



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场



计算机网络之网尽其用

主讲人：聂兰顺

本讲主题

Web缓存/代理服务器技术



Web缓存/代理服务器技术

❖ 功能

- 在不访问服务器的前提下满足客户端的HTTP请求。

❖ 为什么要发明这种技术？

- 缩短客户请求的响应时间
- 减少机构/组织的流量
- 在大范围内(Internet)实现有效的内容分发



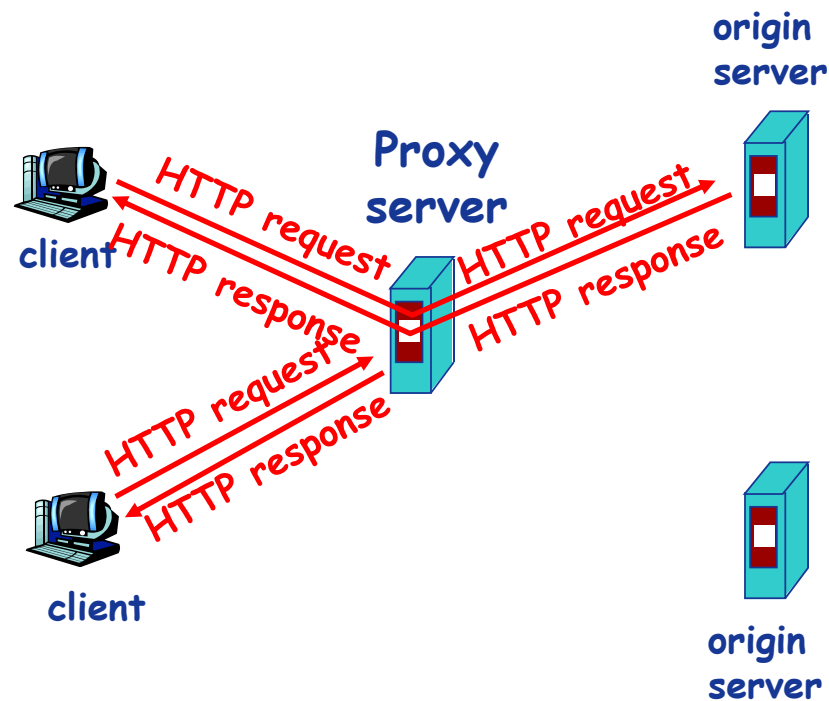
Web缓存/代理服务器技术

❖ Web缓存/代理服务器

- 用户设定浏览器通过缓存进行Web访问
- 浏览器向缓存/代理服务器发送所有的HTTP请求
 - 如果所请求对象在缓存中，缓存返回对象
 - 否则，缓存服务器向原始服务器发送HTTP请求，获取对象，然后返回给客户端并保存该对象

❖ 缓存既充当客户端，也充当服务器

❖ 一般由ISP(Internet服务提供商)架设



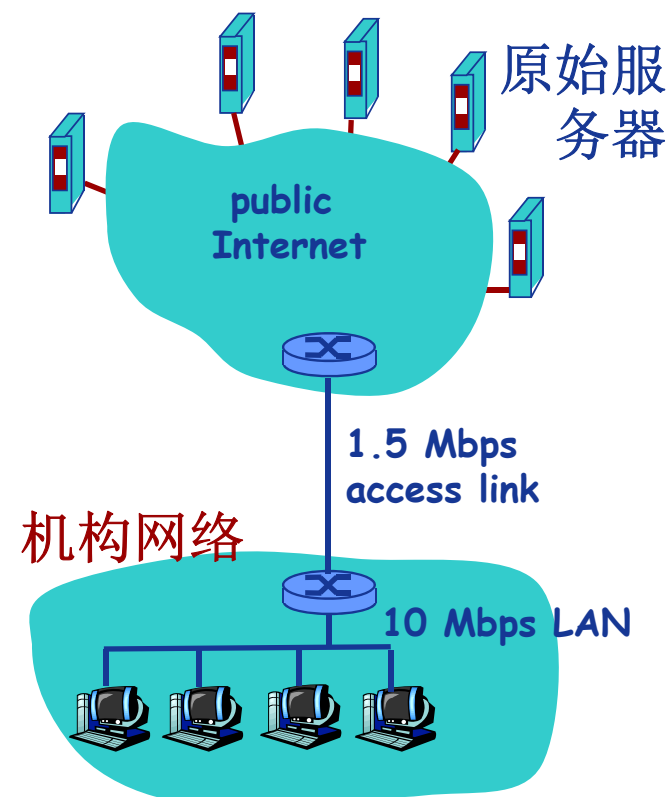
Web缓存示例(1)

❖ 假定:

- 对象的平均大小=100,000比特
- 机构网络中的浏览器平均每秒有15个到原始服务器的请求
- 从机构路由器到原始服务器的往返延迟=2秒

❖ 网络性能分析:

- 局域网(LAN)的利用率=15%
- 接入互联网的链路的利用率=100%
- 总的延迟=互联网上的延迟+访问延迟+局域网延迟=2秒+几分钟+几微秒



Web缓存示例(2)

❖ 解决方案1:

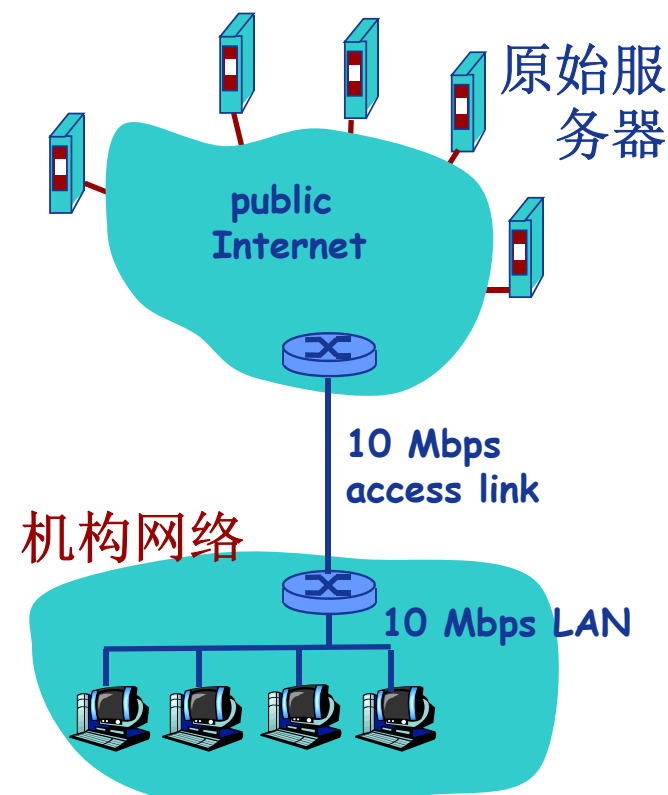
- 提升互联网接入带宽=10Mbps

❖ 网络性能分析:

- 局域网(LAN)的利用率=15%
- 接入互联网的链路的利用率=15%
- 总的延迟=互联网上的延迟+访问延迟+局域网延迟=2秒+几微秒+几微秒

❖ 问题:

- 成本太高



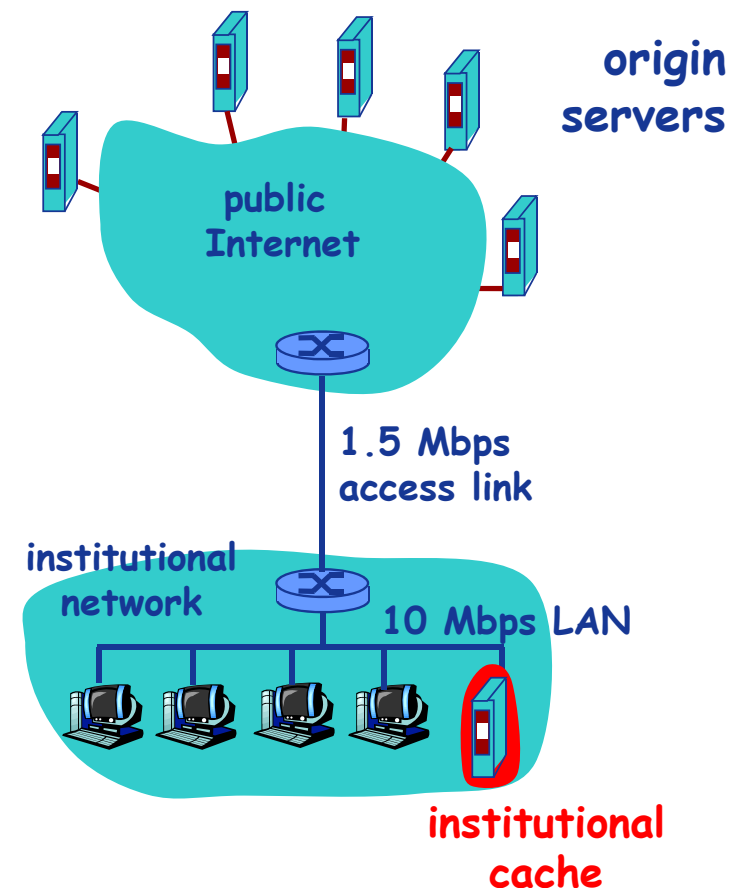
Web缓存示例(3)

❖ 解决方案2:

- 安装Web缓存
- 假定缓存命中率是0.4

❖ 网络性能分析:

- 40%的请求立刻得到满足
- 60%的请求通过原始服务器满足
- 接入互联网的链路的利用率下降到60%，从而其延迟可以忽略不计，例如10微秒
- 总的平均延迟=互联网上的延迟+访问延迟+局域网延迟= $0.6 \times 2.01\text{秒} + 0.4 \times n\text{微秒} < 1.4\text{秒}$



条件性GET方法

❖ 目标:

- 如果缓存有最新的版本，则不需要发送请求对象

❖ 缓存:

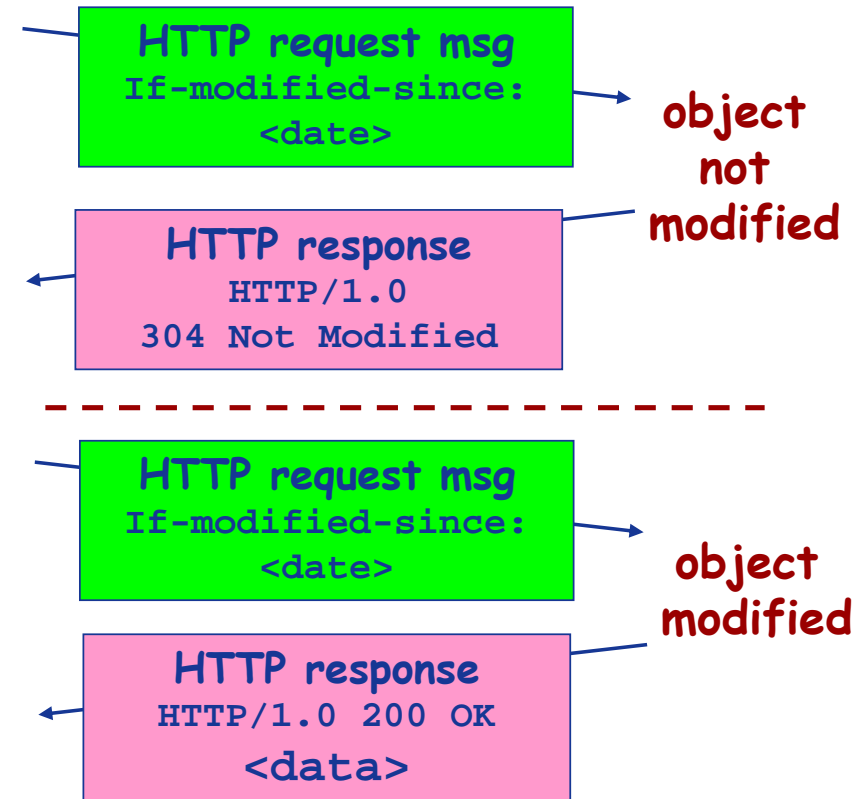
- 在HTTP请求消息中声明所持有版本的日期
- If-modified-since: <date>

❖ 服务器:

- 如果缓存的版本是最新的，则响应消息中不包含对象
- HTTP/1.0 304 Not Modified

cache

server



课后作业

检索文献，分析、总结**Web**技术近年来有哪些新进展？其关键思想和技术是什么？





哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场



谢谢!