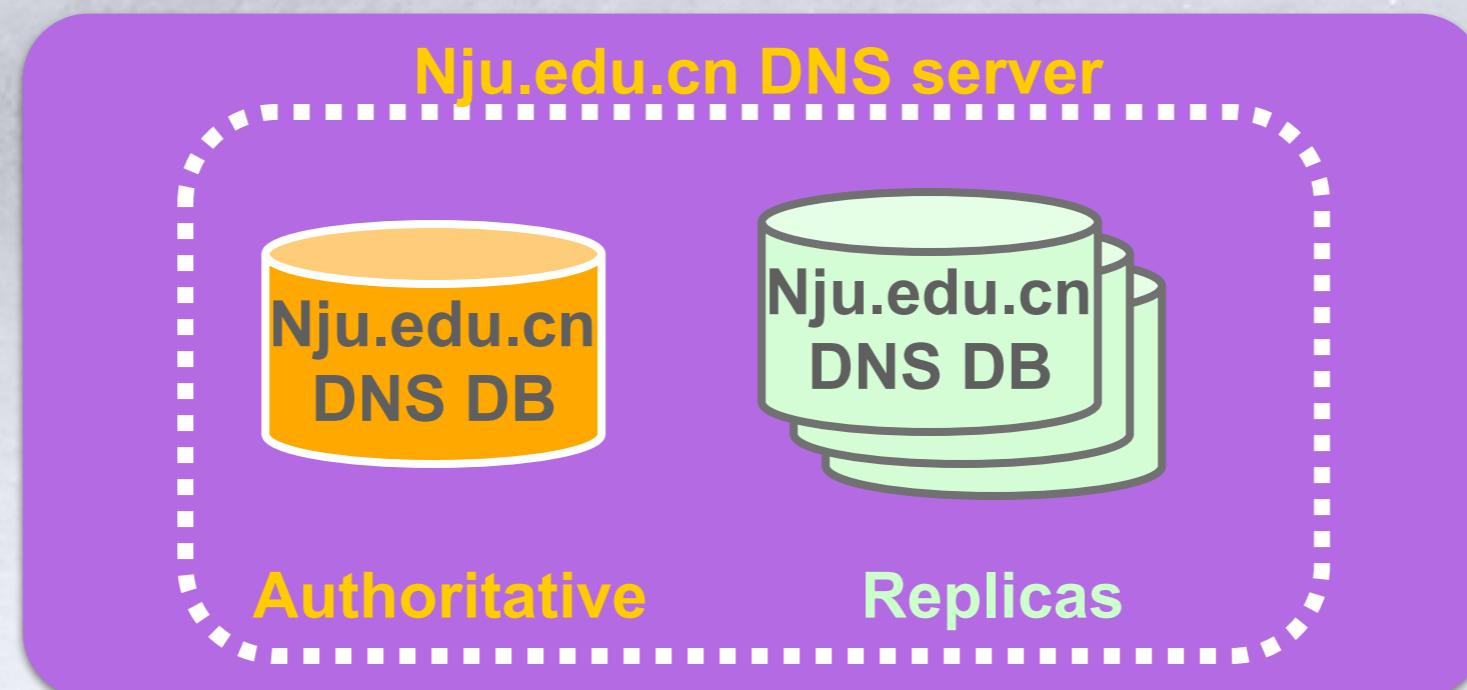
主名称服务器和辅名称服务器

❄ 对于每个区域，必须有一个主名称服务器和辅助名称服务器

- * 主服务器(主服务器)维护一个区域文件，其中包含关于区域的域名信息。对主服务器进行更新。
- * 辅服务器复制存储在主服务器上的数据。

❄ 添加一个主机：

- * 当一个新主机(“xxx.cs.xxx.edu.cn”)被添加到一个区域时，管理员将该主机的IP信息(IP地址和名称)添加到主服务器上的配置文件中

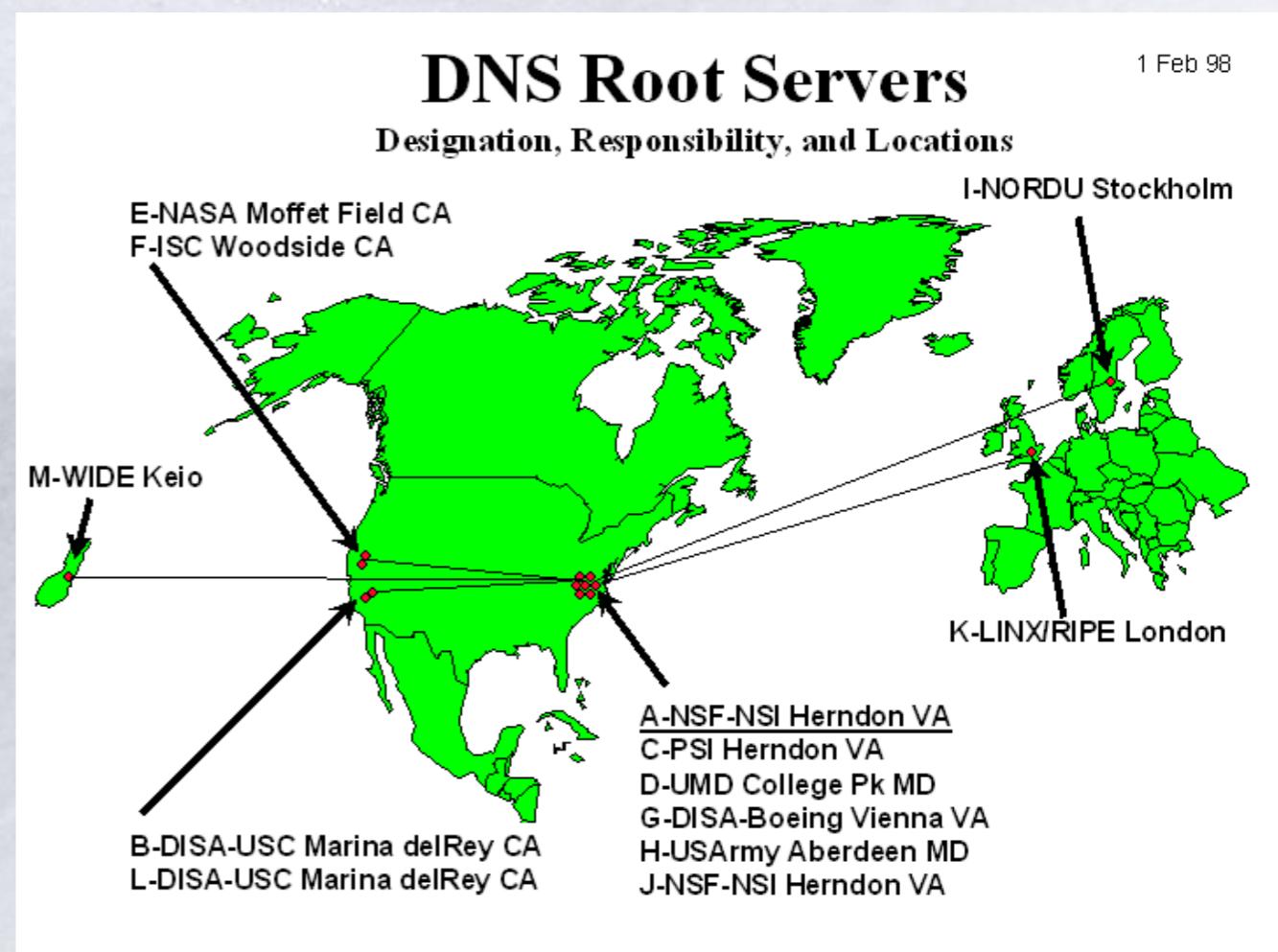


根域名服务器

❄ 互联网的顶级域名解析
服务由根服务器完成

❄ 只有13个根域名服务器

❄ 根服务器对网络安全、
运行稳定至关重要，被
称为互联网的“中枢神
经”



域名解析

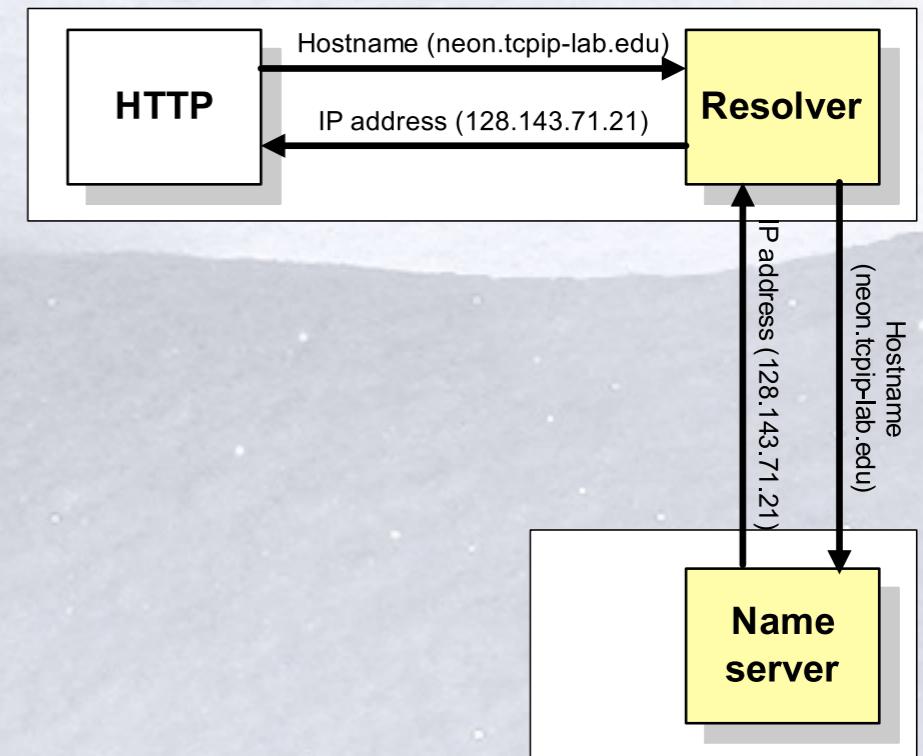
* 用户程序发出主机名到IP地址的请求

* 本地解析器向主机的名称服务器提出DNS查询

* 名称服务器检查它是否被授权回答查询。

- * 如果是，它会响应。
- * 否则，它将从根树开始查询其他名称服务器

* 当名称服务器得到答案时，它将其发送到解析器。



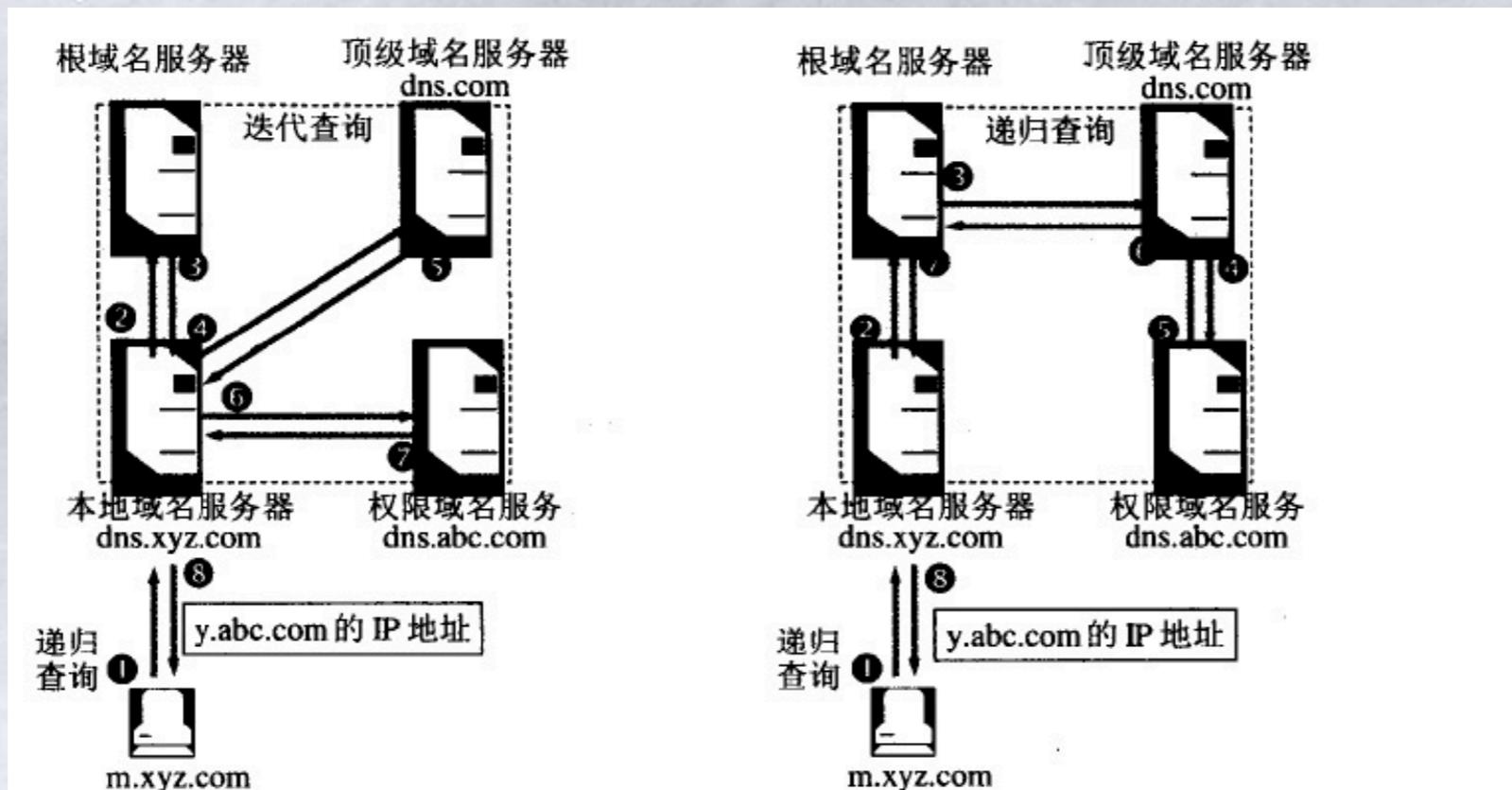
递归和迭代查询

* 查询有两种类型:

- * 递归查询
- * 迭代(非递归)查询

* 递归查询:如果主机所询问的本地域名服务器不知道被查询的域名的IP地址,那么本地域名服务器就以DNS客户的身份,向其它根域名服务器继续发出查询请求报文(即替主机继续查询),而不是让主机自己进行下一步查询。

* 迭代查询:当根域名服务器收到本地域名服务器发出的迭代查询请求报文时,要么给出所要查询的IP地址,要么告诉本地服务器:“你下一步应当向哪一个域名服务器进行查询”。



缓存

- 为了减少DNS流量，名称服务器会缓存域名/IP地址映射信息
- 当查询的条目在缓存中时，服务器不会联系其他服务器

资源记录(Resource Records)

- ✿ 分布式数据库的数据库记录称为资源记录(RR)。
- ✿ 资源记录存储在名称服务器上的配置文件(分区文件)中。
- ✿ DNS数据库包含的不仅仅是主机名到地址的记录:
 - * 名称服务器记录 ns
 - * 主机名别名 CNAME
 - * 邮件服务器 mx
 - * 主机信息 hinfo

资源标识符

统一资源标识符

- * Uniform Resource Identifier, URI
- * 用于标识某一互联网资源名称的字符串，该种标识允许用户对任何（包括本地和互联网）的资源通过特定的协议进行交互操作。

统一资源定位符

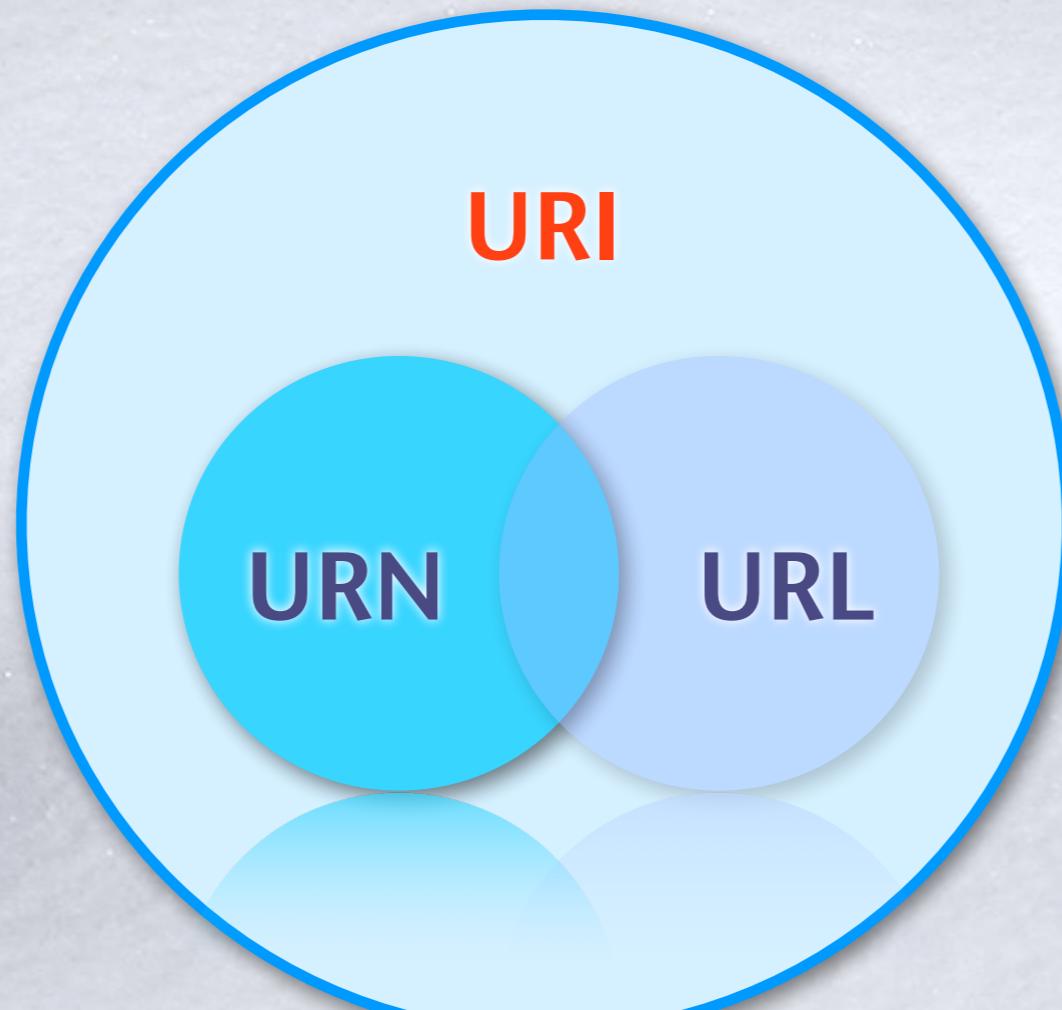
- * Uniform Resource Locator, URL
- * URL是一种URI，它标识一个互联网资源，并指定对其进行操作或获取该资源的方法。

统一资源名称

- * Uniform Resource Name, URN
- * 定义某事物的身份
- * URN仅用于命名，而不指定地址

URI

URI 是 URL 和 URN 的超集



URI

- ❄ RFC 2396
- ❄ URI可被视为定位符（URL），名称（URN）或两者兼备
- ❄ URI格式由URI协议名（例如http、ftp、mailto、file），一个冒号，和协议对应的内容所构成

[协议名]://[用户名]:[密码]@[服务器地址]:[服务器端口号]/[路径]?[查询字符串]#[片段ID]

URN

- ❄ URN是基于某名字空间通过名称指定资源的URI。
- ❄ 位置无关
- ❄ 需要基础设施，因而未被广泛应用

Example URN

❄ urn:ietf:rfc:2141

URL

- 统一资源定位符 URL 是对可以从因特网上得到的资源的位置和访问方法的一种简洁的表示。
- 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。



URL 的一般形式

`http://host.company.com:80/a/b/c.html?user=Alice&year=2018#p2`

- ❄ 协议 (`http:`): 标识用于获取内容的协议.
- ❄ 主机 (`//host.company.com`): 要连接到的计算机的名称.
- ❄ 端口 (80): 允许多个服务器在同一台机器上运行.
- ❄ 路径 (`/a/b/c.html`): 服务器用于查找内容.
- ❄ 查询字段 (`?user=Alice&year=2008`): 提供其他参数
- ❄ 片段标识符(`p2`): 单页面路由, 锚点定位

URL协议

- ❄️ http
- ❄️ https
- ❄️ file
- ❄️ mailto
- ❄️ 还有其他几种协议，比如ftp，但已经不常用了

URL路径 (/a/b/c.html)

✿ 传递到web服务器进行解释。早期的web服务器：

- * 静态HTML文件的路径名。
- * 生成HTML内容的程序的路径名(例如foo.php)。

✿ 路由信息

- * 将层次位置映射到要执行的函数以及函数的参数

✿ 应用程序编程接口(API)设计，示例：

- * /user/create
- * /user/list
- * /user/0x23490
- * /user/0x23433
- * /user/delete/0x23433

查询字段

❄️ 查询字段是指在URL的末尾加上用于向服务器发送信息的字符串（变量）。将“?”放在URL的末尾，然后再加上“参数=值”，想加上多个参数的话，使用“&”。

❄️ 网站访问分析（被动参数）

- * 被动参数对显示的内容没有影响。
- * 为了了解用户从哪里到达了自己的网站，设定固有的参数来统计。
- * 通常，当想要分析网站流入的来源以吸引客户，或者想要知道来自搜索引擎广告的流入客户时，可以使用它。

❄️ 显示动态页面结果（活动参数）

- * 会影响显示的内容。换句话说，添加参数将改变网站上显示的内容。

查询字段

❄️ 查询字段是指在URL的末尾加上用于向服务器发送信息的字符串（变量）。将“?”放在URL的末尾，然后再加上“参数=值”，想加上多个参数的话，使用“&”。

❄️ 网站访问分析（被动参数）

- * 被动参数对显示的内容没有影响。
- * 为了了解用户从哪里到达了自己的网站，设定固有的参数来统计。
- * 通常，当想要分析网站流入的来源以吸引客户，或者想要知道来自搜索引擎广告的流入客户时，可以使用它。

❄️ 显示动态页面结果（活动参数）

- * 会影响显示的内容。换句话说，添加参数将改变网站上显示的内容。

超文本 Hypertext

- ❄ 超文本是用超链接的方法，将各种不同空间的文字信息组织在一起的网状文本。
- ❄ 超链接是Web页面区别于其他媒体的重要特征之一，网页浏览者只要单击网页中的超链接就可以自动跳转到超链接的目标对象，且超链接的数量是不受限制的。
- ❄ 用标签[来实现](#)

```
<a href="https://en.wikipedia.org/wiki/URL">URL</a>
```

种类

网页上的超链接一般分为三种：

- * 绝对URL的超链接
- * 相对URL的超链接
- * 同一网页的超链接，锚点链接/书签链接。

URL的使用

❄ 加载页面，在浏览器地址栏输入URL

❄ 加载图像

- * ``

❄ 加载样式表

- * `<link rel="stylesheet" type="text/css" href="...">`

❄ 嵌入页面

- * `<iframe src="http://www.google.com">`

URL编码

- ❄ URL编码将字符转换为可通过因特网传输的格式。
- ❄ URL只能使用ASCII 字符集通过因特网进行发送。
- ❄ 由于URL通常包含 ASCII 集之外的字符，因此必须将URL转换为有效的ASCII 格式。
- ❄ URL编码使用后跟十六进制数字的 "%" 替代不安全的 ASCII 字符。
- ❄ URL不能包含空格。URL 编码通常使用加号 (+) 或 %20 替代空格。



Thanks!!!

