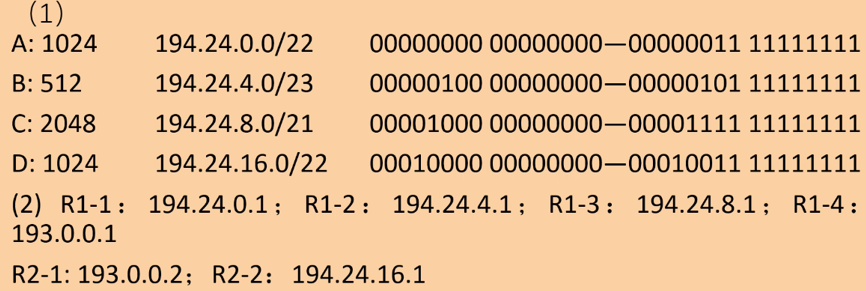
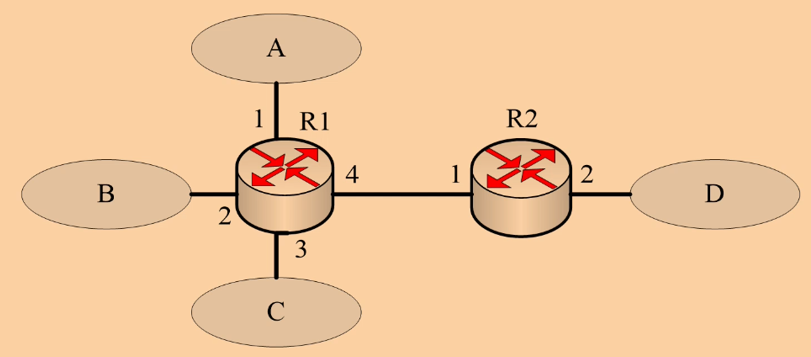
***第四章：网络层（数据平面）自顶向下：应用层、传输层、网络层、数据链路层、物理层***

网络层（主机间）的***转发功能***和***路由选择功能有别*：*转发***是分组在*单一*的路由器中从一条入链路到一条出链路之间的移动，*路由器本地功能、数据平面中唯一实现的、最重要的功能*。***路由***选择涉及多个路由器，它们经路由选择协议交互，控制平面决定分组从源到目的节点采用的路径，*网络范围、控制平面功能*。***概述：*转发和路由选择*网络层的作用***即将分组从一台发送主机移动到一台接收主机。某些分组交换机称***链路层交换机***，基于***链路层****字段*值做转发决定，链路层设备。其他分组交换机称***路由器***,基于***网络层****字段*值做转发决定，网络层设备。某些计算机网络中有***第三种重要的网络功能，即连接建立***（ATM、帧中继、MPLS）。**网络服务模型**定义了分组在发送/接收端系统间的端到端运输特性。在发送主机中，当传输层向网络层*传递分组时*，***可能***由***网络层提供的***服务***包括*①**确保交付**②**有时延上界的确保交付**③**有序分组交付**④**确保最小带宽**⑤**安全性服务。***因特网的网络层***提供了单一的***尽力而为服务（无任何保证）***。***虚电路和数据报网络网络层***能在*两台主机*间提供***无连接服务(数据报网络)***或***有连接服务(虚电路网络)***；与传输层类似（网络层的有连接服务以源/目的主机间的握手开始，无连接服务没有握手预备步骤）**也有差异：网络层**中，这些服务由网络层向传输层提供，*主机到主机*，在网络核心实现；**传输层**的服务由传输层向应用层提供，*进程到进程*，在网络边缘实现；网络层*不能同时提供*这两种服务(传输层可以)，仅提供有连接服务的称为***虚电路网络***，仅提供无连接服务的称为***数据报网络（如IP网络）***；**虚电路网络*因特网是数据报网络，ATM、帧中继是虚电路网络***。网络层连接被称为***虚电路***。虚电路是*端到端路径*，其***行为***：传输分组前建立虚电路，传输结束后拆除虚电路；所有分组携带虚电路标识而不是目的主机地址，均在该虚电路上传输（分组传输是保序的）；每个路由器对经过的连接维护状态，链路和路由器资源（带宽、缓存）**可**分配给虚电路（资源独享），从而虚电路能提供可预期的网络服务。***建立虚电路的本质***是先选好源主机到目的主机的路径，此后分组仅沿选好的路径传输，是否分配资源是可选的。***一条虚电路的组成：1.***源和目的主机之间的端到端路径（一系列链路和路由器）；***2.***VC（虚电路）号，沿着该路径的每段链路的一个号码（用于区分经过该链路的不同虚电路，仅有*本地意义*）；***3.***沿着该路径的每台路由器中的转发表项（进入端口，进入VC号，输出端口，输出VC号）。属于一条虚电路的分组将在它的***首部携带一个VC号***，因为一条虚电路在*每条链路上可能具有****不同****的VC号*，每台中间路由器必须*用一个新的VC号（从该路由器的转发表获得）替代*每个传输分组的VC号。无论何时跨越一台路由器***创建***一条新的虚电路，***转发表就增加***了一个新表项。无论何时***终止***一条虚电路，沿着该路径每个表中的相应项将被***删除***。即使没有VC号的转换，仍有必要维持连接状态信息，该信息将VC号和输出接口号联系起来。***一个分组沿着其路由在每条链路上不保持相同的VC号的原因：***减少分组首部中VC字段的长度，简化虚电路的建立，特别是具有多个VC号的路径，其上的每条链路能够独立于沿着该路径的其他链路选VC号，否则路由器将不得不交换并处理相当大量的报文以约定一个共同的VC号用于一次连接。***注意*：**在虚电路上传输分组时，分组只需携带VC号，不需携带目的地址。***虚电路建立***：发起呼叫-入呼叫-接受呼叫-呼叫连接-数据流开始-接收数据 **数据报网络**不建立网络层连接，分组携带*目的主机地址*，路由器按目的地址*转发*分组；路由器中的*转发表*记录目的地址到输出链路接口的映射；转发表由*路由算法*确定，约1～5min更新一次（转发表项的内容和当前路由器中是否存在相关地址的数据包没有关系），故***同一对主机****之间传输的一系列****分组****可能在通过网络时走****不同的路径****，从而可能****重排序***。路由器转发表用分组的目的地址的***前缀***与该表中的表项进行匹配，若有多个匹配则遵循***最长前缀匹配规则***。在数据报网络中的路由器*不维持连接状态信息*，但*在其转发表*中*维持了转发状态信息*。***路由器工作原理路由器的4个组成部分：输入端口****执行*将入物理链路与路由器连接的*物理层功能*、与入链路远端的数据链路层交互的*数据链路层功能*、通过查询转发表（存储在输入端口的内存中）决定路由器的输出端口的查找功能，又分为**基于目的地的转发**（按报文的目的地地址确定输出端口）和**广义转发**（按任意字段内容确定输出端口）。到达的分组通过路由器的交换结构被转发到输出端口；当**交换结构阻塞**时，分组需在此*排队*；***控制分组***（如携带路由选择协议信息的分组）从输入端口（这里的*端口*指路由器的*物理*输入和输出接口，而非软件端口）转发到*路由选择处理器*。***交换结构***将路由器的输入端口连接到其输岀端口，将数据包从输入端口队列转移到输出端口队列，完全包含在路由器中，是一个网络路由器中的网络。交换可以用许多方式完成：***经内存交换***使用多端口内存连接输入端口和输出端口，控制器在端口之间传输控制消息（如存储地址），性能和代价取决于存储接口数目，仅适合小容量系统。***经总线交换***用一根共享总线（包括地址线、数据线和控制线）将分组直接传送到输出端口，不需路由选择处理器干预，交换带宽受总线带宽限制。***经互联网络交换（交叉结构，crossbar，非阻塞）***能跨越交换结构并行发送多个分组，在输入端口与输出端口间建立内部专用电路，多对端口间可并行传输；交换速率：每秒转移的数据包数量，常用端口线路速度的倍数标识，N个数据端口理想情况下提供N倍线路速度的交换速率。***输出端口****存储*从交换结构接收的分组，并通过执行*必要的链路层*（执行链路层协议，封装）*和物理层功能*在输入链路上传输这些分组。若一条链路是双向的，输出端口通常与该链路的输入端口在同一线路卡上成对出现 ***路由选择处理器***执行*路由选择协议*，*维护路由选择表*以及连接的*链路状态信息*，并为路由器*计算转发表，*还执行网络管理功能。 一台路由器的输入、输出端口和交换结构***共同实现了转发功能***，几乎***总是用硬件实现。***这些转发功能有时总称为***路由器转发平面（ns尺度）*。**路由器的控制功能（执行路由选择协议、对上线或下线的连接链路进行响应、执行管理功能）在ms或s时间尺度上运行，故***路由器控制平面***常用***软件实现***并在路由选择处理器（CPU）上执行。**排队和丢包**在输入端口和输出端口处都能够形成分组队列。***排队的位置和程度***（或者在输入端口排队，或者在输出端口排队）将***取决于***流量负载、交换结构的相对速率和线路速率。若交换结构不能及时将输入端口的分组转移到输出端口，则***输入端口处形成排队***。***丢包：***当输入队列溢出时，发生丢包；若交换结构速率不小于端口速率的n（输入端口数量）倍则消除输入端口的排队，但路由器成本提高。多个输入端口同时向一个输出端口发送时，在***输出端口形成排队***。当输出队列满时，发生丢包；输出端口排队是不可避免的，增大输出队列可减少丢包的发生，但增加了内存消耗和分组延迟（延迟太大的分组还是会被重传，浪费资源），输出队列不是越长越好！**调度策略**FIFO/FSFC；优先调度（基于数据包携带的标识或者头部字段，例如，源/目的地址、端口号等为数据包分出优先级，先传输在最高优先级队列中排队的数据包）；轮询（round-robin）调度（将数据包分到多个队列中，循环检查每个队列，队列非空则传输一个数据包）；加权公平排队（WFQ，广义轮询，每次检查队列时，按权重分配带宽，传输不同队列中的数据包）**路由选择控制平面**网络范围的路由选择控制平面是***分布式的***：不同部分（如路由选择算法）执行在不同的路由器上，通过彼此发送控制报文进行交互。因特网路由器和路由选择算法以这种方式运行。***网际协议：因特网中的转发和编址***因特网编址和转发是***网际协议(IP)的重要组件***。***因特网的网络层有三个主要组件***：**IP协议**；**路由选择部分**（决定数据报从源到目的地所流经的路径，算出转发表）；报告数据报中的差错和对某些网络层信息请求进行响应的设施，因特网里是**互联网控制报文协议(ICMP)**。主机和路由器承担网络层功能。**数据报格式：**网络层分组被称为**数据报**。***IPv4数据报中的关键字段：版本号***4bit，规定数据报的IP协议版本，路由器通过查看版本号确定如何解释IP数据报。不同IP版本使用不同的数据报格式***。首部长度***4bit，***单位为4Byte，***IPv4数据报可包含一些可变数量的选项（位于数据报首部中），需确定数据报中载荷（例如在数据报中封装的传输层报文段）实际开始的地方。大多数IP数据报不包含选项,首部长度为20字节。***数据报长度***16bit，IP数据报的总长度（首部+数据），以**单个字节**计，理论最大长度为65535字节。***标识、标志、片偏移***与IP分片有关。IPv6不允许在路由器上对分组分片。***寿命(Time To Live/TTL)***8bit，确保数据报不会永远在网络中循环：每当一台路由器处理数据报时，该字段的值减1。若TTL减为0则该数据报被丢弃***。协议(upper layer)***8bit，指示IP数据报的数据部分应交给哪个特定的传输层协议。6-TCP，17-UDP。***首部校验和***16bit，将首部中的每2个字节当作一个数，用反码运算对这些数求和，该和的反码（称为因特网校验和）存放在校验和字段中；路由器对每个收到的IP数据报计算其首部校验和，不一致则检测出了差错，一般丢弃；注意，在每台路由器上**必须重新计算校验和并再次存放到原处**，因为TTL字段以及可能的选项字段会改变。***源和目的IP地址***（略）。***选项***允许IP首部被扩展，IPv6挪到其他位置。***数据（有效栽荷）***。如果数据报承载一个TCP报文段，则每个（无分片的）数据报共承载了总长40字节的首部(IP和TCP首部各20字节）和应用层报文。***链路层帧能承载的最大数据字节数***称为**MTU（最大传输单元）**，不同类型的链路的MTU可能不同，不过**IP数据报分片的负载=MTU-20Bytes**。传输过程中，***大于链路MTU的IP数据报可被路由器分片***，即将数据报划分为若干较小的数据报，每个数据块封装成一个独立的链路层帧传输。数据报在传输的过程中可被多次分片，但仅在**目的端系统**上**而非路由器**重组。***分片数目***=负载/（MTU-IP首部）。***分片的首部***取自原始数据报；**与分片有关的字段有**：***长度：*含首部**的报文长度，不大于MTU。***标识：***每个分片携带与原始数据报相同的标识，用于组装分片；***偏移量(offset)：单位为8Byte，***指示分片中数据的第一个字节在原始数据报载荷中的位置**(不考虑任何的IP首部，只考虑该分片的数据在原始报文数据段的位置)**，用来确定分片是否全部到达；***标志位(fragflag)：***最后一个分片为0，其余分片为1。***IPv6取消了路由器分片的功能：***源主机发送探测报文，确定路径上的最小MTU；源主机构造的数据报大小不超过最小MTU，路由器丢弃超大的数据报，并发送错误报告。**IPv4编址：*接口(interface)：***主机/路由器与物理链路的边界。路由器有多个接口，主机通常有一个或两个网络接口(如有线以太网和无线802.11)。***IP地址：***IP地址是一个32位的二进制数，每个主机和路由器接口对应自己的IP地址，通常用点分十进制数表示。在全球因特网中的每台主机和路由器上的接口均必须具有唯一的IP地址，接口的IP地址部分由子网决定。***子网(subnet)：***不需通过路由器就可以相互到达的网络接口构成一个子网，其内的接口有相同子网地址。IP编址为子网分配了形如x.y.z.m/h的子网地址，其中的/h记法称***子网掩码***，指示32bit中从最高侧起h个bit是子网地址。***如何确定子网？***将网络接口与主机/路由器分开，形成的每个分离的网络岛就是一个子网；路由器的每个端口连接一个子网，不同端口连接不同子网；因为路由器是在子网间转发数据包的设备，故子网内部通信不·通过路由器，子网之间通信必通过路由器。因特网采用***CIDR（Classless InterDomain Routing，无类别域间路由）***地址分配策略：地址的子网部分长度任意，低位bit可能含更多子网结构*【在CIDR之前，使用分类地址的方案：地址中子网地址长为8、16和24bit部分的网络被称为A、B、和C类网络；以广播地址：255.255.255.255地址作为目的地址可将报文送到子网中每个主机（****不能越过路由器****）】***子网网络地址**：子网部分不变，主机部分全0；**子网掩码**：网络部分全1，主机部分全0；子网中IP地址AND子网掩码=子网的网络地址，子网部分bit数称掩码长度。**广播地址**：网络部分不变，主机部分全1。目的地为广播地址的报文能被子网中所有接口接收。**最小用户地址**（子网内可分配给主机的最小地址）：网络部分不变，主机部分最小非全0。**最大用户地址**（子网内可以分配给主机的最大地址）：网络部分不变，主机部分最大非全1。按照实际需要的地址数量分配地址空间，提高地址使用效率；允许将若干条转发表项进行聚合，减小转发表规模。***按照实际需要分配地址：***若一个网络需要2000个地址，可为其分配一个具有2048个连续地址的地址块，这些地址的前21位必须相同，从而可将其视作有21位子网地址的网络。**CIDR地址分配原则：**地址块的长度L必须是2的幂次；所有地址的前（）位必须相同。如何利用最长匹配原则挖掉地址空间？例：200-23-00010000-00000000表示为200.23.16.0/21，200-23-00011000-00000000表示为200.23.24.0/24，那么从200-23-00011001-00000000（200.23.25.0）到200-23-00011111-1111111（200.23.63.255）怎么表示？从200.23.16.0/21地址块中把200.23.24.0/24地址块利用最长前缀匹配规则拿掉即可：因为11001000-00010111-00011\*\*\*-\*\*\*\*\*\*\*\*表示为200.23.24.0/21，增加一个有效前缀更长的转发表项200.23.24.0/**24**即满足需求（24>21，假如格式符合，会优先匹配这个端口，注意**最长**前缀匹配保证数据报去向转发表中**前缀匹配最长**的端口。用**更长的前缀**可以从大地址块中**挖掉**小地址块）。***主机/路由器如何获得 IP地址？路由器：***管理员手动配置路由器各接口的IP地址。***主机：***也可手动配置（对服务器），一般使用**动态主机配置协议DHCP**来分配IP地址（网络管理员可配置DHCP以使某给定主机每次与网络连接时得到相同的IP地址或一个临时的IP地址，故主机每次与网络连接时的IP地址或不同）。***DHCP目标：***允许主机加入网络时自动获取配置信息。DHCP允许主机获取IP地址、子网掩码、第一跳路由器（即默认网关）地址、本地DNS服务器等信息。***使用DHCP的好处：***免去手工配置的麻烦（即插即用）；可用少量的IP地址服务较多的客户（地址重用）。DHCP是一个***客户-服务器模式***的应用协议，**子网中**应有一个DHCP服务器或一个DHCP中继代理。***DHCP为新到达主机分配IP地址的4步：1.***主机广播“DHCP discover”报文（src:0.0.0.0,68-dest:255.255.255.255,67-yiaddr:0.0.0.0-transaction ID:114）寻找子网中的DHCP服务器***2.***DHCP服务器也用 “DHCP offer”报文（src:223.1.2.5,67-dest:255.255.255.255,68-yiaddrr:223.1.2.4-transaction ID:114-Lifetime:1145secs）广播响应，给出**可用的IP地址**、**事务ID**、**网络掩码**以及**IP地址租用期**。***3.***主机**选择一个**DHCP服务器，并向选中的服务器用“DHCP request”报文（src:0.0.0.0,68-dest:255.255.255.255,67-yiaddrr:223.1.2.4-transaction ID:115-Lifetime:1145secs）请求IP地址并回显配置参数。***4.***DHCP服务器用“DHCP ack”报文（src:223.1.2.5,67-dest:255.255.255.255,68-yiaddrr:223.1.2.4-transaction ID:115-Lifetime:1145 secs）发送IP地址，服务器响应客户的请求，确认所要求的参数。DHCP服务器使用**UDP端口67**，客户使用**UDP端口68。*DHCP不足：****节*点移动时不能维持与远程应用间的TCP连接，可通过移动IP解决。**网络地址翻译（NAT）：*目的：***使用一个公用IP地址，支持本地网络的多个用户同时上网。NAT使能路由器对外界的行为如同一个具有单一IP地址的单一设备，其可以做到仅为公共可访问的节点分配公用IP地址（减少了需要的公用IP地址数），而网络内部节点分配保留的IP地址（如10.0.0.0/24）。使用NAT后，本地网络上改变设备的IP地址无须对外通知，更换ISP时无需改变设备的IP地址，本地网络内部的设备无法从外面寻址和访问（对外不可见，更安全)。***NAT实现：***将***向外数据报中***的（**源**IP地址，**源**端口号）替换为（NAT路由器的IP地址，NAT端口号）；在***NAT转换表中***记录每个（源IP地址，源端口号）与（NAT的IP地址，NAT端口号）的转换关系；将***向内数据报***中的（**目的**IP地址，**目的**端口号），**查找NAT转换表中的记录**，替换为原来的（“源”IP地址，“源”端口号）【有来有回】。***端口号字段长16bit，所以一个NAT IP地址可同时支持65535个对外连接。NAT的使用争议：***路由器应当只处理网络层以下的包头（端口号在传输层）；违反端到端原则（节点介入修改IP地址和端口号）；***NAT妨碍P2P应用：***NAT只允许内部主动发起的通信，位于NAT后面的主机对外不可见，但P2P应用要求任一对等方可以向任何其它（参与的）对等方发起通信。***可以使用NAT穿越：***静态配置NAT表，总将某端口的连接请求转发到某个内网端口，或***使用UPnP和IGD实现：***内网主机可以主动获得NAT路由器的外网地址(138.76.29.7)，以及请求NAT路由器在(内网IP地址，内网主机端口号)和(NAT地址，公网端口号)之间建立映射表项，自动实现端口映射。 **IPv6：*最初的动机：***IPv4地址耗尽；***进一步的动机：***使用更便于路由器处理转发的报文头部格式，在头部中添加支持服务质量差异的字段。IPv6与IPv4不兼容，但与其它所有因特网协议都兼容。***IPv6地址：***128位，使用冒号十六进制表示，每16位以十六进制的形式写成一组，组之间用冒号分隔，如“8000: 0: 0: 0: 0123: 4567: 89AB: CDEF”。***IPv6数据报格式：***以一个40字节的基本头开始，然后是数据。基本头包括***PRI(或traffic class)：***用于发送方定义数据报的优先级，路由器发现网络拥塞时，按优先级从低到高的顺序丢弃包。***流标签(flow label)：*流(flow)**是具有相同传输特性（源/目的、优先级、选项等）、并要求相同处理（使用相同的路径和资源、有相同的服务质量和安全要求等）的一系列数据包；流由源地址和流标签（flow label）标识，流标签由发送方分配，不支持流的节点忽略该域。***IPv6包格式***不同于IPv4之处包括移除校验和（节约路由器处理报文的时间）、将头部的选项移动到“下一个头部”字段指示（IPv6基本头**固定40字节**）等，拿掉了很多东西，整体上是在做减法。***从IPv4过渡到IPv6：***无法同时升级所有路由器，怎样让IPv4和IPv6在网络中共存→***让IPv6数据包穿越IPv4网络：隧道技术：***把IPv6报文作为负载（**增加**一个IPV4报头）送入IPv4网，由IPv4报文携带并被IPv4路由器处理转发，目的边界路由器取出IPv6包继续传输。注意IPv6包在IPv4网络中传输时未被IPv4路由器检查和处理，报头中TTL未减1。优点：保留原始数据报全部信息。外层IPv4数据报中源IP为**源IPv6路由器**的IPv4地址，目的IP为**目的IPv6路由器**的IPv4地址。本质是封装数据包。***双协议栈方案：***双栈节点（如一些路由器）同时具备收发IPv4和IPv6报文的能力，在将数据报传递给IPv4路由器前，将IPv6报头转换成IPv4报头。缺点：在6-4-6报文转换中一些IPv6字段丢失。

***第五章：网络层（控制平面）***

***两类控制平面实施方案***传统的路由算法（在每个路由器上实现，每个路由器都有单独的路由算法模块，它们相互交互，计算出各自的转发表，一起构成控制平面）和软件定义网络（SDN，逻辑上集中实现，单独的（远程）控制器与路由器本地的控制代理（CA）交互）。***路由选择算法目的：***在由路由器构成的网络上，选择从发送端主机到接收端主机的“好”（“代价”最小、“最快”、“最不拥塞”）的路径（数据包从发送端主机到接收端主机经过的路由器序列）。***分类：全局/局部：***所有路由器均有全网拓扑和链路代价的全部信息**/**路由器只知道直接相连的邻居以及到邻居的链路代价通过相邻路由器之间反复迭代交换信息并计算路由；***静态/动态：***路由随时间缓慢变化，通常人工调整**/**路由器根据网络流量负载或拓扑的变化自动更新路由。**链路状态（LS）路由选择算法：*全局算法***，***核心思想***为**Dijkstra**算法。***可能的问题：选路震荡***（可能出现在任何使用拥塞或基于时延的链路测度算法中）。***解决方案：***确保并非所有路由器都同时运行 LS 算法。**距离向量（DV）路由选择算法：是*局部、迭代、异步、分布式***算法。***核心思想：***通过求解***Bellman-Ford***方程，求解任意两节点间的最小代价路径。***步骤：***对**每个**节点：**①测量**到中所有节点的代价（二者不相邻则记为+），令其为，这些代价构成的距离矢量；其他节点“还未发来的”距离向量的所有分量均初始化为+。**②**对于的每个邻居和中的所有节点，记，并将的距离矢量发送给。**③**（***迭代、异步***）**只有当**到邻节点的**链路代价变化或**邻节点**距离矢量发生改变**时，对所有节点，才利用**B-F方程**重新计算**自己**的距离矢量：（*别人发过来的距离矢量照单全收就行*）。**④**(***分布式***)假如中的任何分量相对上一轮有变化，将发送到的每个邻居节点仅在发现距离矢量有变化时通知其邻居。***链路代价变化：好消息传播快，坏消息传播慢***（**路由选择环路**和**无穷计数问题**）***LS算法和DV算法的比较：链路状态LS：***链路状态信息在全网传播；节点仅计算自己的转发表（LS更健壮）；***收敛速度：***个报文，次计算。***距离矢量DV：***距离矢量仅在发生变化时向邻居发送；节点传播的信息可能不正确：邻居的距离矢量是“道听途说”的；会造成错误扩散；***收敛速度：***较慢，还可能出现路由环路、无穷计算问题。**路由可扩展性、域间和域内路由：**平面结构的网络不具可扩展性：路由器数目多，路由表规模大，信息交换开销足以瘫痪网络，所以网络实际上**不是平面**的；因特网是网络构成的网络，每个网络都希望控制自己网络上的路由。可以通过将路由器组织进***自治系统(AS)***来解决这两个问题。AS由一组**通常**处在相同管理控制下的路由器组成（某些一级ISP在其整个网络中使用**一个**庞大的AS，有的ISP将它们的ISP拆分为**数十个互联**的AS）。每个AS被赋予一个AS编号，由ICANN全局**唯一**分配。***域内(Intra-AS)路由***是**同一AS内部**主机和路由器的路由，**同一AS中**的路由器**必须**运行**相同**的**选路协议（称Intra-AS选路协议）**，**不同AS中**的路由器可以运行**不同的Intra-AS选路协议**。***网关路由器***：位于AS**边界**，通过链路与**其它AS的网关路由器直连**的路由器。***域间(Inter-AS)路由***是AS之间的路由，由网关路由器执行（网关路由器同时**也执行域内路由**）。网关路由器之间运行**Inter-AS选路协议**，所有AS必须运行**相同的**Inter-AS选路协议。**域间路由和域内路由算法共同生成转发表：域内路由决定AS内部目的地的转发表项，域间路由决定外部目的地的转发表项。*Inter-AS的任务：***AS1内部的路由器1a需发送数据报到AS1外，但1a应当具体向AS1的哪个网关路由器转发？AS1的网关路由器必须：**①**了解哪些目的网络通过自己可达**②**将可达性信息传播到AS1内部的所有路由器。**因特网中自治系统内部的路由选择：RIP（应用层协议）**RIP采用***距离矢量***选路算法，**是距离矢量协议**。***距离（代价）***：用跳数(hop count)衡量。***跳(hop)：***相邻路由器之间的链路为一跳（相当于每条链路代价1）。***路径的跳数：***从源路由器到目的子网（含）经过的子网数量。限定一条路径的最大代价和AS的直径为15跳。***RIP通告（RIP响应消息）：***距离向量：路由器到AS内各个子网的最短路径的跳数（估计值）；构造RIP响应报文：距离向量封装在RIP响应报文中传输，称为RIP通告，每个报文携带一个目的子网列表（最多包含25个子网），以及到每个目的子网的最短距离；**发送RIP响应报文：**RIP报文封装在UDP报文中发送，使用**UDP端口520**（RIP是一个**应用层协议**），相邻路由器之间大约每30s通过***RIP通告***交换一次RIP响应报文，**更新RIP路由表中可抵达的子网以及到各个子网的距离**。每台路由器维护一张称为**路由选择表**的RIP表，包括该路由器的距离向量和该路由器的转发表。***RIP链路失效与恢复：***若超过180s未收到某个邻居的RIP通告，认为该邻居不可达：令通过该邻居的路径失效（距离设为16），发送RIP通告；采用***毒性逆转***解决计数至无穷问题：如果Z选择经由Y到达X，则Z向Y通告距离向量时，将设为无穷大，即阻止Y使用这条路由。若选路表中到目的网络x的路由是A通告的，则向A通告该路由时，到x的距离设为16（阻止A使用这条路由）。**因特网中自治系统内部的路由选择OSPF：**OSPF采用***链路状态选路算法***，**是链路状态协议**。路由器利用洪泛链路状态信息和Dijkstra算法，计算以本路由器为根到其他所有**子网**的最短路径树，***链路代价***由管理员配置（将所有链路开销设为1可实现最少跳数路由选择，将链路权值按与链路容量成反比来设置则不鼓励流量使用低带宽链路，反映管理员的选路策略）。***OSPF分组：***OSPF 协议定义了5种分组类型，分别用于探测邻居、通告链路状态等。OSPF通告包含在直接由IP承载的OSPF报文中，OSPF分组被封装在**IP包中传输，协议号为89**。路由器周期性地（至少30min一次）或在链路状态改变时发送***OSPF链路通告，广播链路状态信息***。***OSPF协议负责：***OSPF协议必须实现可靠报文传输、链路状态广播等功能，还要检查链路正在运行（通过向相连的邻居发送HELLO报文），并允许OSPF路由器获得相邻路由器的网络范围链路状态的数据库（路由器根据收到的链路通告分组构造***链路状态数据库***。）。***OSPF最重要的优点是支持AS内部的分层选路。*分层的OSPF：**一个OSPF自治系统可以配置多个层次化的***区域(area)：***每个区域运行**自己的**OSPF链路状态路由选择算法，区域内的每台路由器都向**该区域内**的所有其他路由器广播其链路状态，进行泛洪链路状态通告；每个节点有详细的区域内拓扑，仅知道通向其它区域的（到本区域边界路由器的）最短路径。在每个区域内，一台或多台**区域边界路由器**负责为流向**本区域以外**的分组提供路由选择：将**本区域**的选路信息（子网及路径代价）综合，通告给其它区域（的区域边界路由器）、将收到的其它区域的选路信息（子网及路径代价）通告给本区域的区域内部路由器。一个AS中**只有一个**OSPF区域配置成**主干区域，构成本地区域和主干两个选路层次*。*主干区域（编号0）**总是包含本AS中的**所有区域边界路由器**（还可能包含一些非边界路由器，称为骨干区域路由器，它们在骨干区域内部运行OSPF），其主要作用是为该AS中其他区域之间的流量提供路由选择。***对于在AS中去往其它区域的分组***，要求其先转发到一个本地区域边界路由器（区域内路由选择），然后通过主干转发到位于目的区域的区域边界路由器，再转发到最终目的地。***OSPF的其他优点：*安全**（所有OSPF消息可经过（简单的和MD5的）鉴别认证，避免恶意入侵者发布虚假链路状态）；**允许多条等代价路径同时存在，可以同时使用多条费用相同的路径**（RIP协议仅计算一条），**对单播和多播路由选择的综合支持**（多播OSPF (MOSPF)使用和OSPF一样的网络拓扑信息）**比较RIP和OSPF：*RIP缺点：***不分区，管理网络规模小（直径不超过15跳）；只能用跳数作为路径代价；不支持多条等代价的路径；更新频繁。***OSPF缺点：***路由通告泛洪，代价高（链路状态路由协议固有问题，但RIP的缺点对应的都是OSPF的优点）。**自治系统间的路由选择：BGP（边界网关协议）BGP事实上是唯一的域间路由协议，确保了因特网是一个整体，是一种分布式和异步的协议。BGP只能知道应该“经过”哪个边界路由以及不同AS之间的路径，同一AS里的路径还要靠Intra-AS协议选择。**在BGP中，分组**并不**路由到一个特定的**目的地址**，而是路由到**CIDR化的前缀**，**每个前缀**表示一个子网或一个子网的集合。一台路由器的转发表具有形式为的表项，其中是一个前缀（如138.16.68/22），是该路由器的接口之一的接口号。BGP允许路由器完成以下任务：**①**从邻居AS获得前缀的可达性信息。特别是，BGP允许每个子网向因特网的**其余部分**（通过eBGP和iBGP）**通告自己的存在**。**②**确定到该前缀的“最好”路由。一台路由器可能知道**多条**到特定前缀的不同路由，路由器将本地运行一个BGP路由选择过程（使用经相邻的路由器获得的前缀可达性信息，基于策略以及可达性信息），选出最好的路由。**BGP会话：**两个BGP路由器通过***半永久的TCP连接（端口179）***交换BGP消息，通告通向各网段地址前缀的路径信息，包括**OPEN**：与远端的BGP路由器建立TCP连接并认证对方；**UPDATE**：通告一条新路径(或撤回一条路径)；**KEEPALIVE**：在没有交换UPDATES时保持连接；应答OPEN请求；**NOTIFICATION**：通知前一条消息中的错误并关闭连接。BGP是路径矢量协议，是**应用层协议**。每个AS的路由器可建立**两类连接**，执行两种任务。**eBGP连接（通常路由器物理直连）**是**不同AS**的**边界路由器**间建立的BGP会话，用于从相邻的AS获取子网可达性信息。**iBGP连接（并不总对应物理链路）**向**AS内所有**路由器传播外部的可达性信息，通过可达性信息和策略决定通向外部目的地子网的“好”的路径。若AS3的网关路由器3a向AS2的网关路由器2c通告路径，则**AS3向AS2承诺它可以转发目的地址在前缀内的报文**。每条**BGP路由包含3个组件**：**AS路径**（AS-PATH，所有到达目的地子网需经过的AS的**排列**，如表示**顺序**经过可以到达某子网；BGP路由器使用AS-PATH属性来检测和防止通告**环路**。如果一台路由器在路径列表中看到包含了它自己的AS，它将认为出现了环路，进而拒绝该通告。）、**下一跳**（NEXT-HOP，经过哪个AS内部路由器到达下一跳AS，是**AS-PATH起始的路由器接口**的IP地址）和**目的前缀**。网关路由器可能获得到同一目的子网的多条路径信息，需用BGP路由选择算法选择。当存在**多条**到目的地的路径时，基于以下原则选择路径：**①**本地策略选择路由**②**选择AS-PATH最短的**③**运用热土豆策略，选择到下一跳路由器代价最小的**④**其它。***热土豆路由：***选择**域内**路由**代价最小**的**区域边界路由器**，此时**不考虑域间**路由的代价（即使该路由将导致更多的AS跳数）。**在路由器转发表中增加AS外部目的地的步骤**：**①**从AS间协议学到经多个网关可达子网**②**使用来自AS内部协议的路由选择信息（OSPF），获知到达每个区域边界路由器的最低开销路径**③**按照**热土豆路由**，选择AS内具有**最小开销**的边界路由器**④**从转发表确定通往最低开销网关的接口，并在转发表中加入表项。在实践中，BGP使用了一种结合了热土豆路由选择的特点但更为复杂的算法，通过选择性通告实现选路策略。具体而言，有的时候AS选择采取不向其他AS通告所有连接的**策略**，可能是因为运营商只愿意承担顾客网络的流量，不希望承担**过境流量**，即运营商会选择不通告可能导致自己“赚不到钱”的连接。**比较Intra-AS和Inter-AS选路协议：**Intra-AS选路协议（如OSPF、RIP）用于在AS内部交换选路信息，关注路由性能，总使用某个路由测度（代价）选择到目的节点的最优路径，都是内部流量，无需域间路由的“策略”；Inter-AS选路协议（如BGP）用于在不同的AS间交换选路信息，主要依据策略而不是路由测度去路由，管理员希望通过策略控制哪些流量可以经过自己的网络转发，策略优先，可能导致选择次优路径（如热土豆路由）***为什么会有不同的AS间和AS内部路由选择协议：***该问题的答案涉及AS内与AS间路由选择目标的本质差异：**①策略**不同（AS间策略起主导作用，在AS内部选择路由中策略起着微不足道的作用）**②规模**不同（可扩展性是AS间的关键问题；AS内由于可以进一步划分成别的AS或如OSPF一样分层导致可扩展性不是关注的焦点）**③性能**（AS内关心，AS间路由选择面向**策略**，故不关心）。***选路算法和选路协议：***选路算法是选路协议的一部分；选路协议还包括路径代价的定义、报文格式、报文传输、报文处理、异常事件处理等具体实施方面的问题。**因特网控制报文协议（ICMP协议）：**主机或路由器使用***ICMP协议***传递网络层上的信息。**ICMP报文封装在IP包中传输：**ICMP报文可能需经过多个网络才能到达源节点，跨网络传输借助IP包。***ICMP与IP的关系：***ICMP报文是作为IP有效载荷承载的，故ICMP通常被认为是IP协议的一部分，但从体系结构上讲它是位于IP之上的。***ICMP报文格式：*type:**报文类型，共定义了15种；**code：**对某类型报文作进一步的区分；**Checksum：**ICMP报文的检查和；**标识符**；**序号**；**内容：**与报文类型有关。***Traceroute与ICMP：*Traceroute用ICMP报文实现**。为了判断源和目的地之间所有路由器的名字和地址，源主机中的Traceroute向目的地主机发送一系列普通的IP数据报，每个均携带一个具有不可达UDP端口号的UDP报文段。第一个数据报的TTL为1,第二个的TTL为2,第三个的TTL为3,依次类推。该源主机为每个数据报启动定时器。当第个数据报到达第台路由器时，第台路由器观察到该数据报的TTL正好过期。路由器根据IP协议规则丢弃该数据报并发送一个ICMP告警报文给源主机（类型11编码0），该报文包含路由器的名字和其IP地址。当该报文返回源主机时，源主机从定时器得到往返时延，从报文中得到第台路由器的名字与IP地址。**源主机何时停止发送UDP报文段？**当数据报最终到达目的主机时，因为该数据报包含了一个具有不可达端口号的UDP报文段，目的主机将向源发送一个端口不可达的ICMP报文（类型3编码3）。源主机一旦收到这个特别的ICMP报文即知道不需再发送探测分组了（标准的Traceroute程序实际上用相同的TTL发送3个一组的分组，因此Traceroute输出对每个TTL提供了3个结果）。ICMPv6除了重新组织现有的ICMP类型和编码定义外，还增加了新型IPv6功能所需的新类型和编码，包括“分组太大”类型和一个“未被承认的IPv6选项”差错编码。

***第六章 链路层：链路、接入网和局域网：***

***网络层和链路层的关系：网络层：选路：***路由器确定去往目的节点的下一跳***转发：***在路由器内部将数据报从输入端口转移到输出端口；***链路层：将数据报从一个节点传输到物理上直接项链相邻的下一个节点，***如：源主机→源路由器，路由器→下一跳路由器，目的路由器→目的主机。物理层：多种类型的传输媒体，传输原始比特流（无结构），容易产生传输错误***链路层概述：节点：***运行链路层协议的任何设备（包括主机、路由器、交换机和Wi-Fi接入点）。***链路：***沿着通信路径连接相邻节点的通信信道（有线链路、无线链路、LAN）。不同类型的链路使用不同的协议传输数据包（例：第一段链路以太网、第二段链路帧中继、第三段802.11）；不同的链路层协议提供不同的服务（例：一些协议确保可靠传输，另一些不行）。***帧：***链路层分组称为***链路层帧***。**链路层提供的服务：成帧（基本服务）：**发送：将数据报添加头部、尾部，封装到链路层帧中；接收：从原始比特流中提取出完整的帧；**链路接入**（广播链路需要）当多个节点共享单个广播链路时，协调各节点的发送行为；**差错检测（基本服务）**检测传输错误；**差错纠正（有些提供）**:检测并**纠正**传输错误（**不用重传**）；**可靠交付（部分协议提供）**通过确认、重传等机制确保接收节点正确收到每一个帧（停-等、GBN、SR），低误码率链路（如光纤、某些双绞线）上很少使用，高误码率链路（如无线链路）应当使用；**流控制:**调节发送速度，避免接收节点缓存溢出（提供可靠交付的链路层协议不需要专门的流量控制；不提供可靠交付的链路层协议需要流量控制机制）；***半双工/全双工*：半双工**同一时间只能单向传输数据；**全双工**双方可以同时收发数据。**为何在传输层和链路层都提供可靠传输？**链路层传输数据包误码率较高，可能还没到传输层就已经出现了错误。**链路层在何处实现：链路层主体部分在网络适配器（亦称网络接口卡，简称网卡）**中实现。网络适配器插入主机的总线，其**核心**是链路层控制器（通常是一个实现了许多链路层服务（如成帧、链路接入、差错检测等）的专用芯片），其许多功能是用硬件实现的。链路层和物理层一起实现。链路层的实现方式可以是硬件、固件或软件[***链路层由硬件和软件实现：***网卡中的控制器芯片：成帧、链路接入、检错、可靠交付、流量控制等；主机上的链路层软件：与网络层之间的接口，组装链路层寻址信息、激活控制器硬件、响应控制器中断。]***网络适配器之间的通信：发送端：***将网络层报文封装为帧；添加差错校验、可靠传输、流控制等需要的协议字段；***接收端：***实施差错校验、可靠传输、流控制等；提取报文送交网络层。***差错检测和纠正技术：差错检测的实施：***发送端对要保护的数据（包括帧头字段）生成校验位（差错检测和纠正比特），添加在帧头中；接收端对收到的数据计算校验位，根据和判定是否与相同。检错不是100%有效，某些差错可能无法被检出，但概率很低。越大检错和纠错的效果越好。***传输出错的类型：***单bit差错；突发差错（差错以突发的形式聚集在一起）。***差错控制编码的类型：检错码***：只能检测出传输错误的编码，**不能确定出错位置**，通常与反馈重传机制结合进行差错恢复；***纠错码：***能够**确定错误位置**并**自行纠正**的编码。**码字（codeword）**：由bit的数据加上比特的冗余位（校验位）构成，***检错能力***：为**检测**出bit中所有错误，编码集的海明距离至少应为***纠错能力***：为**纠正**bit中所有错误，编码集的海明距离至少应为**奇偶校验：*单比特奇偶校验：***发送方向在bit信息中包含个附加bit，使得这bit中的总数是偶数，接收方可检测出奇数个bit翻转导致的错误，无法检测偶数个bit翻转导致的错误；在**突发差错**情况下的检错率为50%，编码集海明距离为2。***二维奇偶校验：***将bit信息的bit排列成矩阵，可***检测****2-bit错*和***纠正****单bit错*（包括校验位错），编码集海明距离为3，有利于检测突发错误。**因特网校验和(Internet checksum)：发送端**将整个分段内容（包括头部）视为16-bit整数序列，对分段内容求和取反、溢出回卷得到校验和，将计算出的校验和写入校验和字段；**接收端**基于接收到的分段内容（包括校验和字段）求和，计算结果包含则检测到错误，全则未检出错误，但是仍然可能有错。优：校验和分组开销相对小。劣：与CRC相比差错保护功能相对弱。**循环冗余校验（CRC）：**CRC亦称***多项式编码***，其将bit串看成是一元多项式的系数，检错能力强，可用硬件实现，应用最广泛。**生成多项式：**双方确定用来**计算**的一个bit多项式，次数最高项系数为。**冗余多项式：**由个冗余bit为系数构成的多项式。**计算方法：*的余式，CRC中加减法运算均定义为异或操作***。**检验方法：**若生成的编码除以的余式为0则判定传输正确。例：取，对计算CRC码。**①**，在后面加个0，得到**②**除以的余式为（余数最右边位就是，不够则高位补零)**③**在右边接上得到，此为发送的报文**④**若接收端收到的，因为其除以的余式为，故有传输错误。***循环码性质：***⑴由任何多于一项的生成多项式产生的循环码能够检测所有单bit错误；⑵每个被除尽的多项式都具有偶数项。若生成多项式具有偶数项，则由它产生的编码能检测所有奇数bit错误；⑶若码长不大于生成多项式的指数(即)，则由产生的码能检测所有单个和两个错码；的指数：是使能除尽的最小正整数；⑷若码长不大于的指数，则由生成多项式产生的码能**检测**(但**不能纠正**)所有单个、两个及三个错误；⑸由次多项式产生的任一循环码，能检测所有长度不超过的突发错误；⑹ 长度为的突发错误中：若，则不能检测部分占；若，则不能检测部分占。bit的支持检测出最长连续bit的错误；**连续错个bit是一种突发性错误，CRC可以检测突发性错误。*为什么传输层使用校验和&&链路层使用CRC？*传输层**常在主机中作为OS的一部分用软件实现，所以传输层差错检测也是用**软件**实现的，采用简单而快速的差错检测方案更重要；**链路层**的差错检测在适配器中用专用的**硬件**实现，它能够快速执行更复杂的CRC操作。***多路访问链路和协议：两种类型的链路：*点到点链路：**仅连接了一个发送方和一个接收方的链路，如拨号上网的PPP链路、以太网交换机和主机之间的点对点链路）；一条全双工链路可看成由两条单工链路组成。**广播链路：**连接了许多发送和接收节点的单一共享广播信道，任何一个节点发送的数据可被链路上的其它节点接收到，如老式以太网和802.11无线局域网。***多接入协议冲突（collision）：***在广播链路上，若两个或多个节点同时发送数据，则发送的信号会彼此干扰，导致接收失败。**冲突域：**共享同一条广播链路的主机集合；任何一个主机发送的帧（各种帧），可被冲突域中的其它主机接收到。**广播域**：广播帧能够到达的主机集合。**多接入协议**，也称***媒体接入控制（MAC）协议****，*规定了节点如何共享信道（谁可以发送）。***理想的多接入协议：***在带宽为bps的广播信道上：**①**只有一个节点发送时，它应能以速率发送（信道利用率高）**②**当有个节点发送时，每个节点应能以的平均速率发送（公平性好、信道利用率高）**③**协议是**完全分布式**的：不需一个特殊的节点来协调发送（健壮性好），不需时钟或时隙同步（不需要额外的机制）**④**简单（实现和运行开销小）。***MAC协议的分类：信道划分：***将信道划分为若干子信道，每个节点固定分配一个子信道，不会发生冲突，关注公平性，但轻负载时信道利用率不高；***随机接入（竞争）：***不划分信道，每个节点自行决定何时发送，出现冲突后设法解决，轻负载时信道利用率高，重负载时冲突严重；***轮流使用信道：***不划分信道，有数据的节点轮流发送，不会出现冲突，如果数据量大，轮到的节点可以多传输一会儿。信道利用率介于前两种方法之间，引入额外机制。**信道划分协议：*TDMA（时分多址）：***将信道的使用时间划分成帧，每个节点在帧中被分配一个固定长度的***时隙（1个时间对应N个时隙）***，每个时隙可以发送一个分组，节点只能在分配给自己的时隙内发送，若节点不发送，其时隙轮空。***FDMA（频分多址）：***将信道频谱划分为若干子频带，每个节点被分配一个固定的子频带（**带宽**），若节点不发送，其子频带空闲。***TDMA和FDMA的优劣：***消除碰撞而且非常公平，**但**节点被限制于 bps的平均速率，必须总是等待它的轮次。***CDMA（码分多址）：***对每个节点分配不同的编码，节点用它来对它发送的数据进行编码。CDMA网络允许不同节点同时传输信息，并且它们各自的接收方仍能正确接收它们编码的数据（假设接收方知道发送方的编码），屏蔽其他节点的干扰传输。**CDMA允许所有节点同时使用整个信道！随机接入的MAC协议：*随机接入的基本思想：***当节点有数据要发送时，以信道速率R发送，发送前不需要协调；若发送某帧时发生冲突，则节点等待一个随机时延再重发该帧。**随机接入MAC协议**规定了如何检测冲突，以及如何从冲突中恢复。***时隙（Slotted）ALOHA：***假设所有帧长bit、时间被划分为长度为秒的时隙（即每个时隙传帧）、节点只在时隙开始时发送帧、节点是**时钟同步**的（知道时隙**何时开始**）、所有节点均可在**时隙结束前**检测到该时隙中是否有冲突发生。**操作：**节点从上层收到数据后，在下一个时隙发送；若时隙结束前未检测到冲突，节点可在下一个时隙发送新的帧；若检测到冲突，节点在随后的每一个时隙中以概率重传，直至发送成功。**优点：**单个活跃节点可以信道速率全速发送数据；分布式（节点仅需要时隙同步，可自行决定什么时候发送），简单。**缺点：**发生冲突的时隙被浪费了；由于概率重传，有些时隙被闲置；需要时钟同步。**时隙多路访问协议的效率：**当有大量的活跃节点且每个节点总有大量的帧要发送时，长期运行过程中成功时隙所占的比例。**时隙ALOHA的效率：**假设有个活跃节点，每个节点在每个时隙开始时以概率发送；节点在一个时隙中发送成功的概率为；该时隙中有节点发送成功的概率为；令最大的概率可求得为(求导)；代入得；再令即得到最大效率为。***纯（无时隙）ALOHA：基本思想：“想发就发”，***取消同步时钟。当某帧到达节点（即一个网络层数据报在发送节点从网络层传递下来）时，节点**立刻**将该帧**完整地**传输进广播信道。如果该帧传输时**经历了碰撞**，该节点在完全传输完它的碰撞帧**之**后立即以概率**重传**该帧；若**没有经历碰撞**，则该节点**等待一个帧传输时间**，再以概率传输其他帧（即以概率保持空闲）。**发生冲突的情形：**在时刻发送的帧与在时段内发送的所有帧冲突。**纯ALOHA的效率：**P(给定节点发送成功)=P(节点发送)P(无其它节点在内发送)P(无其它节点在内发送)=；求出令节点发送成功概率最大的，并令，得最大效率为。**载波侦听多接入（CSMA）MAC协议两个重要规则：载波侦听：发送前监听信道**，信道空闲则发送**整个帧**，信道忙则推迟发送（等待至少一段时间，没有传输再发送）；**碰撞检测：**传输节点在传输时一直侦听用到的信道，如果检测到另一个节点正在传输干扰帧，则**立即停止发送剩余的部分**，在重复“侦听－当空闲时传输”循环之前等待一段**随机时间**。即使这样，***冲突仍可能发生：*①**由于存在端到端信道传播时延，节点可能没有监听到其它节点正在发送，认为信道是空闲的**②**即使忽略传输延迟，当多个节点同时发现信道由忙变为空闲、并都决定立即发送时，仍会发生冲突。信道传播时延在决定CSMA性能方面起关键作用。***（没有碰撞检测的）载波侦听多接入（CSMA）：*仅执行载波侦听**，**检测到碰撞仍然继续传输**，一旦发生碰撞干扰则整个帧的传输时间被浪费。***具有碰撞检测(Collision Detection)的载波侦听多接入CSMA/CD：*同时执行碰撞检测**，**检测到碰撞就立刻终止传输**，浪费时间更少。**以太网采用CSMA/CD协议：①**网卡从网络层接收数据报，构造以太帧**②**若网卡监听到信道空闲，立即发送帧；若信道忙则持续监听，直至发现信道空闲再发送帧**③**若网卡发送整个帧而没有检测到冲突，认为发送成功**④**若网卡在传输过程中检测到冲突，立即停止发送帧**⑤**中止传输后等待一个**随机时间间隔**然后返回**②**。***希望时间间隔满足：***当碰撞节点数较少时，时间间隔较短；当碰撞节点数较多时，时间间隔较长***。*二进制指数后退算法*：***以太网采用该算法确定随机时间间隔。当传输一个给定帧时，在该帧连续经历了次碰撞后，节点从{}中随机选择一个值，等待个发送1bit的时间后返回上述算法的**②**。时，算法能够取的最大值为1023。每次适配器准备传输一个新的帧时均要运行CSMA/CD算法，不考虑近期过去的时间内可能已经发生的任何碰撞。因此，当几个其他适配器处于指数后退状态时，有可能一个具有新帧的节点能够插入一次成功的传输。**指数回退的目的：**根据网络负载调整重传时间，负载越重（冲突次数越多），重传时间的选择范围越大，再次发生冲突的可能性越小。***CSMA/CD的效率：***当有**大量的活跃节点**，且每个节点有**大量的帧**要发送时，帧在信道中**无碰撞地传输的那部分时间**在长期运行时间中**所占的份额**。记为信号在任意两个节点之间传播延迟的最大值，为最大帧的传输时延，则效率为。**在以下情况下，以太网的效率趋近于1**：t\_prop0，或t\_trans。**结论：**应控制以太网的规模。**轮流MAC协议：ALOHA和CSMA协议具备理想多路访问协议的特性①，但不具备特性②（带宽均分）。*轮询协议：***节点之一被指定为主节点，主节点能够通过观察信道上是否缺乏信号来获知一个节点何时完成了帧的发送。主节点以循环的方式轮询每个节点：主节点向节点1发送一个报文，告诉它能够传输帧的最多数量。在节点1传输了某些帧后，主节点再告诉节点2其能够传输帧的最多数量。**优点：**消除了碰撞和空时隙，效率高；**缺点：**存在**轮询时延**，即通知一个节点“它可以传输”所需的时间，例如，如果只有一个节点是活跃的，那么这个节点将以小于bps的速率传输，因为每次活跃节点发送了它最多数址的帧时，主节点必须依次轮询每一个非活跃的节点；若单点失效（主节点有故障）则整个信道都不可操作。***令牌传递协议：*没有主节点**，一个称为**令牌(token)**的小的特殊帧在节点之间以某种固定的次序进行交换。当一个节点收到令牌时，仅当它确有帧要发送时，它才持有这个令牌；否则，它立即向下一个节点转发该令牌。一个节点收到令牌时，若它确有帧要传输，则发送最大数目的帧数，然后把令牌转发给下一个节点。**优点：**令牌传递是分散的，效率高。**缺点：单点失效**：一个节点的故障可能会使整个信道崩溃；如果一个节点偶然忘记了释放令牌，则必须调用某些恢复步骤使令牌返回到循环中来；令牌传递也有延迟。***MAC协议比较：*信道划分MAC协议（时分、频分、码分多址）:**重负载高效，没有冲突，节点公平使用信道；轻负载低效：即使只有一个活跃节点也只能使用 的带宽。**随机接入MAC协议(纯ALOHA, 时隙ALOHA, CSMA/CD（早期以太网）CSMA/CA（802.11））：**轻负载高效：单个活跃节点可以使用整个信道；重负载低效：频繁发生冲突，信道使用效率低。**轮流MAC协议(中心节点轮询（蓝牙）、令牌传递)（试图权衡以上两者）：**按需使用信道（避免轻负载下固定分配信道的低效）；消除竞争（避免重负载下的发送冲突），博采众长。***交换局域网：局域网LAN（Local Area Network）：***将小范围内的计算机及外设连接起来的网络，范围在几公里以内，通常为个人或机构所有；***城域网MAN：***通常覆盖一个城市的范围（几十公里），支持数据、音频和视频在内的综合业务，服务质量好，支持用户数量多；***广域网WAN（Wide Area Network）：***通常覆盖一个国家或一个洲（以上），规模容量可任意扩大。**链路层寻址和ARP：**不是主机或路由器具有链路层地址，而是每一块它们的网络适配器（网卡）固定分配一链路层地址（具有多个网络接口的主机或路由器将具有与之关联的多个链路层地址），称为***物理地址、LAN地址或MAC地址***，用于将链路层帧从一个接口转移到物理相连的另一个接口，接口在同一子网中。MAC地址具有**扁平结构**，不论适配器到哪里用都**不会变化**，而IP地址具有**层次结构**（即一个网络部分和一个主机部分），当主机移动时，其IP地址**需要改变**，即改变它所连接到的网络。以太网和802.11无线局域网的MAC地址长**6B**，形如A1-2B-3C-D4-E5-6F，由IEEE保证每块适配器的地址全球唯一（网卡生产商向IEEE购买一块MAC地址空间（前3Byte），其再确保生产的每一块网卡均MAC地址不同）；MAC地址固化在网卡的ROM中，不过也可能用软件改变网卡的MAC地址。**链路层帧的发送：**当某适配器要向目的适配器**发送**一个帧时，发送适配器将**目的适配器的MAC地址**插入到该帧中，并将该帧**发送到局域网**上。当适配器接收到一个帧时，将**检查**该帧中的目的**MAC地址**是否与它自己的MAC地址匹配。**如果匹配**，该适配器**提取**出封装的数据报，并将该数据报沿协议栈向上传递。**如果不匹配**，该适配器**丢弃**该帧，而不会向上传递该网络层数据报；交换机偶尔会将一个帧**广播**到它的所有接口，802.11也广播帧，因此一块适配器可以接收一个并非向它寻址的帧；有时，某发送适配器要让局域网上所有其他适配器来接收并处理它打算发送的帧。在这种情况下，它在该帧的目的地址字段中插入一个特殊的MAC广播地址（对于使用6字节地址的局域网来说，MAC广播地址为“全”）。**为什么有了IP还需要MAC？**为了使网络体系结构中各层次成为独立的构建模块，不同的层次需要有不同的寻址方案。**①**局域网是为任意网络层协议而设计的，而不只用于IP和因特网。如果适配器仅被指派IP地址的话，则不能方便地支持其他网络层协议（如IPX或DECnet）**②**如果适配器使用网络层地址而不是MAC地址，则网络层地址必须存储在适配器的RAM中，并且在每次适配器移动或加电时要重新配置，或者是在适配器中不使用任何地址，让每个适配器将它收到的每帧数据（通常是IP数据报）沿协议栈向上传递，由网络层核对网络层地址是否匹配，但此时主机将被局域网上发送的每个帧中断，包括被目的地是在相同广播局域网上的其他节点的帧中断。**地址解析协议（ARP）(同一子网内)：**ARP只为在同一个子网上的主机和路由器接口进行IP地址到MAC地址间的转换。每台主机或路由器在其内存中具有一个包含IP地址到MAC地址映射关系的ARP表，表项格式为<IP, MAC, TTL(过期时间，常为20min)>。在以太网上，ARP报文封装在以太帧中传输**。*地址解析的过程：***A希望发送数据包给B，且A的ARP表中当前没有B的表项：**①**A构造一个称为**ARP分组**的特殊分组，在**发送方**字段填入**自己的MAC地址**和IP地址，在接收方字段填入**广播MAC地址**和**B的IP地址②**A将ARP请求封装**广播**发送到子网中**③**每个收到ARP请求的节点用自己的IP地址与目标IP地址**比较**，**IP相同**的节点（B）进行**响应**。**④**B构造一个**响应ARP分组**：在发送方字段填入自己的MAC地址和IP地址，接收方MAC地址为A的MAC地址。然后B向A发送它**⑤A更新其ARP表，**并将其IP数据报封装在一个链路层帧中并发送，该帧的目的MAC地址就是B的MAC地址**⑥**A在其ARP表中缓存该信息，直到其超时。**不在同一局域网内的机器不适用上述过程的ARP广播查询*。*ARP是即插即用的**，ARP表是自动建立的。**最好把ARP看成是跨越链路层和网络层边界两边的协议**，虽然这不完全符合分层协议栈：一个ARP分组封装在链路层帧中，因而在体系结构上位于链路层之上；但一个ARP分组具有包含链路层地址的字段，因而可认为是链路层协议；又因为它也包含网络层地址，因而也可认为是为网络层协议。***数据报到达子网之外：***子网1中的主机A经过路由器R向子网2中的B发送数据报：**①**A的下一跳地址为**R的端口R-1**，R知道其**端口R-2可达B**（查询转发表）**②**A创建IP数据报（src\_IP＝A.IP, dest\_IP＝B.IP）；**③**A利用ARP获得**下一跳**（R-1端口）对应的**MAC地址**，**而非获得B的MAC④**A创建链路层帧（src\_MAC=A.MAC, dest\_MAC=R-1.MAC），封装IP数据报，发送**⑤**R接收帧，取出IP数据报，发现目的地址为B.IP；R利用ARP获得B的MAC地址**⑥**R创建链路层帧（src\_MAC=R-2.MAC, dest\_MAC=B.MAC），封装IP数据报，发送**⑦**B的网卡接收帧，取出IP数据报，交给网络层。**ARP与DNS的一个重要区别：**DNS为在因特网中任何地方的主机解析主机名，ARP只为在同一个子网上的主机和路由器接口解析IP地址。ARP广播可穿过交换机，但不能穿过路由器，会被路由器丢掉，因为两个LAN之间肯定是有路由器的，所以**主机不可能有其他LAN中主机的ARP表项**。**以太网：**第一个广泛部署的高速局域网技术，也是目前占主导地位的有线局域网技术；与其它的局域网技术相比，技术简单、成本低；以太网技术不断演化和发展，数据速率不断提高。***物理总线拓扑(mid 1970s）：***总线：以同轴电缆作为共享传输媒体（总线），所有节点通过特殊接口连接到这条总线上，是一个**广播局域网**。***物理星形拓扑***（1990s 后期）***：***基于**集线器**（即hub，物理层设备，作用于各个bit而不是帧，相当于共享电缆）。从一个端口进入的物理信号被集线器放大后立即从其它端口输出，因此也是一个**广播局域网**。如果集线器同时从两个不同的接口接收到帧，将出现一次碰撞，生成该帧的节点必须重新传输该帧。基于集线器的以太网为物理星型拓扑，逻辑总线拓扑。**交换式以太网**（21世纪初）**：**主机通过双绞线或光纤连接到交换机，交换机在端口之间存储转发帧（**链路层设备**），主机与交换机之间为**全双工链路**，**交换式以太网不会产生冲突，不需使用 CSMA/CD 协议！**交换式以太网实际上是**逻辑上**的星形拓扑（各节点仅与中心节点直接通信，各节点之间不直接通信）。**以太网帧结构：（按顺序）**发送端网卡将IP报文（或者其它网络层报文）封装为以太网帧。**前导码：**7个10101010字节加1个10101011字节，用于同步发送端、接收端的时钟速率（**链路层有流控制**）。**源地址/目的地址：**均为6Bytes的MAC地址。**Type（2Bytes）：**指出**Data**所属的上层协议（如IP、Novell IPX、AppleTalk和ARP等），每个协议有一个编号，用于多路分解（和网络层数据报中的协议字段、传输层报文段的端口字段类似）。**Data：**46～1500 Bytes，不足46B必须填充至46B；以太网的最大传输单元(MTU)是1500B，所以如果IP数据报超过了1500字节，则主机必须将该数据报分片；当采用填充时，传递到网络层的数据包括**IP数据报**和**填充部分**，网络层使用**IP数据报首部中的长度字段**来去除填充部分。**CRC（4Bytes，循环冗余检测）：**对源/目的地址、类型和数据4个字段计算得到CRC码。**以太网技术都向网络层提供无连接服务**：发送数据前，网卡间无需握手。**以太网技术都向网络层提供不可靠服务**：适配器对收到的帧执行CRC校验，但当该帧通过CRC校验时它不发送确认帧（ACK），当该帧未通过CRC校验时，它也不发送否定确认帧（NAK），而只是丢弃该帧。被丢弃的数据依靠**上层协议**（例如TCP）的**可靠数据传输机制**恢复（以太网并不知道它是正在传输全新的数据报还是已经被传输过至少一次的数据报），如**上层没有**可靠传输机制（如UDP），则**数据丢失**。在链路层缺乏可靠的传输使得以太网简单和便宜，但也意味着传递到网络层的数据报流能够有“间隙”（出现“缺漏”）。以太网的最小帧长为64B（不包括前导码），长度小于64B的帧为无效帧。***为什么有最小帧长的要求：***节点检测冲突需要时间。为确保节点在发送结束前检测到冲突，帧的发送时间必须足够长。假设信号在相距最远的两个适配器之间的**往返**延迟为，则帧的发送时间不应小于，即帧的最小长度链路速率×（根据早期以太网的最大直径（2500m）和数据速率（10Mbps）计算可得到最小帧长度）。以太网帧格式**30年未变**，是以太网一个**真正重要的特征**。**802.3以太网标准：链路层&物理层：使用CSMA/CD**历史上出现过许多不同的以太网技术，他们的链路层（MAC协议，帧格式）相同，物理层不同：传输媒体：光纤（X），同轴电缆，双绞铜线（T）；数据速率：10Mbps，100Mbps，1Gbps；物理层编码方式不同。“共同的链路层，不同的物理层”。以太网技术由IEEE 802.3工作组标准化，形成IEEE 802.3标准族。**MAC协议的必要性：**采用总线拓扑和基于集线器的星形拓扑的以太网是一种广播链路，其中多个节点同时传输时会出现帧碰撞。为了处理碰撞，以太网标准包括了CSMA/CD协议（对于小半径有线广播局域网很有效）；基于交换机的星形拓扑的以太网采用存储转发分组交换，交换机协调其传输，在任何时候都不会向相同的接口转发超过一个帧。此外，现代交换机是全双工的，一台交换机和一个节点能够同时向对方发送帧而没有干扰。所以，在基于交换机的以太局域网中，不会有碰撞，因此没有必要使用MAC协议。**共享式以太网和交换式以太网：共享式以太网(属于802.3)：**集线器的所有端口位于同一个碰撞域，任一时刻最多只允许一个主机发送，网络规模（节点数量）与网络性能的矛盾无法解决；**交换式以太网(不属于802.3)：**交换机每个端口为一碰撞域，多端口可以同时通信，网络的集合带宽=各个端口的带宽之和，从根本上解决了网络规模与网络性能的矛盾。***交换式以太网的最小帧长及规模：***交换式以太网**不再使用CSMA/CD协议，理论上不再需要限制帧的最小长度**；但**为了向后兼容**，帧格式及最小帧长度的限制仍然保持不变；由于交换式以太网不再使用CSMA/CD协议，**网络直径不再受到信号最大往返时间的限制**。**链路层交换机：交换机的任务**是接收入**链路层帧**并将它们转发到**出链路**。**交换机**自身对子网中的主机和路由器是**透明**的，它们无需知道交换机的存在：某主机/路由器向**另一个主机/路由器**寻址一个帧，**而不是**向交换机寻址该帧；将该帧发送进局域网后，它们也**并不知道**某交换机将会接收该帧并将它转发到另一个节点。**交换机输出接口有缓存：**因为帧到达交换机输岀接口的速率可能会暂时超过该接口的链路容量（类似于路由器接口为数据报设有缓存）。***过滤***：决定一个帧应该**转发**到某个接口还是应当将其**丢弃**的**交换机功能**。***转发***：决定一个帧应该被**导向哪个接口**，并把该帧**移动**到那些接口的**交换机功能**。交换机的过滤和转发**借助于*交换机表*完成**，它包含某局域网上某些主机和路由器的（但不必是全部的）表项。交换机表中的一个**表项包含①一个MAC地址②通向该MAC地址的交换机接口③表项放置在表中的时间**。**帧转发的描述类似于数据转发**，但**重要差异**是交换机转发分组基于MAC地址而不是基于IP地址，且交换机表与路由器的转发表的构造方式有很大差别。**交换机的过滤、转发（①、②）和自学习（③）：**当帧到来时，**①**记录帧的**到来端口②**用帧的**目的MAC地址**查找**端口转发表（转发*决策*）**：if(在表中**找到目的MAC地址**（节点已知）)then{if(帧的**到来**端口==目的地址所在端口)then丢弃（**过滤**不需转发的帧）else{转发帧到表项指定的端口（的缓存中）（按转发表转发帧）}}else{向输入端口以外的**所有**端口（的缓存）转发该帧（节点未知，采用扩散法泛洪广播）}**③【**自动、动态和自治地用帧的**源MAC地址**建立（维护）交换机表（**交换机是自学习的**）：**①**交换机表初始为空。**②**接收到帧时，若**源MAC地址不在**交换机表中，则该交换机**在其表中存储**：<该帧**源MAC地址**, 输入该帧的接口, 当前时间>，否则，交换机更新其表项：<该帧**新的源MAC地址**, 输入该帧的接口, 当前时间>。交换机以这种方式在表中记录了发送节点所在的局域网网段。**③**如果在一段时间（称为老化期）后，交换机没有接收到源地址为该地址的帧，则它在表中删除该项。】只要交换机的表是完整和准确的，它无须任何广播即可向目的地转发帧。交换机是**即插即用设备**(只需将局域网网段与交换机的接口相连，不用手动设置)，**也是双工的(**任何交换机接口能够**同时发送和接收)。**数据包要从A发往F，交换机S1如何知道应转发给S4，而S4如何知道应转发给S2？自主学习，与单交换机情形相同！**链路层交换机的性质：**交换机不同于如总线或基于集线器的星形拓扑那样的广播链路，**优点为①消除碰撞：**交换机缓存帧且不会在网段上同时传输多于一个帧**②链路异质：**交换机将链路彼此隔离，故局域网中的不同链路能以不同的速率和在不同的媒体上运行（但MAC协议要相同，如都是以太网或者WiFi）**③便于管理：**交换机提供强化的安全性，使得网络管理更加方便。**交换机和路由器比较：**交换机和路由器不同，它用MAC地址转发分组，是第二(链路)层的分组交换机，而路由器是第三(网络)层的分组交换机。***交换机优点*①**即插即用**②**分组过滤和转发速率相对高***交换机缺点*①**交换机**不能阻断广播帧传播**，对广播风暴并不提供保护措施**②**交换机**不能连接MAC协议不同**的网络（比如WiFi和以太网），因为交换机只是按原样转发帧**③**为了防止广播帧循环，交换网络的活跃拓扑限制为一棵**生成树④**大型交换网络的主机和路由器中ARP表规模大，ARP流量和处理量也大。***路由器优点*①**分组不会被限制到一棵生成树上，可以使用源和目的地之间最佳路径**②**对第二层的**广播风暴**（广播帧在网络中大量传播，消耗大量资源）提供防火墙保护，**能阻断广播帧传播③**路由器**可以连接MAC协议不同**的网络，因为路由器需重新封装链路层帧***路由器的点*①**路由器不是即插即用的（即路由器和连接到它们的主机都需配置IP地址）**②**路由器对每个分组的处理时间通常比交换机更长，因为它们必须处理高达第三层的字段。***虚拟局域网：***为了安全、隐私和改善局域网的性能，需要**限制广播流量的范围**，可以通过支持***虚拟局域网***（VLAN）的**交换机**来完成。支持VLAN的交换机允许一个物理局域网基础设施定义多个虚拟局域网，**同一个VLAN**内的主机**彼此通信**，仿佛**只有**它们（而没有其他主机）与交换机连接。在**基于端口**（也可基于终端MAC地址划分VLAN）的VLAN中，**交换机的端口**由网络管理员划分到不同的***组***中，**每个组均构成一个VLAN**，**属于**每个VLAN中的**端口**构成一个***广播域***，来自一个端口的**广播流量**仅能到达**该组中的**其他端口。**来自VLAN A的流量怎样才能发送到VLAN B中**？可以将VLAN交换机的一个端口与一台外部路由器相连，并且将该端口配置为属于VLAN A和VLAN B（就像真有两个交换机通过路由器连接，通常VLAN功能的交换机和路由器在一个设备里）。**VLAN交换机应当如何互联？**使用***VLAN干线连接（VLAN trunking）：***每台交换机上的一个特殊端口被配置为干线端口，该干线端口**属于所有VLAN**，发送到任何VLAN的帧**经过干线链路**转发到**其他交换机**。**一个交换机怎样知道到达干线端口的帧属于哪个VLAN？**对跨越VLAN干线的帧，IEEE定义**扩展的以太网帧格式802.1Q**，802.1Q帧包括标准以太网（802.3）帧与加进首部的4字节**VLAN标签**（承载该帧所属的VLAN标识符，插入到802.3的源MAC地址和类型中间）。VLAN标签由一个2字节的标签协议标识符字段（固定为0x8100）和一个2字节的标签控制信息字段（含12bit的VLAN标识符字段和3bit优先权字段）组成，由VLAN干线**发送侧**的交换机加进帧中，解析后由在VLAN干线**接收侧**的交换机删除。***链路虚拟化：网络作为链路层：MPLS（多协议标签交换）***是一种分组交换的**虚电路网络**，其目标是为在**不放弃**基于**目的IP地址**转发数据报的基础设施的前提下，当可能时，通过选择性地**标识数据报**，允许路由器基于**固定长度的标签**（**而不是**目的IP地址）转发数据报。支持MPLS的路由器被称作**标签交换路由器**，它们之间传输的链路层帧具有额外的MPLS首部（插入到了第二层（如以太网）首部和第三层（即IP）首部之间，算是2.5层协议但肯定不是IP的上层协议）。**标签交换路由器**通过在其**转发表**中查找**MPLS标签**来将**MPLS帧**转发给**适当的接口**，**不需要**提取目的IP地址和在转发表中执行最长前缀匹配。**MPLS转发表不同于IP转发表**，故**MPLS的转发决策不同于IP的转发决策**：**标签交换路由器**基于**MPLS标签值**转发IP报文（**而不检查IP地址**），源地址***和***目的地址可以共同决定转发路径，因此到同一目的地的帧可以经由不同路径，实现流量工程。入口MPLS路由器（和普通路由器直连的标签交换路由器）使用RSVP-TE协议建立到下游路由器的MPLS转发表项。MPLS的其他作用：链路故障时快速地重新路由（事先计算后备路径，经过一条预计算的无故障路径重路由流量来对链路故障做出反应）；用于实现VPN。***回顾：Web页面请求的历程* 一：*DHCP、UDP、IP和以太网*** 设主机A由一根以太网电缆连接到学校的以太网交换机，交换机又与学校的路由器相连，这台路由器与一个ISP连接，DNS服务器驻留在ISP网络中而非学校网络中且DHCP服务器运行在路由器中。A与网络连接时一开始没有IP地址，故其运行DHCP协议以从本地DHCP服务器获得一个IP地址以及其他信息：**①**A的OS生成一个DHCP请求报文并将其放入UDP报文段中（目的端口67（DHCP服务器），源端口68（DHCP客户）），该UDP报文段被放置在一个IP数据报中（目的IP地址为广播地址，源IP地址0.0.0.0，因为A还没有IP地址）**②**包含DHCP请求报文的IP数据报被放置在以太网帧中（目的MAC地址为广播MAC地址），它将广播到与交换机连接的所有设备（包括DHCP服务器）；该帧的源MAC地址是A的MAC地址MAC-A。**③**包含DHCP请求的广播以太网帧是第一个由A发送到以太网交换机的帧，该交换机在所有的出端口广播它，包括连接到路由器的端口。**④**路由器在其接口（MAC地址为MAC-1）接收到该广播以太网帧（包含DHCP请求），并且从该帧中抽取出IP数据报。该数据报的广播IP目的地址指示该IP数据报应当由该节点的高层协议处理，因此该数据报的载荷（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服务器拿到DHCP请求报文。**⑤**假设运行在路由器中的DHCP服务器能够以CIDR块68.85.2.0/24分配IP地址，且DHCP服务器分配地址IP-A给A。DHCP服务器生成包含IP-A以及DNS服务器IP地址IP-DNS、默认网关路由器IP地址IP-Def和子网块（68.85.2.0/24，等价为“网络掩码”）的一个DHCP ACK报文；该报文又被放入一个UDP报文段中，后者又被放入一个IP数据报中，IP数据报再被放入一个以太网帧中。这个以太网帧的源MAC地址是MAC-1，目的MAC地址是MAC-A。**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A收到过（包含DHCP请求的）以太网帧，所以该交换机知道寻址到MAC-A的帧仅从通向A的输岀端口转发。**⑦**A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。A的DHCP客户端记录下其IP地址和其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。A将向该默认网关发送目的地址为其子网68.85.2.0/24以外的所有数据报。此时A已经初始化好它的网络组件，并准备开始处理Web网页获取。**二：*DNS和ARP***当A将www.wjb.com的URL键入其Web浏览器时，谷歌主页最终将显示在其Web浏览器上。A的Web浏览器通过生成一个TCP套接字开始该过程，套接字用于向www.wjb.com发送HTTP请求。为生成该套接字，A需要知道www.wjb.com的IP地址。**①**A上的OS生成一个DNS查询报文，将“www.wjb.com”放入DNS报文的问题段中，它则被放置在一个目的端口为53号（DNS服务器）的UDP报文段中，该UDP报文段则被放入IP数据报中（目的地址IP-DNS，源IP地址IP-A）。**②**A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A学校网络中的网关路由器。为了获得该网关路由器的MAC地址，A需要使用ARP协议。**③**A生成一个具有目的IP地址IP-Def的ARP查询报文，将该ARP报文放置在一个具有广播目的MAC地址的以太网帧中，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。**④**网关路由器在通往学校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IP-Def匹配其接口的IP地址。网关路由器因此准备一个ARP回答，指示它的MAC地址MAC-1对应IP地址IP-Def。它将ARP回答放在一个以太网帧中，其目的地址为MAC-A，并向交换机发送该帧，再由交换机将帧交付给A。**⑤**A接收含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MAC-1。**⑥**A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。该帧中的IP数据报中IP目的地址为IP-DNS，而该帧的目的MAC地址为MAC-1（网关路由器）。A向交换机发送该帧，交换机将该帧交付给网关路由器。**三：*域内路由选择到DNS服务器* ①**网关路由器接收该帧，抽取含DNS查询的IP数据报，路由器查找该数据报的目的地址IP-DNS，并根据其转发表决定该数据报应当发送到哪里，这里设为路由器B。**②**路由器B接收到该帧，抽取IP数据报，检查该数据报的目的地址（IP-DNS），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，而转发表已根据ISP的域内协议（RIP、OSPF等）以及因特网的域间协议BGP填写。**③**包含DNS查询的IP数据报最终到达DNS服务器，后者抽取出DNS查询报文，在它的DNS数据库中查找名字www.wjb.com，找到包含对应www.wjb.com的IP地址（IP-wjb）的DNS源记录（假设它缓存在DNS服务器中；这种缓存数据源于google.com的权威DNS服务器。该DNS服务器形成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，该报文段放入寻址到A（IP-A）的IP数据报中。该数据报将转发到学校的路由器，并从这经过以太网交换机到A。**④**A从DNS报文抽取出服务器www.wjb.com的IP地址。**四：*Web客户-服务器交互——TCP和HTTP* ①**A有了www.wjb.com的IP地址，其生成TCP套接字，后者将用于向www.wjb.com发送HTTP GET报文。当A生成TCP套接字时，A中的TCP必须首先与www.wjb.com中的TCP执行三次握手。A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IP数据报中（目的IP地址为IP-wjb），将该IP数据报放置在MAC地址为MAC-1的帧中，并向交换机发送该帧。**②**在学校网络、ISP网络和谷歌网络中的路由器朝着www.wjb.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①②③那样。支配分组经ISP和谷歌网络之间域间链路转发的路由器转发表项由BGP协议决定。**③**最终包含TCP SYN的数据报到达www.wjb.com，从数据报抽取出TCP SYN报文并分解到与端口80关联的欢迎套接字。对谷歌HTTP服务器和A之间的TCP连接生成一个连接套接字，产生一个TCP SYN ACK报文段，将其放入向A寻址的一个数据报中，最后放入链路层帧中，该链路将www.wjb.com连接到其第一跳路由器。**④**包含TCP SYN ACK报文段的数据报通过谷歌、ISP和学校网络，最终到达A的以太网卡。数据报在OS中分解到①生成的TCP套接字，进入连接状态。**⑤**借助A上的套接字，准备向www.wjb.com发送字节。A的浏览器生成包含要获取的URL的HTTP GET报文，后者被写入套接字，其中GET报文成为一个TCP报文段的载荷。该TCP报文段被放置进一个数据报中，并交付到www.wjb.com，如前面步骤①②③所述。**⑥**在www.wjb.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。**⑦**包含HTTP回答报文的数据报通过谷歌、ISP和学校网络转发，到达A。A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Web网页。

***第七章 无线和移动网络：***

***概述：构成无线网络的要素：*无线主机：**运行网络应用，可能静止或移动（**无线并不一定意味着移动**）。**基站：**常连接到有线网络，在无线终端和固定网络之间传递数据包，通常负责协调与之关联的多个无线主机的传输。**无线链路：**连接无线终端和基站，需要MAC协议协调无线链路的使用，使用多接入协议避免碰撞，不同的无线链路数据传输速率和传输距离不同。**自组织网络**中，无线主机没有这样的基础设施与之相连（网络中没基站），主机本身必须提供路由选择、地址分配和类似于DNS的名字转换等服务，节点只能与通信范围内节点通信，节点相互帮助转发分组，每个节点既是终端又是路由器。***切换：***无线终端接入到不同基站的过程。***无线网络的分类：*①无线网络中的分组跨越了一个还是多个无线跳②网络中是否有基站这样的基础设施。1.单跳有基础设施**(如WiFi)主机连接到基站，基站连接到更大的固定网络；**2.多跳有基础设施**（如无线网状网络）主机可能需通过多个无线节点的中继才能到达固定网络**3.单跳无基础设施**（如蓝牙网络）无基站，不连接到更大的固定网络，节点间通信不需要中继**4.多跳无基础设施**（如自组网，车载网）无基站，不连接到固定网络，节点间通信需通过其它节点中继。***无线链路和网络特征：无线链路的特性：*信号强度损失：**无线信号能量随传播距离衰减（称为路径损耗）**和其他信号源互相干扰：**受到其它发送源或电磁噪声的干扰；**多径传播：**由于地面或物体的反射作用，信号沿多条不同长度的路径到达接收端，使接收方收到的信号模糊；**以上特性导致无线链路的传输距离受限、误码率很高。信噪比(SNR)**，信噪比大则更容易提取出信号。***信噪比与误码率的权衡：*同一物理层:**增加功率→提高信噪比→降低误码率；**给定信噪比:**选择满足误码率要求的物理层，再从中选出吞吐量实现最高的；注意**信噪比可随移动性变化**，即移动场景下可动态选择物理层（调制技术、速率）***无线网络的特性：*隐藏终端问题：**A和C同时向B发送，由于**信号强度衰减**，A和C所处的位置使得**A、C**的信号强度**不足以**使他们相互**检测到对方**，**但足以在B产生冲突**，因此***CSMA/CD不适合多跳无线网络***（通过载波侦听，发送节点只能知道**其周围**是否有节点在发送；但真正影响此次通信的是**接收节点周围**是否有节点在发送）。***解决方案是CDMA：***所有用户共享相同的频率，但每个用户都用自己的CDMA码编码数据，多个用户共存并以最小干扰(CDMA码是**“正交的”**)同时传输。设传输的bit是，CDMA码是形如的向量，则**编码结果为**，其中，**解码过程为：**，这里。当有多个发送端和接收端时，它们的编码结果、等在传输过程中互相干扰，**相加**形成，但**只要**所有的CDMA码**正交**（**内积为0**），直接用替代解码过程中的即可**正常解码**，不受影响。***WIFI：802.11无线LAN：*802.11b**：2.4-5 GHz，最高11 Mbps；**802.11a：**5-6 GHz，最高54Mbps；**802.11g：**2.4-5 GHz，最高54 Mbps；**802.11n:**多天线，2.4-5 GHz，最高200Mbps；**802.11ac：**5GHz，最高1300bps；**802.11ax**最高9.6 Gbps。上述标准均使用**CSMA/CA**作为MAC协议；都支持**基站模式**和**自组织模式**，但物理层不同。**802.11体系结构：*802.11无线LAN的基本组成单元***是***基本服务集（BSS）。***一个BSS**包括：**若干个无线站点、一个中央基站（称为AP，即接入点）。不同BSS中的AP可连接到互联设备（交换机/路由器）上，互联设备又可连接到因特网中。每个802.11无线站点都具有一个6B的MAC地址，该地址存储在该站适配器（802.11网络接口卡）的固件中。每个AP的无线接口也具有一个MAC地址，与以太网类似，这些MAC地址是全球唯一的。**注意：**AP与路由器相连的有线端口**没有MAC地址**，即AP对于路由器是透明的。配置AP的无线LAN常被称作**基础设施无线LAN**。***802.11信道与关联：***802.11将通信频段划分成若干信道（在2.4-2.4835GHz定义了11个部分重叠的信道），每个BSS**分配一个**信道：管理员安装每个AP时，均为其分配一个**服务集标识符（SSID）**，并选择AP使用的信道号；距离近的多个AP使用相同信道**可能**相互干扰（此时通过CDMA码避免干扰）。**每个无线站点若想发送或接收网络层数据必须与一个AP关联：**主机扫描11个信道，监听各个AP发送的**beacon帧**（周期性发送，包含AP的SSID和MAC地址），选择一个AP进行关联（可能需要身份鉴别），使用DHCP从AP所在的子网获得一个IP地址。***802.11主动/被动扫描：被动扫描：*①**多个AP广播beacon帧**②**主机扫描信道和监听AP发送的beacon帧，选择一个AP发送关联请求帧**③**AP向主机发送关联响应帧***主动扫描*：①**主机向位于其通信范围内的所有AP广播探测请求帧**②**若干个AP用探测响应帧应答**③**主机从收到的探测响应帧中选择一个AP发送关联请求帧**④**AP发送关联响应帧**802.11MAC协议：采用CSMA**：发送前监听信道，不与当前正在进行的发送冲突；**不检测冲突**：发送过程中检测冲突很困难（接收信号的强度衰减，远小于发送信号的强度），不能检测出所有的冲突（隐藏节点）。因此，一旦开始发送一个帧，就发送完该帧。***802.11目标：*避免**冲突，**选用CSMA/C(ollision)A(voidance)**。***以太网和802.11都使用CSMA，但MAC协议有重要的区别：***802.11使用**碰撞避免(竞争)而非碰撞检测（非竞争）**；由于无线信道相对较高的误比特率，802.11(不同于以太网）使用链路层确认／重传(ARQ)方案。**链路层确认方案：****802.11的CSMA/CA（不使用信道预约）：发送方：①**若初始时站点监听到信道空闲，则在一个分布式帧间间隔（DIFS）的短时间段后发送该帧。**②**否则，该站点选择一个随机回退值（类似二进制回退），并在监听信道空闲时递减该值；如果监听到**信道忙**则**该值不变（被冻结）**。**③**若计数器减为0，该站点发送整个帧并等待确认。**④**如果收到确认且想要发送第二个帧，回到②；若给定时间内**没有收到确认**，回到②，并从**更大**的范围内选择**随机回退值**。**接收方：**如果成功收到帧，则在短帧间间隔(SIFS)后发送ACK。***采用随机时间递减的原因：***尽量避免碰撞冲突。若两个站点同时进入随机回退状态，则时间短的先发送，并且阻塞时间长的，可有效避免碰撞。***为什么需要链路层ACK：***难以检测碰撞且节点不能中断传输；接收方只会在没有碰撞时确认帧。***链路层确认可以完全避免碰撞吗：***不能，隐藏节点（检测不到对方）或两节点选择了接近的回退时间均可能导致冲突。**加上信道预约特性以处理隐藏终端：加入*RTS和CTS*（←*二者均可选*，只对长数据帧使用）操作方法：①**若A欲AP发送一个数据帧，则A先向AP发送一个RTS（请求发送）帧，其中给出随后要发送的数据帧及确认（ACK）帧需要的**总时间**（虽然避免了数据帧的碰撞，但RTS帧仍可能造成碰撞，不过RTS帧小，浪费资源少）；若A和B同时发送RTS帧，则产生冲突，所有不成功的发送方**随机等待**一段时间后重试。**②**AP收到RTS帧后广播一个CTS（允许发送）帧，帧中给出相同的时间**③**A收到CTS帧后开始发送帧，其他收到CTS帧的节点均沉默指定的时间，让出信道让A完成发送。**④**AP收到帧后，发送一个ACK帧进行确认。***CSMA/CA与CSMA/CD的不同：*最根本的不同：**CSMA/CD在发送过程中**检测冲突**，而CSMA/CA在发送过程中**不检测冲突**；**由此带来的协议处理方面的不同：**在CSMA/CD中，节点侦听到信道空闲时**立即发送**；在CSMA/CA中，节点侦听到信道空闲后要**随机回退**，因为冲突对无线网络损害很大，要尽可能避免。***暴露终端问题：***A、B分别向A1、B1发送帧，且A、B**能干扰对方**，但A1、B1**并不互相干扰**。尽管如此，A、B**还是会因为对方而随机退避**，浪费传输资源，降低传输效率，该问题无法解决。***DCF/PCF信道接入机制***DCF采用上述的CSMA/CA+RTS/CTS（大的数据帧可选），而PCF中AP轮询终端是否数据需要传输，无线终端之间无竞争。802.11中PCF和DCF交替工作，一个交替周期被称为CFP。*802.11可作为点对点链路。***802.11帧格式：*四个地址字段：*Address 1：**要**接收该帧的无线站点**的（目的）MAC地址，可以是移动无线站点或AP的；**Address 2：**是传输该帧的站点的（源）MAC地址，可以是移动无线站点或AP的；**Address 3：**AP连接的路由器接口的MAC地址；**Address 4：**只在自组织模式中使用***。802.11帧寻址举例（路由器R1向主机H1发送数据包，H1回应）*①**R1知道H1的**IP地址**IP-H1，使用**ARP**获得H1的MAC地址MAC-H1，将**802.3帧（有线）**的目的地址设置为MAC-H1，源地址设置为MAC-R1，发送给AP**②802.3帧**到达AP，AP将其**转换为802.11帧（无线）**，设置MAC-H1为地址1，MAC-AP为地址2，MAC-R1为地址3**③**H1构造链路层802.11帧回复，其中地址1为MAC-AP、地址2为MAC-H1地址3为MAC-R1 **④**帧到达AP，AP将这个该**802.11**帧**转换**为**802.3帧**，以MAC-H1为源地址，MAC-R1为目的地址。注意，AP连接路由器的有线端口**没有MAC地址**！**AP仅对无线终端可见，对于固定网络上的设备不可见。其他字段：**Duration（预留的传输时间，用于RTS/CTS）；seq control（帧序列号，用于可靠传输）；Type（RTS, CTS, ACK, 数据）；To AP/From AP（方向）；WEP(是否加密)**802.11终端在同一IP子网内移动：*切换：***终端从一个BSS移动到另一个BSS。**发生切换时，终端要关联到新的AP上①**当H1检测到来自AP1的信号逐渐减弱时，开始扫描新的beacon帧**②**当H1收到信号更强的beacon帧时（假设来自AP2），先解除与AP1的关联，然后关联到AP2**③**发生切换时，交换机（同时连接AP1和AP2）中的**转发表**也需要更新。交换机***通过自主学习更新转发表：***交换机收到H1发送的帧时，更新H1所在的端口；若转发表未及时更新，可能产生丢包。**802.11f规定了AP间漫游的方法。主机停留在同一个IP子网中，因而IP地址保持不变**。切换过程中，终端上的应用正常运行：由于IP地址没变，网络层及以上层次感觉不到这个移动，切换过程中产生的延迟及丢包，在上层协议看来是正常的。**802.11终端在不同IP子网内移动：**若AP1和AP2同时连接到路由器R1，则AP1和AP2属于不同的子网，H1的IP地址在移动过程中会发生变化，其IP地址不能保留：H1通过DHCP申请新地址，伴随会话中断。**802.11先进功能：*速率适配：***当节点移动或信噪比变化时，基站和节点动态改变传输速率（物理层调制技术）；**实现：**连续两帧未收到ACK，则降低到前一个较低的速率；连续收到10帧的ACK或上次降速后计时器超时，则提升到上一个较高的速率。***能耗管理：*节点设置功率管理位，告知AP它将进入休眠状态：**功率管理bit置1，则节点进入休眠，并正好在下一个beacon帧之前醒来（AP常每100ms发送一个beacon帧）；在节点休眠期间，AP缓存发往该节点的帧；AP在发送信标帧前，恰好唤醒节点，节点迅速进入全面活动状态（用时约250）；由AP发送的信标帧包含了帧被缓存在AP中的节点的列表。如果某节点无缓存帧，则它返回睡眠状态。否则，该节点通过向AP发送一个探询报文请求发送缓存帧。对于beacon帧之间的100ms时间来说，250的唤醒时间（以及很短的接收beacon帧与检查是否由缓存帧的时间）使得没有帧要发送和接收的节点能够睡眠99%的时间，节省了大量能源。***移动用户的编址和路由原理：***移动性：一直使用同一AP接入网络→接入新网络使用DHCP分配的新地址，会话中断→漫游多个网络时保持会话不中断，移动性**不断变强**。终端进入到一个新的子网后，必须分配该子网上的一个地址（DHCP），并使用新的地址通信，不能保留其IP地址。然而，当终端改变IP地址后，终端上正在运行的应用将中断：通信的对方不知道终端的新地址，无法与其通信；终端上正在运行的网络应用也将中断，因为通信的对方不知道终端的新地址，无法与其通信；即使对方获知了终端的新地址，应用也必须重新建立连接，因为通信的套接字变了。***术语：*本地网络：**移动节点的永久“居所”(e.g.128.119.40/24)**。永久地址：**移动节点**在本地网络中的地址**，**总是**可以使用这个地址与移动节点通信，即使移动也保持不变（e.g.128.119.40.186）。**本地代理(“老家”的代理)：**当移动节点在外地时，本地网络中代表移动节点执行移动管理功能的实体。**被访问网络：**移动节点**当前所在**的网络(e.g.79.129.13/24)。**外地代理(设备目前处在的路由)：**被访问网络上为移动节点执行移动管理功能的实体。**转交地址（COA，亦称外部地址）：**移动节点在**被访问网络**上的地址(e.g.79.129.13.2)。**通信者：**希望与移动节点通信的节点。***寻址交给终端系统解决：间接路由：***通信对象与**本地代理**通信，再由**本地代理转交**给移动节点。***直接路由：*通信对象**获得**移动节点**当前在被访问网络中的**转交地址**，与其**直接通信**。应当注意到移动节点也能承担外地代理的责任。例如，某移动节点可在外部网络中得到一个COA（使用DHCP之类的协议），且由它自己把其COA通告给本地代理。***移动节点注册：*①**移动节点进入**被访问网络**后，向**外地代理**注册（移动节点离开该外部网络时将向外部代理取消注册）**②**移动节点通过**外地代理**向**本地代理**注册移动节点的COA（移动节点离开其网络时外部代理不需要显式地注销COA，因为当移动节点移动到新网络时，随之而来就要注册一个新的COA，自动完成了注销）。**结果：**外地代理可以和移动节点通信、本地代理知道移动节点当前位置（在哪个被访问网络上，转交地址是什么）。***路由选择到移动节点：间接选路到移动节点：*①**通信者在数据包中使用移动节点的**永久地址②**本地代理截获该数据报，封装在一个数据报内（**隧道技术**：外加一层**IP头**，src为**本地代理IP地址**，dest为**COA**），转发给外地代理**③外地代理**收到隧道数据报，解封装，取出通信者的**原始数据报**，转发给移动节点**④**移动节点**直接**将响应发送给通信者。***间接选路特点：***移动节点**使用两个地址**：**永久地址：**通信者用来向移动节点发送数据报（通信对象无需知道节点是否发生移动，移动节点的位置对于通信者是透明的）；**转交地址：**本地代理用来向外地代理转发数据报；***存在“三角”选路*：**通信者→本地网络→移动节点，当通信者和移动节点在同一个网络中时很低效。***间接选路中终端在被访问网络间移动：***设节点移动到**另一外地网络**，并**希望保持**和通信对象的会话，则需要**①**节点向新的外地代理注册**②**新的外地代理向本地代理注册**③**本地代理更新移动节点的转交地址**④**本地代理经过新的转交地址向移动节点转发包（**仍然是隧道技术**）。上述**①②④均经过广域网**。***优点：*间接选路中**，节点移动及变换被访问网络等**对通信者都是透明的：**已建立的会话不受影响，通信对象始终和移动节点的永久地址通信。***直接选路到移动节点：*①**通信者向本地代理请求，获知移动节点的转交地址（此步以后不必再做）**②通信者**通过**隧道技术**将包发送给外地代理（**外层**dest为**COA**，**内层**dest为**永久地址**）**③**外地代理解封装，将包转发给移动节点**④**移动节点直接将响应发送给通信者。***直接选路特点：***克服了三角选路问题，但对通信者不透明。**若移动节点移动到新网络且希望保持会话**，则需要将移动节点**一开始**访问网络的**外地代理**标记为**锚点外地代理：①**假设节点发生新的移动前已建立会话**②**节点移动到**新的被访问网络③**移动节点向**新的外地代理**注册**④**新的外地代理向**锚点外地代理**提供移动节点的**新转交地址⑤**通信者向锚点外地的代理发送**隧道数据包**（**外层**dest为**锚点外地代理IP地址**，**内层**dest为**永久地址**）**⑥**锚点外地代理收到**隧道数据包**，**解封装**提取原始数据包，**再封装一层新头部（**src为**锚点外地代理的IP地址**，**dest**为**COA）**，转交给移动节点的新转交地址（“跑得了和尚跑不了庙”）。**⑦**同间接选路的③&④。间接路由vs.直接路由**①对通信对象是否透明？间接透明**，通信对象仅需要知道移动节点的永久地址；**直接不透明**，通信对象需要知道移动节点当前的转交地址，并自己建立隧道；**②效率？间接低**，存在三角路由问题；**直接高**；**③**在**移动中保持会话**？**间接简单**，本地代理更新移动节点新的转交地址；**直接复杂**，涉及锚点外地代理和新的外地代理两次隧道。***移动IP（RFC 3344）：***支持移动性的因特网体系结构与协议合称移动IP，包含很多特性：本地代理，外地代理，注册，转交地址，隧道；**标准化了三个部分：**代理发现，向本地代理注册，数据报间接选路（当前的标准RFC 5944规定到移动节点使用**间接路由选择**的方法）。**代理发现：**通过新的网络地址使移动节点中的网络层知道其已进入一个新的外部网络（发现新的外部代理），可经**代理通告**或**代理请求**实现**：①**愿意当本地代理或外地代理的路由器发送**代理通告**，宣布自己的存在及IP地址**②**愿意当外地代理的路由器在代理通告中提供若干转交地址（常为自己的IP地址）**③**移动节点接收分析代理通告，判断自己是否处于被访问网络&是否切换了网络**④**如果发现自己在被访问网络上，它从外地代理提供的转交地址中选择一个作为自己的转交地址。**若使用代理请求**则移动节点不必等待代理通告，其可广播代理请求报文，收到的代理将直接向它单播一个代理通告。***向本地代理注册：*①**移动节点向外地代理发送注册请求，给出自已的永久、转交、本地代理地址和认证信息等**②**外地代理记录节点永久IP地址，向本地代理转发注册请求**③**本地代理处理注册请求，若认证通过，将移动节点的永久及转交地址保存在绑定表中，发回注册响应**④**外地代理收到注册响应后，将移动节点记录在自己的转发表中，向移动节点转发注册响应**⑤**当移动节点回到本地网络时，向本地代理注销注册。***数据报间接选路：*①**数据包先被本地代理得到**②**本地代理查找地址绑定表，获得移动节点当前的转交地址**③**本地代理将数据包发送到转交地址**④**外地代理将数据包转发给移动节点。***本地代理如何得到数据报？*若通信者不在本地网络上①**数据包首先到达移动节点本地网络上的路由器**②**路由器查表得知可以直接转发，于是查找ARP缓存或者发送ARP请求，以获取移动节点的MAC地址**③**利用得到的MAC地址将数据报封装到链路层帧中发送。**若通信者在本地网络上：**通信者查表得知移动节点直接可达，于是查找ARP缓存或者发送ARP请求，利用得到的MAC地址封装数据报发送。***数据报如何被本地代理得到？***链路层帧的dest=本地代理的MAC地址。即移动节点的永久地址应当映射到本地代理的MAC地址。***ARP代理：***本地代理为位于**被访问网络**的移动主机用自己的MAC地址发送ARP响应；即将移动主机的永久地址映射到本地代理的MAC地址。***免费ARP：***接收到移动主机的注册请求后，本地代理主动发送ARP请求，刷新其它节点的ARP缓存。**本地代理如何将数据报发送到转交地址？**本地代理收到的数据报的dest为移动节点的永久地址，而移动节点的转交地址位于被访问网络。将目的地址在本地网络的数据报送达被访问网络需**使用隧道技术**（不改变目的地址，对上层协议透明）。***本地代理通过隧道转发数据包：*本地代理向外地代理发送：**Src=本地代理IP，Dest=COA，内封装通信者发送的包（dest是永久地址）；**外地代理**解封装收到的数据包，得到原始数据报并向移动节点发送。*外地代理如何获得移动节点的MAC地址？***①**在移动节点注册阶段，外地代理获知移动节点的永久地址和MAC地址，记录在其转发表中**②**外地代理根据目的IP地址查转发表，得移动节点的MAC地址**③**外地代理用移动节点MAC地址将数据报封装到链路层帧中发送给移动节点。***移动节点如何发送数据包？***移动节点将数据包发送给外地代理（缺省路由器）：SrcIP/SrcMAC=自己的永久/MAC地址，DestIP=通信者IP地址，DestMAC=外地代理MAC，外地代理再按正常方式转发数据包。***移动节点如何得知外地代理的MAC地址？***代理通告报文的源MAC是外地代理MAC。***无线和移动对上层协议的影响：***无线链路带来**误码率、丢包率、延迟增大**问题，节点移动带来**丢包、延迟增大**问题。**逻辑上影响小**（为上层协议仍然提供的尽力而为的服务，TCP和UDP可运行在无线网络上），**性能上影响大**（丢包率高，传输延迟增大，TCP将丢包和长延迟解释为拥塞而减小拥塞窗口，吞吐率不必要地降低；无线&有线-无线混合链路的TCP拥塞控制是一个问题；高时延对实时业务影响大；无线链路往往带宽有限而导致拥塞。

***第八章 计算机网络中的安全：***

***什么是网络安全：***安全目标—**保密性**只有发送者和指定的接收者可“理解”消息内容，发送者加密消息，接收者解密消息；**完整性**消息在传输过程中没有被篡改或被篡改后能发现；**可用性**安全作为一种服务被用户使用；**真实（可信）性**发送者和接收者需向对方证明身份。Bob和Alice均可代表真人或DNS服务器等；攻击者无处不在（偷听/插入消息/模仿/劫持/拒绝服务攻击）。***密码学原理（消息安全性）：明文：***欲加密的原始数据；***密文 ：***明文经加密算法作用后的输出；***密钥：***加密/解密使用的参数。***针对加密系统的密码分析攻击：*仅依赖密文的攻击：**密码分析者仅能根据截获的密文进行分析，以得到明文或密钥（对密码分析者最不利）；**已知明文攻击：**密码分析者除有截获的密文外，还有一些已知的“明文-密文对”来帮助破译密码；**选择明文攻击：**密码分析者可以指定一些明文，用被攻击的加密算法加密，得到相应的密文。**一个安全的加密系统必须能抵御选择明文攻击。*按加密/解密密钥是否相同，加密算法分为：*对称加密算法**（加密/解密密钥相同）和**非对称加密算法**（加密/解密密钥不同）。***对称密钥密码体制：替换密码：***用密文字母替换明文字母，但字母位置不变。例：单表置换（凯撒）密码/多表置换密码（按特定顺序循环加密用字母表）；密钥：字母映射表。**块密码：**将bit的块映射为bit密文，越大，越难暴力破解（种可能的密钥），但维护个项的表很难，且若每个明文块被独立加密则相同的明文块生成的密文块相同，易被利用。***块密码链（CBC）：***发送方生成一个随机的初始向量(IV)，记为，用明文发送给接收者，则明文块的加密结果（为XOR，代表对“”查CBC表得到的结果）。优点：相同的明文块几乎不可能得到相同的密文块；只额外传输，不因引入随机性产生大量额外带宽，但需要在协议中提供机制分发并保证按顺序传输。**块加密算法举例：对称密钥算法DES**采用CBC，用56bit对称密钥加密64bit明文，输出64bit密文。DES是基于迭代(16轮)的算法，每轮均执行加密换位。没有有效的统计分析DES破解方法，但暴力破解56bit密钥，仅需不到一天，更安全的方法是**3DES**（使用不同密钥的DES加密3次，安全性指数级上升）。**对称密钥算法AES**加密128 bit明文块，密钥长度128/192/256 bit，暴力破解AES需149万亿年。**对称加密算法的问题：**要求发送者和接收者使用同一个密钥。**优点：**加密速度快，便于硬件实现和大规模使用；**缺点**是存在**密钥分发问题：**发送者和接收者未必能安全地共享密钥。***非对称密钥密码体制：***发送者和接收者不共享密钥。发送者用加密密钥（公开），接收者用解密密钥（私有），不存在密钥传递问题。**优点：不存在密钥分发问题**（不必保持信道的保密性）。**缺点：**加密速度慢，不便于硬件实现和大规模使用。***公开密钥算法的使用：***每个用户生成一对加密密钥和解密密钥：**加密密钥**放在一个**公开**的文件中，**解密密钥妥善保管**。**公开密钥和私有密钥：**公开密钥（）：加密密钥，发送者使用；私有密钥()：解密密钥，接收者使用。***公开密钥算法要求：***且给定公钥，不可能计算出私钥。运算性质：。**RSA算法：**消息作为bit串可唯一地表示为整数，对消息加密等价于对整数加密。**步骤：①生成密钥：**选择两个大素数和（如1024bit的），计算和；再选一个与互质的数，并找到满足的，则公钥为，私钥为（这里代表有序对）。**②加密方法：**将明文bit串看成十进制整数（）。计算，即为的密文。**③解密方法：**对密文，计算，即为所求的明文。**重要性质：**，即**公钥加密**、**私钥解密**和**私钥加密**、**公钥解密**的**结果相同**。**优点：**安全性好（对大数质因数分解目前不可行）；使用方便（不必传递密钥）。**缺点：**计算开销大，速度慢（DES比RSA快100倍）。***RSA的应用：***RSA一般用来加密少量数据，如用于鉴别、数字签名或分发会话密钥等（例：Bob和Alice使用RSA交换对称密钥：发送方用计算，接收端用解密获得，双方接下来用对称密钥加密会话）。**Diffe-Hellman算法：**发送端和接收端都知道大质数和一个大数，是的原根（）。密钥只有自己知道，公钥所有人都知道，则Alice、Bob分别选择密钥、，计算公钥、，则Alice和Bob可以分别计算共享对称密钥：、，二者结果都是！***认证（消息真实性）：端点鉴别：***一个实体经过计算机网络向另一个实体证明其身份的过程。鉴别应当在报文和数据交换的基础上，作为某***鉴别协议***的一部分独立完成。鉴别协议常在两通信实体运行其他协议（如可靠数据传输协议、路由选择信息交换协议或电子邮件协议）**之前**运行。鉴别协议首先建立相互满意的各方的标识，仅当鉴别完成之后，各方才继续后面的工作。**目标**：Bob要Alice证明她是Alice。**鉴别协议 ap1.0：**Alice直接发送一个报文说“我是Alice”，但Trudy可伪装成她。**鉴别协议 ap2.0：**Alice说“我是Alice”，并从Alice的IP地址发出IP包，但Trudy可用Alice的IP地址伪造数据包（IP地址欺骗）。**鉴别协议 ap3.0：**Alice说“我是Alice”，并向Bob发送密码证明，但Trudy可窃听到Alice发送的明文口令并发送给Bob（回放攻击）。**鉴别协议 ap3.1：**Alice说“我是Alice”，并将加密的密码发送给Bob，但Trudy可截获数据包、记录口令加密版本，并向Bob回放（又是回放攻击）**鉴别协议 ap4.0：*目标：***避免**回放攻击（通过不重数）。*不重数(nonce)：***在一个协议的生存期中只使用一次的数（一旦某协议使用了一个nonce，则永不会再使用该nonce）。***ap4.0以如下方式使用一个nonce：*①**Alice说“我是Alice”**②**Bob选择一个不重数R发给Alice**③**Alice使用她与Bob共享的对称密钥（假设有）来加密它并把发回给Bob**④**Bob解密接收到的报文，若得到的不重数等于之前发送给Alice的那个不重数，则可鉴别Alice的身份。**缺点：**需要一个**共享**的**对称密钥**。**鉴别协议 ap5.0（采用公开密钥算法加密不重数）：①**Alice说“我是Alice”**②**Bob选择一个不重数R发给Alice**③**Alice使用她的私钥加密R并把发回给Bob**④**Bob对Alice说“给我你的公钥”，Alice发送。**⑥**Bob计算，若其为，则因为只有Alice拥有，故对面一定是Alice。但Trudy可向Bob扮演Alice，向Alice扮演Bob（中间人攻击）。***消息完整性：***使用公钥加密长消息计算代价高，需要易于计算、固定长度的数字“指纹”来鉴别消息完整性。对消息应用哈希函数可获得固定长度（128/160bit）的报文摘要。**哈希函数性质：①**多个变量值对应一个函数值**②**产生**固定**长度的摘要**③**给定摘要，反过来得计算上不可行，即不可求逆。**哈希函数算法：MD5：**4步计算128bit消息摘要；**SHA-1**：160bit消息摘要。**消息认证码1.0**（有漏洞的消息完整性保护）**①**Alice创建消息并计算哈希②Alice把附在后面，形成扩展消息发给Bob③Bob收到扩展消息并计算，发现**④**Bob认为消息未经篡改。但Trudy可构造假消息并计算，声称她是Alice并向Bob发送。**消息认证码2.0：①设**Alice和Bob有一个共享秘密，称为认证密钥**②**Alice将附在后，形成，计算（被称为**消息认证码（Message Authentication Code，MAC）**）。**③**Alice把MAC附在后形成扩展消息并发送给Bob。**③**（知道的）Bob收到扩展消息并计算MAC（）**⑤**若则Bob认为未被篡改。怎样分发认证密钥？手动配置或公钥分发（D-H等）。**数字签名：**发送方数字签名文件，证明ta是文件的所有人/创建者，具有可验证、不可抵赖性：接收者（Alice）可以证明Bob而不是其他人（包括Alice自己）签名了这个文件。**过程：①**Bob用自己的私钥加密消息，创建签名。**②**设Alice收到消息和签名**③**Alice用Bob的公钥解密签名，并检查是否有成立。**④**若，则签名方一定持有Bob的私钥，Alice可以验证：Bob对消息签名、Bob对而非消息签名、不是其他人签名。**不可抵赖：**Alice可用和签名证明Bob签名了消息。**问题：**消息很长时，用私钥加密后得到的签名过长。**签名数字摘要：**Bob仅对的摘要用签名（仅长128/160bit）。此时，Bob同时发送和，Alice用解密，并与从直接计算出的比较。如果两者相符，表明报文是真实的。**回顾ap协议5.0的漏洞：**Bob发布公钥，Alice获取Bob公钥的请求被Trudy截获，Trudy将假冒的Bob公钥（实为Trudy的）发送给Alice，Alice用加密会话密钥并发送给Bob，Trudy截获会话密钥，用Bob的公钥加密后再发送给Bob→Alice和Bob之间通信的报文都被Trudy破译。**Alice从公开途径得到Bob的公钥后，如何确认所得确为Bob的公钥，而不是别人的？需要可靠的第三方→*认证中心（CA）***将公钥与实体绑定，实体（网站、服务器、路由器等）向CA注册自己的公钥：实体向CA提供身份证明，CA创建实体到其公钥的绑定，CA用**自己的私钥**为**实体的公钥**签名，形成**公钥证书**，好比CA说**“我证明这是某实体的公钥”**。***公钥证书***包含主体的公钥和公钥所有者身份标识信息（全局唯一），并由CA进行数字签名（用**CA的私钥**），任何人无法伪造或篡改证书的内容（本质是加密的公钥）。实体获得其公钥证书后可将其放在任何可公开访问的位置。**当Alice希望获得Bob的公钥时**，Alice从Bob或别处获得Bob的证书，再用**CA的公钥**验证Bob的证书，获得。小结：**①**安全性：对大量消息加密用对称密钥，分发对称密钥用RSA或D-H。**②**真实性（Alice向Bob证明她真是Alice）：nonce或CA。**③**完整性：MAC（使用哈希函数和认证密钥）或数字签名摘要。【到这里了】***安全email*（&满足保密性、满足完整性、满足真实性）：**Alice希望发送保密邮件给Bob。Alice：**①**用哈希函数生成邮件摘要**②**用加密（签名）**③**用对称密钥加密和，得到。**④**用Bob的公钥加密，得到**⑤**将和发给Bob。Bob：**①**使用恢复对称密钥**②**用解密得到和**③**用解密得到计算，验证二者相等。Alice使用三个密钥：、和新创建的对称密钥。安全email的协议是**PGP**（无标准号，是事实标准）。***安全TCP连接（SSL协议）：***SSL位于TCP之上、应用层之下，依靠TCP提供可靠的端到端连接，向**所有**基于**TCP**的网络应用提供**安全的传输层**服务（如支持Web浏览器和服务器之间的安全通信（https）），支持保密、完整和真实性。**过程：①握手：**客户先与服务器创建一条TCP连接（完成TCP三次握手），验证服务器是真实的服务器，再生成并发送给服务器一个用加密的主密钥，客户和服务器用生成SSL所需的所有对称密钥。**过程：**客（发）TCP SYN→服TCP/SYNACK→客TCP ACK；客SSL hello→服certificate→客。**②密钥导出：两者都用KDF处理生成密钥**：（客到服加密）（客到服的MAC密钥）；、类似，不过是服到客的。**③数据传输：**将数据流分割为**记录**，对每个record附加MAC用于完整性检查，将“记录+MAC”加密传给服务器。这里MAC=Hash（记录+MAC密钥+记录（功能）类型（0-传数据/1-关连接）+TCP序号），以防止重排序、重放报文段和关闭连接攻击。**④SSL记录：**由类型字段(hello or 应用)、版本字段(自解释)、长度字段、数据字段和MAC字段组成。前三个字段不加密。***网络（IP）层安全：*IPsec和VPN**目标：通过对流量进行加密，将公共网络当成本地局域网络用。**IPv4在设计时未考虑安全性**，缺少对通信双方身份的鉴别，易受地址欺骗攻击；缺少对数据完整性和机密性的保护，数据易被窃听、修改和劫持。**IPSec有两种模式**：传输模式（终端系统收发IPsec报文，仅保护IP的负载）和隧道模式（可用于路由器间或终端系统间，保护整个IP报文，加密后再封装一层IP头部）**IPSec有两种协议**：**AH协议**（提供数据源认证、数据完整性，但不提供保密性）和**ESP协议**（比AH多提供保密性，应用更广）。**隧道模式+ESP**应用最多。***安全关联（SA）：***发送数据前，发送实体和接收实体间需建立安全关联（SA）。**SA是两个通信端点间从发到收的*单向*连接**（如果需要双向安全通信则需建立2个SA），由32bit的安全参数索引（SPI）**标识**。**发送和接收实体均维持SA的状态，故IPsec是有状态、有连接协议。**SPI、初始接口和SA目的接口、加密类型、加密密钥、完整性检查类型、鉴别密钥等SA的状态信息存储在IPSec实体的**安全关联数据库（SAD）**中。SA可建立在一对主机间、一台主机与一个安全网关间和一对安全网关间，所以个在外地的终端和总部双向连接需要个SA。***IPsec数据报：取决于IPSec使用模式：*运输模式：**IPSec头被插入到原始IP头和传输层头之间，路由器根据原始IP头转发数据包；**隧道模式（关注）：**原始数据包被封装在一个新的IP包中，**IPSec头被放在新的IP头和原始IP头间**，路由器根据**外层**IP头的信息转发数据包。隧道的端点（外层IP头中的地址）通常是一个支持IPSec的安全网关。***两种模式的比较：***传输模式占用较少的带宽，隧道模式更安全（隐藏内部网络的细节（原始IP头不可见））；内部网络上的主机可以不运行IPSec，它们的安全性由安全网关来保证；隧道模式可以将一对端点间的通信聚合成一个**加密流**，从而有效地防止入侵者。**ESP数据包(载荷)大致分为以下几个部分：ESP头：**包含SPI和**序号**（新SA序号初始化为0，并随数据包发送而递增，**防止重放攻击**）；**载荷：**原始数据包中被加密部分的密文（初始IP首部+初始载荷）；**ESP尾部：**包括填充（若需要）、填充长度和下一个首部，ESP尾部也要被加密；**ESPMAC**（覆盖ESP头、载荷和ESP尾的报文鉴别码，**完成真实性和完整性**）。***安全无线LAN：802.11WEP（Wired Equivalent Privacy有线等效保密）***是最初的802.11规范使用的安全协议，在主机和基站之间提供较弱的加密及鉴别服务，没有密钥分发机制。**WEP设计目标**：**对称密钥加密**（保密、终端接入授权、数据完整性）；**每个数据帧独立加密**（前一数据帧丢失不影响后一数据帧解密）；**高效**（硬件或软件实现）。实现简单：明文和keystream按位XOR得密文、密文和keystream按位XOR得明文（细节略）。WEP有漏洞，现在用的是WPA和WPA2。***防火墙和入侵检测系统：*防火墙：**在可信的内部网络与不可信的外部网络之间执行访问控制策略的硬件或软件系统（硬软结合体），隔离机构内部网络和公共互联网，放行某些数据包，阻止其它数据包；**目的：**保护内部网络免受来自外部网络的攻击。为何需要防火墙？**①**阻止拒绝服务攻击（SYN泛洪：攻击者建立大量TCP伪连接，服务器没有足够资源留给真正的连接，但该问题实际上无解）**②**阻止非法的内容修改（攻击者篡改机构的网站）**③**内部网络仅允许授权访问（用户/主机认证）***防火墙的类型：*无状态数据包过滤器、有状态数据包过滤器、应用网关**。***无状态数据包过滤器：***内部网络通过**路由器防火墙**连接到公共互联网，路由器对数据包进行逐包过滤，基于以下字段决定转发还是丢弃包：源IP地址，目的IP地址；TCP/UDP源端口号和目的端口号；ICMP消息类型；TCP SYN标志和ACK标志。***包过滤策略的例子：*①**不允许访问外部Web网站→丢弃所有外出的、目的端口为80的包**②**不允许外部发起的TCP连接，除非访问的是内网的公共web服务器→丢弃进入的TCP SYN包，除非去往某IP的端口80**③**防止因特网广播吞噬网络带宽→丢弃除了DNS和路由广播外的所有进入的UDP**④**防止网络拓扑被探测（traceroute）→丢弃所有外出的ICMP TTL过期消息**⑤**阻止外部客户发起到内部服务器的连接→过滤进入的所有ACK比特设为0的报文段→阻止所有外部的客户端连接内部的服务器，但是允许内部发起TCP连接**⑥**阻止所有到达和离开的，IP协议字段=17，或源或目的端口号是23的数据包→阻止所有UDP流和telnet连接。***访问控制列表（Access Control Lists/ACL）：***访问控制列表是一个规则表，包含一系列<匹配条件，动作>。对于每个进出的包，**无状态数据包过滤器孤立地**从上到下地匹配规则来**过滤每个包**，其***仍会允许一些异常的包进入***(例：80端口上**并未建立TCP连接**，但**允许**dest port=80, ACK=1（表示已经建立TCP连接）的包进入)。***有状态数据包过滤器***可跟踪**TCP连接**的建立（SYN）和关闭（FIN）等**状态**。实现：扩展ACL，在允许放行一个包前检查连接的状态，判断收到的包是否有意义，有意义才放行。***应用网关：***应用网关除检查网络层及传输层协议头（IP、TCP/UDP），还检查**应用层**数据。**例：**允许特定的内部用户使用telnet连接外部服务器**①**所有telnet用户必须连接到应用网关**②**对授权用户，网关建立到目的服务器的telnet连接，并在2个连接之间中继消息**③**路由器丢弃所有不是从网关发起的telnet数据包。**防火墙的局限性：①**无法抵御IP欺骗攻击：路由器**无法验证**数据包是否真实来自**源地址②**每个被代理的应用都需要单独的应用网关，而应用网关处理开销大，速度慢**③**应用网关对于用户不透明，客户端需知道怎样连接应用网关**④**防火墙通常完全允许或完全禁止UDP，应考虑数据传输量而非全部允许或禁止。**更高级的做法**→**入侵检测系统（IDS，不是防火墙）：**防火墙**不检查**数据包之间的**关联**，而IDS执行深度数据包检查来查看包内容（如检查内容是否匹配已知病毒特征和攻击特征），并**检查多个包之间的关联**性（包括端口扫描、网络拓扑映射、DoS攻击等情况）。**拒绝服务（denial of service/DoS）**的攻击方式有多种，任何通过合法的方式使服务器不能提供正常服务的攻击手段都属于广义DoS攻击的范畴。**最基本**的DoS攻击手段是利用合理的服务请求来占用过多的服务器资源，使服务器无法响应合法用户。**网络中可以设置多个IDS**在不同位置进行不同类型的检查。***为什么使用多个IDS传感器？***IDS不仅需做深度分组检查，还必须将每个过往的分组与很多“特征"进行比较，可能导致极大的处理量。将IDS传感器进一步向下游放置，则每个传感器仅看到该机构流量的一部分，更易维护。***基于特征的IDS的限制：***要求根据以前的攻击知识来产生准确的特征，换言之，对不得不记录的新攻击缺乏判断力。另一个缺点：即使数据包与特征匹配，它也可能不是一次攻击，即产生虚假告警。最后，因为每个分组必须与较大的特征集合比较，IDS可能处于**处理过载**状态并因此**难以检测**出许多**恶意分组**。***为什么在因特网的多个层次上提供安全性功能？仅在网络层提供安全性功能并加以实施不够吗？*①**尽管可以通过加密数据报中的所有数据（所有的传输层报文段）以及通过鉴别所有数据报的源IP地址在网络层提供“地毯式覆盖”的安全性，但并不能提供用户级的安全性（如一个商业站点不能依赖IP层安全性来鉴别一个在该站点购买商品的顾客）。**②**在协议栈的较高层上部署新的因特网服务（包括安全性服务）通常较容易，而在网络层上广泛地部署安全性可能还需若干年。

链路层单位为**帧**，网络层单位是**分组**。**不是**每个子网**都有**一个**DHCP服务器**。**并非**每个**CIDR**地址块都是一个**子网**。**路由器**转发表是被**域内**路由协议**和域间**路由协议**共同**配置的。域内路由器的**转发表不会仅包含域内**地址前缀的表项。**数字签名**是**非对称**的。**交换机**的每个**端口**是独立的**冲突域**，**不阻止**广播帧，对于目的MAC地址**不在转发表**中的帧会**泛洪**。**CRC**的检错能力取决于**生成多项式**。IPv4首部中“标识”字段的作用是分片。采用**NAT**技术解决IPv4地址不足的最主要依据：数据包转发时主要考虑将数据包转发到目的**网络**。CIDR路由表最长前缀匹配**不可降低**路由查找算法的**时间复杂度**。同一IP地址中的网络地址可能不同（子网掩码不同）。**路由器**不会将一个**子网**/**LAN**中的**链路层广播**扩散到其他**子网**/**LAN**中。IPv6不分片，头部长固定，支持option。AS内部可划分为区域（OSPF）。所有网关路由器执行相同域间路由协议，边界路由器**同时运行**BGP和域内路由协议。CSMA**不可避免**冲突，适用于小规模广播网络。链路层协议**有控制流**，可**检错**。MAC地址是**扁平化**而非层次化的。**随机接入协议**：ALOHA及其时隙版本，CSMA及其衍生。主机**没有**关于**其他LAN**中主机的ARP表项。MPLS**不算**IP的上层协议。4G/5G是**单跳**（仅需手机到基站），**有基础**设施的。802.11MAC协议**不解决暴露终端**问题，**不检测冲突**。使用**CTS**通知**隐藏终端避让**，**RTS/CTS不能确保**不发生碰撞，**BSS和LAN没有直接关系**（BSS是一个AP和关联到该AP的所有移动节点的集合，BSS间可以由交换机或者路由器相连，假如BSS之间是通过交换机相连，则BSS则不构成LAN。所以，BSS和LAN之间没有直接的关系，取决于具体的组网方式）。AP接入点**不可以**在属于同一个BSS的无线终端之间转发802.11帧，因为**AP**只是一个无线网络到有线网络的接入点，**只执行信号转换（协议转换）功能**，不执行转发功能。**RTS/CTS**机制其实**不是必须**的，而是可选项（假如无线网络里就一个移动节点，显然没必要）。【不同局域网的**安全网关**之间通过IPSec实现安全通信。如果采用最安全的ESP+隧道模式，则两个安全网关之间传输的原始IP包会被整个加密，作为安全载荷封装到一个新的IP包中传输。其结果是，两个网络之间传输的流量与其它流量隔离，外网节点无法获知原始IP包的任何信息，也无从探测和发起针对局域网的攻击，提高了内网及网间通信的安全性。】