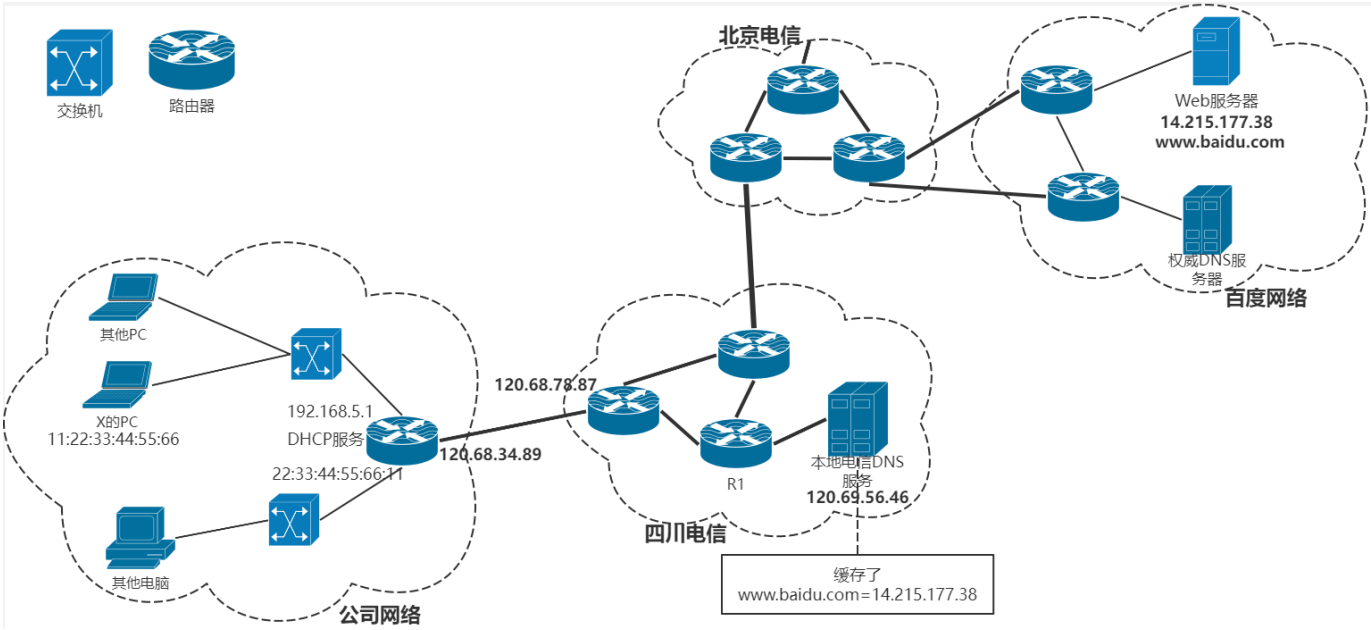
一台新 PC 进行 Web 页面请求的历程

注意：因为本文档属于补充资料，课外阅读部分，所以并不会提供额外的技术支持和答疑，敬请谅解。

场景和网络拓扑说明

场景：一名同学 X，入职成都一家新公司 NewCompany，年薪 50 万，公司福利很好，给他派发了一台全新的笔记本电脑，现在 X 同学将他的电脑接入公司的网络，准备打开百度的页面 www.baidu.com（IP 地址：14.215.177.38），NewCompany 公司的 ISP 服务由四川电信提供，百度公司的 ISP 服务由北京电信提供。



假设其中笔记本电脑的 mac 地址是：11:22:33:44:55:66，网关路由器对内的网关地址 192.168.5.1、对内的 mac 地址 22:33:44:55:66:11 和对外的 Internet 地址 120.68.34.89，因为 mac 地址主要用于局域网内的寻址，对外的的 Mac 地址无关紧要。同时路由器还承担着 DHCP 服务器的职责。

同时在我们下面的描述中，默认百度所有的数据内容都放在百度公司内部的服务器之上，并没有使用 CDN 之类的机制。

交换机

工作在链路层,负责接收链路层入帧并将它们转发到链路层另一出口，交换机自身对子网中的主机和路由器是透明的。交换机内部存在着交换机表，里面的每一个表项都至少包含了①一个 MAC 地址;②通向该 MAC 地址的交换机接口。

对于从接口 X 接收到的一个链路层入帧，交换机的处理是：获得入帧中的目的 MAC 地址，并在自己内部的交换机表寻找

1、如果没有对于目的地址的表项，交换机广播该帧。

2、表中有一个表项将目的 MAC 地址到达接口 X 联系起来。交换机丢弃该帧。

3、表中有一个表项将目的 MAC 地址与接口 Y≠X 联系起来。交换机通过将该帧放到接口 Y 前面的输出缓存完成转发功能。

交换机是即插即用设备，自学习的。交换机表初始为空，对于收到对于在每个接口接收到的每个入帧，该交换机在其交换机表中存储 MAC 地址和接口的对应关系。

在实际工作中，常会听到类似于“汇聚交换机”、“核心交换机”之类的名词，其实这些本质上都是我们下面要说到的路由器，因为这些交换机工作在网络层而非链路层。

路由器

路由器工作在网络层，在输入端口接受到数据包后，解析后根据 IP 地址，在内部的路由表中寻找后，经过内部的交换结构往输出端口输送，使数据包可以到达正确的 IP 地址。在路由寻址算法上有集中式路由选择算法和分散式路由选

择算法两种。集中式路由选择算法最出名的就是图论中的 Dijkstra 算法，这种算法必须知道整个网络的情况。

两种算法各有优劣，所以实际工作中两种算法会结合工作。可以这样理解，一个 ISP 内部，使用了基于 Dijkstra 算法的 OSPF 协议；多个 ISP 之间采用的 BGP 协议，则算法思想接近分散式路由选择算法。

比如我们上面的网络拓扑图中，四川电信和北京电信内部的路由器就可能使用 OSPF 协议进行路由规划，四川电信和北京电信之间的路径可能就使用 BGP 协议规划。当然路由表也可以手工配置。

准备：DHCP、IP 等等

当 X 同学首先将其笔记本 PC 与网络连接时，没有 IP 地址他就不能做任何事情。所以，X 同学的笔记本 PC 所采取的一个网络相关的动作是运行 DHCP 协议，以从本地 DHCP 服务器获得一个 IP 地址以及其他信息。

名称解释 DHCP：动态主机配置协议，为一个新接入的主机分配一个 IP 地址。

1 ） X 同学笔记本 PC 上的操作系统生成一个 DHCP 请求报文，并将这个报文放入 UDP 报文段，该 UDP 报文段则被放置在一个具有广播 IP 目的地址（255.255.255.255）和源 IP 地址 0.0.0.0 的 P 数据报中，因为 X 同学的笔记本 PC 还没有一个 IP 地址。

2）包含 DHCP 请求报文的 IP 数据报则被放置在以太网帧中。该以太网帧具有目的 MAC 地址 FF:FF:FF:FF:FF:FF，使该帧将广播到与交换机连接的所有设备；该帧的源 MAC 地址是 X 同学笔记本 PC 的 MAC 地址 11:22:33:44:55:66。

3）包含 DHCP 请求的广播以太网帧是第一个由 X 同学笔记本 PC 发送到以太网交换机的帧。该交换机在所有的出端口广播入帧，包括连接到路由器的端口。

4）路由器在它的具有 MAC 地址 22:33:44:55:66:11 的接口接收到该广播以太网帧，该帧中包含 DHCP 请求，并且从该以太网帧中抽取出 IP 数据报。该数据报的广播 IP 目的地址指示了这个 IP 数据报应当由在该节点的高层协议处理，因此该数据报的载荷（一个 UDP 报文段）被分解向上到达 UDP，DHCP 请求报文从此 UDP 报文段中抽取出来。此时 DHCP 服务器有了 DHCP 请求报文。

5）假设 DHCP 服务器分配地址 192.168.5.10 给 X 同学的笔记本 PC。DHCP 服务器生成包含这个 IP 地址以及 DNS 服务器的 IP 地址（120.69.56.46）、默认网关路由器的 IP 地址(192.168.5.1）和子网掩码（255.255.255.0）的一个 DHCP ACK 报文。

该 DHCP 报文被放入一个 UDP 报文段中，UDP 报文段被放入一个 IP 数据报中，IP 数据报再被放入一个以太网帧中。这个以太网帧的目的 MAC 地址是 X 同学笔记本 PC 的 MAC 地址（11:22:33:44:55:66）。

6）包含 DHCP ACK 的以太网帧由路由器发送给交换机。因为交换机是自学习的，并且先前从 X 同学笔记本 PC 收到（包含 DHCP 请求的）以太网帧，所以该交换机知道寻址到 11:22:33:44:55:66 的帧仅从通向 X 同学笔记本 PC 的输出端口转发。

7） X 同学笔记本 PC 接收到包含 DHCP ACK 的以太网帧，从该以太网帧中抽取 IP 数据报,从 IP 数据报中抽取 UDP 报文段，从 UDP 报文段抽取 DHCP ACK 报文。X 同学 的 DHCP 客户则记录下分配给它的 IP 地址和它的 DNS 服务器的 IP 地址。它还在其 IP 转发表中安装默认网关的地址。

X 同学笔记本 PC 将向该默认网关发送目的地址为其子网 192.168.5.X 以外的所有数据报。

准备：DNS 和 ARP

名词解释： DNS，Domain Name System，专门提供将主机名转换为其背后的 IP 地址，DNS 协议运行在 UDP 之上，使用 53 号端口。

当 X 同学 将 www. baidu. com 的 URL 键入其 Web 浏览器时，X 同学笔记本 PC 需要知道 www. baidu. com 的 IP 地址。

8 ） X 同学笔记本 PC 上的操作系统因此生成一个 DNS 查询报文，将字符串 www. baidu. com 放入 DNS 报文中，该 DNS 报文则放置在 UDP 报文段中。该 UDP 报文段则被放入具有 IP 目的地址 120.69.56.46(在第 5 步中 DHCP ACK 返回的 DNS 服务器地址）和源 IP 地址 192.168.5.10 的 IP 数据报中。

9）X 同学笔记本 PC 则将包含 DNS 请求报文的数据报放入一个以太网帧中。该帧将发送(在链路层寻址）到 X 同学公司网络中的网关路由器。然而，即使 X 同学笔记本 PC 经过上述第 5 步中的 DHCP ACK 报文知道了公司网关路由器的 IP 地址（192.168.5.1），但仍不知道该网关路由器的 MAC 地址（尽管网关路由器和 DHCP 服务器是同一个路由器，但是 X 同学笔记本 PC 并不知道，而且实际生活中，DHCP 服务器和网关路由器可能是分开的）。为了获得该网关路由器的 MAC 地址，X 同学笔记本 PC 将需要使用 ARP 协议。

名称解释：ARP 协议，地址解析（Address Resolution）协议，用以在同一个子网内网络设备在网络层地址（如 IP 地址）和链路层地址（即 MAC 地址)之间的转换，每台主机或路由器在其内存中具有一个 ARP 表(ARP table），这张表包含 IP 地址到 MAC 地址的映射关系。RARP 以与 ARP 相反的方式工作，RARP 发出要反向解析的物理地址并希望返回其对应的 IP 地址。

10） X 同学笔记本 PC 生成一个具有目的 IP 地址 192.168.5.1（默认网关）的 ARP 查询报文,将该 ARP 报文放置在一个具有广播目的地址(FF:FF:FF:FF:FF:FF）的以太网帧中，并向交换机发送该以太网帧，交换机将该帧交付给所有连接的设备，包括网关路由器。

11）网关路由器在通往公司网络的接口上接收到包含该 ARP 查询报文的帧，发现在 ARP 报文中目标 IP 地址 192.168.5.1 匹配其接口的 IP 地址。网关路由器因此准备一个 ARP 回答，报文中说明了本机的 MAC 地址 22:33:44:55:66:11 对应 IP 地址 192.168.5.1。它将 ARP 回答放在一个以太网帧中，其目的地址为 11:22:33:44:55:66 (X 同学笔记本 PC），并向交换机发送该帧，再由交换机将帧交付给 X 同学笔记本 PC。

12）X 同学笔记本 PC 接收包含 ARP 回答报文的帧，并从 ARP 回答报文中抽取网关路由器的 MAC 地址（22:33:44:55:66:11）。

13）X 同学笔记本 PC 现在能够使包含 DNS 查询的以太网帧寻址到网关路由器的 MAC 地址。在该帧中的 IP 数据报具有 IP 目的地址 120.69.56.46 (DNS 服务

器），而该帧具有目的 mac 地址 22:33:44:55:66:11（网关路由器），X 同学笔记本 PC 向交换机发送该帧,交换机将该帧交付给网关路由器。

名词解释：NAT 网络地址转换

上面交付的数据帧中，IP 数据报内的源地址是 192.168.5.10（X 同学笔记本 IP），但是这不是一个 Internet 使用的 IP 地址，是一个私有 IP 地址，在局域网中使用的 IP 地址（私有 IP 地址包括：10.0.0.0 ~ 10.255.255.255、172.16.0.0 ~ 172.31.255.255、192.168.0.0 ~ 192.168.255.255）。

于是网关路由器会进行 NAT（网络地址转换），简单来说，NAT 路由器收到运输层报文，为该数据报生成一个新的源端口号，假设为 5002，将源 IP（192.168.5.10）替代为路由器广域网一侧接口的 IP 地址 120.68.34.89，且将报文中本来的源端口更换为新端口 5002。当生成一个新的源端口号时，并在路由器的 NAT 转换表进行记录。

应答服务器并不知道请求数据报已被 NAT 路由器进行了改装，它会发回一个响应报文，其目的地址是 NAT 路由器的 IP 地址，其目的端口是 5002。当该报文到达 NAT 路由器时，路由器使用目的 IP 地址与目的端口号从 NAT 转换表中检索出 X 同学笔记本 IP 地址（192.168.5.10）和发出 DNS 报文时使用的端口号。于是，路由器重写该数据报的目的 IP 地址与目的端口号，并向 X 同学笔记本转发该数据报。

注意，后面的描述中，经由网关路由器在 Internet 和局域网内交换的报文都默认进行了 NAT。

准备：域内路由选择到 DNS

14） 网关路由器接收该帧并抽取包含 DNS 查询的 IP 数据报。路由器查找该数据报的目的地址（120.69.56.46），并根据其转发表决定该数据报应当发送到四川电信网络中 R1 路由器。IP 数据报放置在链路层帧中，并根据寻找出来的链路发送。

15）在四川电信网络中 R1 路由器接收到该帧，抽取 IP 数据报，检查该数据报的目的地址（120.69.56.46），并根据其转发表确定出接口，经过该接口朝着 DNS 服务器转发数据报。

16）最终包含 DNS 查询的 IP 数据报到达了本地电信 DNS 服务器。DNS 服务器抽取出 DNS 查询报文，在它的 DNS 数据库中查找名字 www. baidu. com，找到包含对应 www. baidu. com 的 IP 地址(14.215.177.38）的 DNS 源记录。该 DNS 服务器形成了一个包含这种主机名到 IP 地址映射的 DNS 回答报文，将该 DNS 回答报文放入 UDP 报文段中，该数据报将通过四川电信网络反向转发到公司的路由器，并从这里经过以太网交换机到 X 同学笔记本 PC。

17 ） X 同学笔记本 PC 从 DNS 报文抽取出服务器 www. baidu. com 的 IP 地址。最终，在大量工作后，X 同学笔记本 PC 此时准备访问 www. baidu. com 服务器。

终于可以上网了：TCP 和 HTTP

18）既然 X 同学笔记本 PC 有了 www. baidu. com 的 IP 地址，它能够生成 TCP 套接字），该套接字将用于向 www. baidu. com 发送 HTTP GET 报文。

在 X 同学笔记本 PC 中的 TCP 必须首先与 www. baidu. com 中的 TCP 执行三次握手。X 同学笔记本 PC 因此首先生成一个具有目的端口 80（针对 HTTP 的）的 TCP SYN 报文段，将该 TCP 报文段放置在具有目的 IP 地址 14.215.177.38 (www. baidu. com）的 IP 数据报中，将该数据报放置在目的 MAC 地址为 22:33:44:55:66:11(网关路由器）的帧中，并向交换机发送该帧。

19）在公司网络、四川电信网络、北京电信网络和百度网络中的路由器朝着 www. baidu. com 转发包含 TCP SYN 的数据报，使用每台路由器中的转发表，如前面步骤 14~16 那样。

20）最终，包含 TCP SYN 的数据报到达 www. baidu. com。从数据报抽取出 TCP SYN 报文并分解到与端口 80 相联系的欢迎套接字。对于百度 HTTP 服务器和 X 同学笔记本 PC 之间的 TCP 连接生成一个连接套接字。产生一个 TCP SYNACK 报文段。

21）包含 TCP SYNACK 报文段的数据报通过百度、北京电信、四川电信和公司网络，最终到达 X 同学笔记本 PC 的以太网卡。数据报在操作系统中分解到步骤 18 生成的 TCP 套接字，从而进入连接状态。

22）借助于 X 同学笔记本 PC 上的套接字 socket，X 同学的浏览器生成包含要获取的 URL 的 HTTP GET 报文。HTTP GET 报文则写入套接字，其中 GET 报文成为一个 TCP 报文段的载荷。该 TCP 报文段放置进一个数据报中，并交付到 www. baidu. com，如前面步骤 18～20 所述。

23）在 www. baidu. com 的 HTTP 服务器从 TCP 套接字读取 HTTP GET 报文，生成一个 HTTP 响应报文，将请求的 Web 页内容放入 HTTP 响应体中，并将报文发送进 TCP 套接字中。

24）包含 HTTP 回答报文的数据报通过百度网络、北京电信、四川电信 和公司网络转发，到达 X 同学笔记本 PC。X 同学的 Web 浏览器程序从套接字读取 HTTP 响应，从 HTTP 响应体中抽取 Web 网页的 html，并最终显示了 Web 网页。