第四章：网络层

网络层的**转发功能**和**路由选择功能**做**重要区分**：转发涉及分组在单一的路由器中从一条入链路到一条出链路的传送。路由选择涉及一个网络的所有路由器，它们经路由选择协议共同交互，以决定分组从源到目的地结点所采用的路径。**概述：转发和路由选择：网络层的作用**即将报文段从一台发送主机移动到一台接收主机。**分组交换机**是指一台通用分组交换设备，它根据分组首部字段中的值，从输入链路接口到输出链路接口转移分组。某些分组交换机称为**链路层交换机**，基于链路层字段中的值做转发决定。其他分组交换机称为**路由器**,基于网络层字段中的值做转发决定。某些计算机网络中，实际上有**第三种重要的网络功能，即连接建立**。**网络服务模型：网络服务模型**定义了分组在发送与接收端系统之间的端到端运输特性。在发送主机中， 当运输层向网络层传递一个分组时，能由**网络层提供的**特定服务**包括：**•确保交付• 具有时延上界的确保交付；能够为给定的源和目的地之间的分组流**提供下列服务：**•有序分组交付•确保最小带宽•确保最大时延抖动•安全性服务。**因特网的网络层**提供了单一的服务，称为**尽力而为服务**。**虚电路和数据报网络：网络层**也能够在两台主机之间提供**无连接服务**或**连接服务**；**与运输层类似：**网络层连接服务以源和目的主机间的握手开始；网络层无连接服务则没有任何握手预备步骤；**重大差异：**在网络层中，这些服务是由网络层向运输层提供的主机到主机的服务。在运输层中，这些服务则是运输层向应用层提供的进程到进程的服务；网络层不同时提供这两种服务，仅提供连接服务的称为**虚电路网络**，仅提供无连接服务的称为**数据报网络**；二者实现面向连接的服务是根本不同的，运输层面向连接服务是在位于网络边缘的端系统中实现的，网络层连接服务除了在端系统中，也在位于网络核心的路由器中实现。**虚电路网络：因特网是一个数据报网络**。网络层连接被称为**虚电路**。虚电路是一条端到端路径，其**行为**：传输分组前建立虚电路，传输结束后拆除虚电路，每个路由器为经过它的虚电路维护状态；路由器资源（带宽、缓存等）可以分配给虚电路，从而虚电路能提供可预期的网络服务。**建立虚电路的本质**是预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**如下**：**源和目的主机之间的路径（即一系列链路和路由器）；VC 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）；沿着该路径的每台路由器中的转发表表项（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 号替代每个传输分组的 VC 号。该新的 VC 号从转发表获得。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**。**一个分组沿着其路由在每条链路上不简单地保持相同的 VC 号的原因：**减少了在分组首部中 VC 字段的长度；大大简化了虚电路的建立，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。**信令报文：**专门用于建立、维护、拆除虚电路的控制报文。**信令协议：**交换信令报文的协议。**数据报网络：**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的映射；转发表被选路模块修改，约 1～5 分钟更新一次。**同一对主机之间传输的分组可能走不同的路径，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；当有多个匹配时，路由器使用**最长前缀匹配规则**，即在该表中寻找最长的匹配项。虽然在数据报网络中的路由器不维持连接状态信息，但它们无论如何在其转发表中维持了转发状态信息。**虚电路和数据报网络的由来：ATM（虚电路网络）**由电信网发展而来；注重用户体验（用户付费），追求高质量服务；终端无智能或很少智能；复杂工作由网络完成，以保持终端简单。**Internet（数据报网络）**为计算机通信而设计；早期的网络应用均为弹性应用，对网络服务没有严格要求；用户免费使用网络；终端（计算机）具有智能；可将复杂的工作（如差错控制）推到网络边缘，以保持网络核心简单。**数据报网络只提供最小服务的好处：**可运行在各种链路之上；增加新服务只涉及终端。**路由器工作原理：路由器的 4 个组成部分：输入端口。**将一条输入的物理链路与路由器相连接的物理层功能；需要与位于入链路远端的数据链路层交互的数据链路层功能；查找功能，通过查询转发表决定路由器的输出端口，到达的分组通过路由器的交换结构将转发到输出端口；排队，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）从输入端口转发到路由选择处理器；注意这里的端口一词，是指路由器的物理输入和输出接口。**交换结构。**交换结构将路由器的输入端口与输出端口相连接，这种交换结构完全包含在路由器中， 即它是一个网络路由器中的网络！**输出端口。**输出端口存储从交换结构接收的分组；组装：若需要，将交换结构输出的信元组装成分组；排队：若输出端口来不及发送，分组在此排队；调度：若有多个等待队列，选择一个队头分组发送；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。**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，执行网络管理功能。一台路由器的输入端口、输出端口和交换结构**共同实现了这种转发功能**，并且**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**。这些**路由器控制平面**通常用**软件实现**并在路由选择处理器（CPU）上执行。**输入端口：**转发表的一份**影子副本**通常会被存放在每个输入端口。转发表从路由选择处理器经过独立总线（例如一个 PCI 总线）复制到线路卡。有了影子副本，转发决策能在每个输入端口**本地做出**，无须调用中央路由选择处理器，因此**避免了集中式处理的瓶颈**。必须要用硬件执行查找，而且需要对大型转发表使用超出简单线性搜索的技术，同时必须对内存访问时间给予特别关注。**采取许多其他动作：**检查分组的版本号、检验和以及寿命字段，并且重写后两个字段；更新用于网络管理的计数器。**交换结构：**交换可以用许多方式完成：**经内存交换：**第一代路由器：由传统计算机构成，选路和交换都由 CPU 完成，交换通过拷贝完成；现代路由器：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；交换过程：输入端口将一个包放入内存，其接口硬件通过控制器发送一个消息，给出包在内存中的存放地址，输出端口从指定的内存位置读取包，发回响应消息；性能和代价取决于存储接口数目，仅适合小容量路由器，一次只能转发一个分组。**经总线交换：**[illegible]……每个端口都有一个……总线协议防止多……分组，分组中……发给本端口的……**经互联网络交换：**输入端口与输出端口……**排队：**输入排队和输出排队两种，阻塞……入交换结构，……目的地为相同的……移形成分组队列……取决于流量负载……分组转移到输出……其后分组的转发……的 n 倍时（n……**输入排队交换机中**……结构发送（即……输入端口同时……输出端口排队……减少丢包的发……浪费资源）；输……**主动队列管理**……与 TCP 拥塞控制一起使用；$AvgLen = (1-Weight)×AvgLen + Weight×SampleLen$……丢弃概率 p 丢弃……的分组；概率……丢弃间隔越大，……的，即不同部……进行交互。**因特网中的转发和编址：**因特网的网络层有三个主要组件：IP 协议；选路协议；……数据报中的差错……**网际控制报文协议(ICMP)**。**数据报格式：**•版本号。•首部长度……数据报不包含……•数据报长度（以字节为单位）总长度……字节。•标识、标志、片偏移……（如由于长时……•寿命：字段的值减 1；……•上层协议：……数据报的数据部分……TCP、UDP；类似多……•首部检验和：用反码运算……路由器要对每……丢弃；注意到……•选项：……及可能的选项……对 IP 首部计……与 IP 不一定……上，而 IP 能……•数据（有效载荷）。……40 字节的首部……**IP 数据报分片：**一个链路层帧能承载的最大……较大的 IP 数据报……装成一个独立……在端系统上而非路由器……必须携带与原……位置；**标志位：**MF (more fragments)：最后一个分片的 MF=0，其余分片的 MF=1；DF (don't fragment)：

DF=1 表示不允许对数据报分片；分片报头中的**以下字段需要修改：**总长度，偏移量，MF，头部检查和。由于偏移量只有 13 比特，除最后一个分片外，其余分片的数据长度应为 8 字节的整倍数，假设原始数据报的报头长度为 H，则分片的数据长度 N 应为满足以下条件的最大整数：$N ≦ MTU – H$，N 为 8 的倍数。**数据报分片的处理过程：**根据报头长度 H 和输出线路的 MTU，确定分片的最大数据长度 N；将数据报的载荷划分成长度为 N 的若干数据块（最后一个数据块可能不足 N 字节）；将原始报头加到每一个数据块的前面，修改报头中的以下字段：总长度 = H + 数据块长度，最后一个报头的 MF 位置 0，其余报头的 MF 位置 1，偏移量 = 数据块在原始数据报载荷中的字节序号/8，计算头部检查和。**重组：**将收到的分片重新组装成原始数据报的过程称为重组，重组在目的主机中进行：收集分片：目的主机使用 <源 IP 地址，标识> 确定属于同一个数据报的分片；利用最后一个分片计算原始数据报长度：原始数据报长度=偏移量×8 + 分片总长度；组装：将各分片中的数据块按照其在原始数据报载荷中的偏移量重组。**分片的问题：**分片的开销：降低了路由器的吞吐量；消耗了目的主机的资源，每个重组的数据报需要一个重组缓冲区和一个重组定时器；使得系统更为复杂。针对分片的 DoS 攻击：攻击者发送一系列奇怪的分片，消耗目的主机的资源。**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，主机通常只有一个接口。**IP address:**每个网络接口对应一个 IP 地址，IP 地址是一个 32 位的二进制数，通常用点分十进制数表示。**基于类的编址（早期）：**A 类 8bit（第一位 0）子网地址+24bit 网内地址，B 类 16bit（前两位 10）+16bit，C 类 24bit（110）+8bit，D 类（1110）+广播地址，E 类（1111）+保留后续使用。**单播地址结构：**单播地址除类别标识外，其余比特被划分成网络号和主机号两部分：网络号：标识一个物理网络；主机号：标识该物理网络上的一个网络接口；同一个物理网络上的网络接口，它们的 IP 地址具有相同的网络号。**地址分配：**因特网中的每个接口必须具有唯一的 IP 地址，为在因特网范围内保证 IP 地址的全局唯一性：网络号由 ICANN 统一分配，主机号由网络管理员统一分配；建立私有网络的组织可以自己选择网络号，但同样必须保证每个网络号在私有网络内的唯一性。**特殊的地址：**全 0 或全 1 的网络号及主机号是特殊地址，从不分配给特定的网络接口：网络号有效、主机号全为 0 的地址：保留给网络本身。网络号有效、主机号全为 1 的地址：保留作为定向广播，即在网络号指定的网络中广播（仅用作目的地址）。32 位全 1 的地址：本地广播地址，表示仅在发送节点所在的网络中广播（仅用作目的地址）。32 位全 0 的地址：指示本机（仅用作源地址）。网络号为 0、主机号有效的地址：指代本网中的主机。形如 127.xx.yy.xx 的地址：保留作为回路测试，发送到这个地址的分组不输出到线路上，而是送回内部的接收端。**网络数量与地址数量：接口地址数：**A 类地址：2^24-2 = 16777214；B 类地址：2^16-2 = 65534；C 类地址：2^8-2 = 254。**地址的个数：**A 类地址：27-2 = 126；B 类地址：214-2 = 16382；C 类地址：221-2 = 2097152。**子网（subnet）：**具有相同子网地址、不需要通过路由器就可以相互到达的网络接口构成一个子网。**子网编址：**从概念上说，引入子网仅略微改变了 IP 地址的解释：主机号被进一步划分成子网号和主机号两部分；子网号标识网络内的一个子网，主机号标识子网中的一个网络接口；管理员可以根据需要，如子网的数量和规模，选择合适的子网号长度。**子网掩码：**用于指示 IP 地址中子网号与主机号的边界：子网掩码是一个 32 比特的数，其中对应主机号的比特为 0，其余比特为 1；子网掩码也采用点分十进制表示，如 255.255.252.0；**如何从 IP 地址中获取子网地址？**将 IP 地址与子网掩码做“与”运算，例如：128.10.1.1 AND 255.255.255.0 = 128.10.1.0。**子网掩码将 IP 地址划分为两部分:**子网地址：对应子网掩码中“1”的部分；主机地址：对应子网掩码中“0”的部分。**如何确定子网？**将网络接口与主机/路由器分开，形成一些分离的网络岛，每个网络岛就是一个子网。路由器的每个端口连接一个子网，不同的端口连接不同的子网。路由器是在子网之间转发数据包的设备。子网内部通信不需要通过路由器，子网之间通信必须通过路由器。**网络层转发数据报的两种情形：**直接交付：节点（源主机或目的路由器）将数据包直接发送给目的主机（不需要其它路由器转发）；间接交付：节点将数据包转发给一个路由器去处理。**如何判断使用直接交付还是间接交付？**直接交付：数据包的目的地址与节点的某一端口在同一个子网中；间接交付：数据包的目的地址不与节点的任何一个端口在同一个子网中。**间接交付的实现：**节点查找转发表，将数据包发送给下一个路由器。**直接交付的实现：**第五章。**转发表：**记录目的地址到输出端口的映射；取决于目的地址类型的不同，有**三类转发表项：**目的地址是一个**子网地址**，称**地址前缀表项**；目的地址是一个**特定的网络接口地址**，称特定主机表项；缺省项：不匹配所有其它表项的地址，这些地址被映射到一个**默认的**路由器端口。**IP 采用逐跳选路：**每个转发表项只记录去往目的地址的下一跳信息（下一个要到达的路由器端口），而不是一条完整的端到端路由。**每个转发表项包括：目的地址/掩码、下一跳地址、输出端口等。**下一跳地址必须与输出端口在同一个子网中（不需要通过其它路由器就可以直接到达）。**IP 数据报的转发过程（主机/路由器）：**从数据报中提取目的 IP 地址 D，利用子网掩码计算地址前缀 N；if D 与自己的任何一个 IP 地址匹配（本节点是数据报的目的节点）then 将数据包交给 protocol 域指定的协议实体处理 elseif N 与自己的任何一个直连网络的地址匹配（直接交付）then 通过该直连网络把数据包直接交付到目的节点 D elseif 表中包含到 D 的特定主机表项 then 把数据包发送到表中指定的下一跳 elseif 表中包含到 N 的一个地址前缀表项 then 把数据包发送到表中指定的下一跳 elseif 表中包含一个缺省项 then 把数据包发送到表中指定的默认路由器端口 else 宣告选路出错，向数据包的源地址发送一条错误报告消息（ICMP）。**分类编址的缺点：**只能按照三种固定的大小分配地址空间，地址浪费严重；转发表必须记录每个已分配的网络，转发表规模爆炸式增长。**CIDR（Classless InterDomain Routing）：**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，从而可将其看成是一个具有 21 位子网地址的网络。**CIDR 地址分配的原则：**地址块的长度 L 必须是 2 的幂次；所有地址的前（32-log_2_L）位必须相同。**网络地址的表示方法：**用**掩码**指示**网络地址**的长度，如 194.24.0.0，255.255.248.0；用“/长度”指示**子网地址**的长度，如 194.24.0.0/21。**机构如何获得网络地址？**机构通常从 ISP 的地址空间中分配地址。**地址聚合：**转发表中符合以下条件的若干个表项可以合并成一个表项：这些表项的目的地址可以聚合成一个前缀更短的地址；这些表项使用相同的下一跳。地址聚合的过程可以**递归进行。若个别表项不满足路由聚合的条件：**仍然可以在转发表中给出一条聚合表项；**同时**给出不能被聚合的表项；**最长前缀匹配：**在所有匹配的路由表项中，选择前缀最长的表项。**查找转发表：使用分类地址：**转发表分为 A、B、C 三张表，分别记录 A、B、C 三类地址的转发表项，用哈希表组织；路由器收到数据报后：根据目的地址的类型确定要查找的转发表，根据目的地址的类型提取网络地址，用网络地址在相应的转发表中进行哈希查找（精确匹配）。**采用 CIDR 后出现的问题：为与某个转发表项 Dest_addr/prefix_len 进行匹配运算：**路由器需要先从表项中读出地址掩码（或 prefix-len 值），计算包的目的地址前缀（用地址掩码和包的目的地址相与），与 Dest_addr 的地址前缀（Dest_addr 与地址掩码相与）进行比较；**引入的问题：**地址前缀的长度 prefix_len 可以是任意值，Prefix_len 无法从地址本身得到，只能从转发表项中得到，必须从所有匹配的表项中选择前缀最长的表项。在大规模转发表中进行**快速查找**是一个难题（已经解决）。**主机/路由器如何获得 IP 地址？路由器：**管理员手工配置路由器各个接口的 IP 地址。**主机：**管理员手工配置主机 IP 地址，服务器通常采用这种方法；主机使用**动态主机配置协议 DHCP（Dynamic Host Configuration Protocol）**获取 IP 地址、子网掩码、缺省路由器、本地 DNS 服务器等配置信息。**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 目标：**允许主机加入网络时自动获取配置信息。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 代理。**DHCP 为主机发现新 IP 地址的四个步骤：**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:654），寻找子网中的 DHCP 服务器；DHCP 服务器用“DHCP offer”报文（src:223.1.2.5,67 \dest:255.255.255.255,68 \yiaddrr:223.1.2.4 \transaction ID:654 \Lifetime:3600secs）进行响应，给出推荐的 IP 地址及租期、其它配置信息；主机用“DHCPrequest”报文（src:0.0.0.0,68 \dest:255.255.255.255,67 \yiaddrr:223.1.2.4 \transaction ID:655 \Lifetime:3600 secs）请求 IP 地址，主机选择一个 DHCP 服务器，向其请求 IP 地址；DHCP 服务器用“DHCP ack”报文（src:223.1.2.5,67 \dest:255.255.255.255,68 \yiaddr:223.1.2.4 \transaction ID:655 \Lifetime:3600 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 不足之处：**从移动性角度看，结点移动时，不能维持与远程应用之间的 TCP 连接。**网络地址转换（NAT）：Motivation:**使用一个公用 IP 地址支持许多用户同时上网，仅为公共可访问的节点分配公用 IP 地址（减少需要的公用 IP 地址数），网络内部节点对外是不可见的（安全考虑）。地址空间 10.0.0.0/8 用于家庭网络等专用网络或具有专用地址的**地域**。**NAT 实现：外出的数据报:**将数据报中的（源 IP 地址，源端口号）替换为（NAT IP 地址（不变），NAT 端口号（不断分配））**NAT 转换表**：记录每个（源 IP 地址，源端口号）与（NAT IP 地址，NAT 端口号）的转换关系；**进入的数据报**：取出数据报中的（目的 IP 地址，目的端口号）查找 NAT 转换表，然后用转换表中对应的（IP 地址，端口号）进行替换。**16 比特端口号:允许一个 NAT IP 地址同时支持 65535 个对外连接。NAT 的使用有争议:**路由器应当只处理三层以下的包头（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 后面的主机对外是不可见的；但 P2P 应用要求任何对等方可以向任何其它（参与的）对等方发起通信。**使用 UPnP 实现 NAT 穿越：**假设主机在端口 3345 上运行一个 BT 程序：BT 程序请求 NAT 产生一个“洞”，将<10.0.0.1,3345>映射到<138.76.29.7,5001>上；BT 程序向追踪器通告它在<138.76.29.7,5001>上可用；其它主机通过追踪器可以看到该主机，并能向<138.76.29.7, 5001>发起 TCP 连接；NAT 将<138.76.29.7,5001>上收到的 SYN 包转发给主机。**使用中继服务器实现 NAT 穿越：**在 Skype 中使用：NAT 后面的服务器与中继器建立连接，外部客户与中继器建立连接，中继器在两个连接之间转发分组。**因特网控制报文协议：**主机或路由器使用 **ICMP 协议**传递网络层上的一些信息。ICMP 报文有**询问和错误报告两类**：询问：用来请求一些信息，通常采用请求-响应模式交互；错误报告：发现错误的节点向源节点报告错误信息，不需响应。**ICMP 与 IP 的关系：**ICMP 报文被封装在 IP 包中传输：这是因为 ICMP 报文可能需要经过几个网络才能到达源节点；ICMP 通常被认为是 IP 协议的一部分，因为 IP 协议使用 ICMP 向源节点发送错误报告。**ICMP 报文格式：**type：报文类型，共定义了 15 种；code：

对某类报文作进一步的区分；Checksum：ICMP 报文的检查和；内容：与报文类型有关，报告错误的 ICMP 报文包含触发该错误的数据报的头部和前 8 个数据字节。**Ping 与 ICMP:** Ping 利用 ICMP 报文测试目的主机是否活跃，以及去往目的主机的路径是否正常：源主机发送 Type=8, Code=0 的 Echo Request 报文；若目的主机收到，发送 Type=0, Code=0 的 Echo Response 报文；源主机计算 RTT，并报告；若源主机连续几次超时（收不到 Echo Response），向调用者报告目的主机不可达。**Traceroute 与 ICMP:** Traceroute 测试到达目的主机的路由（经过的路由器）：源主机的 Traceroute 程序向目的主机发送一个 Echo Request 报文（包含了一个具有不可达端口号的 UDP 报文段），IP 报头的 TTL 设为 1；第一跳路由器对 TTL 减 1，发现 TTL 变为 0，向源主机发送一个 TTL expired 报文（IP 报头中有路由器的 IP 地址）；Traceroute 记录第一跳路由器的 IP 地址，然后向目的主机发送第二个 Echo Request 报文，IP 报头的 TTL 设为 2；若收到第二跳路由器的 TTL expired 报文，记录第二跳路由器的 IP 地址；接着发送一个 TTL 为 3 的 Echo Request 报文；该过程不断重复，直至收到目的主机发送的 Echo Response 报文（该目的主机向源主机发送一个端口不可达的 ICMP 报文）。**IPv6：最初的动机：**IPv4 地址将很快耗尽；**进一步的动机：**简化头部格式，加快数据报处理和转发；提供服务质量；支持多播；支持移动性；增强安全性。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”；**地址表示的零压缩技术：**可将连续的多组 0 压缩为一对冒号，如以上地址可表示为：“8000::0123: 4567: 89AB: CDEF”。IPv6 定义了**三种地址类型：**单播地址：一个特定的网络接口；多播地址：一组网络接口；任播地址（anycast）：一组网络接口中的任意一个（通常是最近的一个）。**IPv6 数据报格式：**以一个 40 字节的基本头开始，后面跟零个或多个扩展头，然后是数据。**PRI（或 traffic class）：作用：**发送方在该域定义数据报的优先级，路由器检测到网络拥塞时，按优先级从低到高的顺序丢弃包。**IPv6 将网络流量划分为两大类：受拥塞控制的流：**非实时流属于这一类，优先级 0～7，按照重要性及用户体验设定；**不受拥塞控制的流：**实时多媒体流属于这一类，优先级 8～15，尚无标准，可以按照用户要求的服务质量等级定义。**流（flow）：**流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、具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）唯一标识；流标签由发送方分配，不支持流的节点忽略该域；支持流的路由器维护一张**流表（flow table）**，记录每一个流需要的处理；收到数据包后，根据源地址和流标签查找流表，进行相应的处理；流的引入使得 IPv6 具备了对数据包进行区分处理的能力。**IPv6 包格式：**与 IPv4 固定头相比，IPv6 的基本头中**去掉了以下一些字段：**选项：IPv6 的基本头总是 40 字节长（但可以通过“下一个首部”）；与分片相关的字段：IPv6 路由器不负责分片；头校验：计算校验和太花时间，现在的网络非常可靠，并且链路层和传输层上往往都有校验和。IPv6 基本头中**增加了：**流标签：支持对数据包区分处理；**改变了以下字段的作用：**Type of Service：代之以 Traffic Class；总长度：代之以载荷长度；Protocol：代之以 Next header，允许任意扩展选项。**ICMPv6：**ICMPv6 合并了 IPv4 中的 ARP 和 IGMP，并取消了 RARP（该协议的功能已被其它协议取代），ICMPv6 仍然使用差错报告和查询两类报文，为 IPv6 增加了新的类型，如“分组太大”和“未识别的 IPv6 选项”；去掉了源抑制报文，优先级和流标签允许路由器控制拥塞，丢弃不重要的数据包，增加了一些查询报文，用于实现 ARP（地址解析）和 IGMP（多播组管理）的功能。**从 IPv4 过渡到 IPv6：双协议栈方案：**支持 IPv6 的主机和路由器同时运行 IPv4 和 IPv6；运行双栈的源节点先对目的节点查询 DNS：若 DNS 返回 IPv4 地址，发送 IPv4 分组；若返回 IPv6 地址，发送 IPv6 分组；双栈节点同时拥有 IPv4 和 IPv6 地址。**IPv6 数据包如何穿越 IPv4 网络？报头转换：**双栈节点（如路由器 B）在将数据报传递给 IPv4 路由器（如路由器 C）之前，将 IPv6 报头转换成 IPv4 报头；缺点：报头转换不完全，有信息丢失。**建立隧道：**IPv6/IPv4 边界路由器将 IPv6 包封装到一个 IPv4 包中，送入 IPv4 网络，目的边界路由器取出 IPv6 包继续传输；**优点：**保留原始数据报的全部信息。**路由选择算法：什么是最佳路径：**应用评价：路径长度、数据速率、通信费用、安全性等；ISP 关心：网络吞吐量最大、平均包延迟最小、平均通信费用最低、网络负载均衡、路由稳定、健壮等；路由评价指标通常是矛盾的，需要折衷。**选路算法分类：全局算法 or 分布式算法？全局算法:**所有路由器具有关于拓扑和链路代价的全部信息，集中式计算；**分布式算法:**路由器仅知道邻居节点以及到邻居节点的链路代价，通过与邻居交换信息，进行迭代计算。**静态算法 or 动态算法？静态算法:**路由随时间不变或缓慢变化（手工配置）；**动态算法:**路由器根据拓扑及链路代价的变化而自动更新路由。**链路状态（LS）选路算法：链路状态选路算法**为**全局算法**，其**基本思想为：**每个节点利用可靠方法获得全网拓扑信息，抽象成一个带权拓扑图，计算到各个节点的最短路径。链路状态选路算法包括**五个步骤：发现邻居：**有链路直接相连的节点称为邻居，**探测链路代价：**探测到每个邻居的代价，**构造链路状态（LS）分组：**利用邻居及链路代价信息，**扩散 LS 分组：**向网络中所有节点发送 LS 分组（**链路状态广播算法**），**计算路由：**利用收到的 LS 分组构造网络拓扑，计算从本节点到其它各个节点的最短路径（Dijkstra）。**可能出现的问题：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**一个是强制费用不依赖拥塞，一个是确保并非所有路由器都运行 LS 算法。**距离向量（DV）路由选择算法：距离矢量算法：**利用 **Bellman-Ford 方程**求解任意两个节点之间的最小代价路径，主要贡献在于给出了**分布式（迭代、异步地）**求解 B-F 方程的方法。**算法的基本思想:**节点 x 测量其到各个邻居 v 的链路代价 c(x,v)；节点 x 估计其到达各个节点 y 的最小代价 Dx(y)，这些代价构成了自己的距离矢量 Dx=[Dx(y):y∈N]；每个节点周期性地将它的距离矢量 Dx 发送给邻居；节点 x 拥有每个邻居 v 的距离矢量 Dv=[Dv(y):y∈N]；当节点 x 从各个邻居收到它们的距离矢量后，利用 B-F 方程更新自己的距离矢量：Dx(y)←minv{c(x,v)+Dv(y)}。**节点的本地计算由以下两种事件引起:**本地某条链路的代价 c(x,p)发生了变化；收到了某个邻居节点的距离矢量 dp(y)。节点仅在发现距离矢量 dx(y)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；节点仅传播可靠的信息（LS 更健壮）：亲自测量的本地链路代价；节点计算的路由不传播，错误不扩散；**收敛速度：**O(|N||E|)个报文，O(N2)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；节点计算的路由要传播，会造成错误扩散；收敛较慢，还可能出现路由环路、计算至无穷问题。**层次路由选择：平面结构的网络不具有扩放性:**路由器数目扩大，路由表规模，信息交换开销；**网络管理员希望对于网络有更多的控制权（管理自治）：**选路算法的选择，隐藏网络内部组织。**自治系统（AS）：**自治系统（AS）是由处于同一个管理域下的网络和路由器组成的集合；每个 AS 被赋予一个 AS 编号，由 ICANN 分配；同一个 AS 中的路由器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：**在一个 AS 内，直接连接到其它 AS 的路由器；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。**Inter-AS 的任务：**AS1 内部的路由器需发送数据报到 AS1 外，路由器应当向哪个网关路由器发送？AS1 的网关路由器必须了解哪些目的网络通过自己可达&&将可达性信息传播到 AS1 内部的所有路由器。**热土豆路由选择：**AS 尽可能快地扔掉分组，这通过让路由器向某个网关路由器发送分组来完成，同时该网关路由器在到目的地的路径上的所有网关路由器中有最低的路由器到网关的费用。**举例:路由器 1d 设置去往外网 x 的转发项：**1c 和 1b 分别从 3a 和 2a 了解到网络 x 可达，于是各自将“网络 x 可达”的信息传播到 AS1 内部（inter-AS）；1d 决定使用 1c 还是 1b 路由器：**热土豆协议：**选择离 1d 最近的网关路由器（intra-AS）；从转发表中得到该网关的接口 I，添加表项 (x, I) 到转发表中。**涉及两个层次的选路：**通过哪个网关路由器到达外网 x（AS 间选路），如何到达该网关路由器（AS 内选路）。**转发表由 intra-AS 和 inter-AS 配置：**intra-AS：设置到 AS 内部网络的路由；inter-AS & Intra-As：设置到外部网络的路由。**因特网中的路由选择：**Intra-AS 选路协议也称**内部网关协议 IGP**，常见的有:RIP，较低层 ISP 和企业网使用；OSPF，较顶层 ISP 使用。Inter-AS 选路协议也称**外部网关协议 EGP**，目前只有：BGP。**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。**因特网中自治系统内部的路由选择：RIP：**RIP 采用**距离矢量**选路算法。**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。**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为 15 跳。**RIP 通告（RIP 响应报文）：**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 30 秒发送一次 RIP 响应报文。每台路由器维护一个称为**选路选择表**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 180 秒未收到某个邻居的 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：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**，**链路代价：**由管理员配置（反映了管理员的选路策略）。**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 分组被**封装在 IP 包中传输，协议号为 89**。路由器周期性地、或在链路状态改变时发送 **OSPF 链路通告。OSPF 协议负责：**链路通告分组在网络中的广播及可靠传输。路由器根据收到的链路通告分组构造**链路状态数据库**。路由器利用链路状态数据库及 Dijkstra 算法，计算以本路由器为根的最短路径树。**OSPF 最重要的优点是支持 AS 内部的分层选路；**一个 OSPF 自治系统可以配置成多个**区域（area）：**每个区域运行自己的 OSPF 协议，区域内部的链路状态仅在本区域内广播，区域边界路由器负责区域间的选路。一个 OSPF 自治系统配置为若干区域：**一个特殊区域称为主干**，所有区域必须连接到主干上；每一个区域都有**区域标识**，主干的区域标识为 0；**路由器：区域边界路由器：**连接本地区域和主干的路由器；**主干路由器：**主干上的路由器，可以同时是区域边界路由器；**内部路由器:**AS 内部的非区域边界路由器。**分层的 OSPF 路由器。两个选路层次：**本地区域，主干；每个区域（包括主干）运行自己的 OSPF 协议；**每个区域边界路由器：**将本区域的选路信息汇总（子网及路径代价），通告给其它区域，将收到的其它区域的选路信息（子网及路径代价）通告给本区域的内部路由器；**对于去往其它区域的分组：**首先转发到本区域的边界路由器，在主干上转发到目的区域边界路由器，然后再转发到目的子网。**OSPF 的其他优点：**安全，可以同时使用多条费用相同的路径，对单播和多播路由选择的充分支持。**自治系统间的路由选择：BGP：AS 间选路的困难与目标：**因特网规模极其庞大且结构非常复杂；每个 AS 可运行自己的内部路由协议，使用自己的路由测度确定到目的网络的最佳路由，不同网络判断最佳路由的标准不同；一个 AS 可能不信任来自某个 AS 的选路信息；一个 AS 可能不愿意为其它 AS 发数据包；AS 间选路试图找到能够到达目的网络的路由，但还不可能（也不可能）找到最佳路由。**边界网关协议 BGP：**当一对 AS 同意交换选路信息时，每个 AS 指定一个接近 AS 边缘的路由器（或主机），使用 BGP 协议交换选路信息。**运行 BGP 协议的边界路由器（或主机）称为 BGP speaker。**一对 BGP speaker 通过一条**半永久的 TCP 连接（端口 179）**建立 BGP 会话，交换 BGP 报文（BGP 是应用层协议！），BGP 会话的两个端点互为 **BGP 对等方**。不同 AS 的两个边界路由器之间建立的 BGP 会话，称为**外部 BGP（eBGP）会话；**一个 AS 可能有多个边界路由器，这些边界路由器必须通过半永久 TCP 连接成全连通，它们之间的 BGP 会话称为**内部 BGP（iBGP）会话**。**BGP 定义了 4 种类型的报文：**打开报文：BGP 路由器用来启动与邻居 BGP 路由器的联系；保活报文：BGP 路由器定期交换保活报文，告知对方自己处于工作状态；通知报文：当检测到某种差错时，发送通知报文；更新报文：BGP 路由器使用该报文宣布新路由，以及撤销以前通告的路由。**可达性信息：**以 AS 枚举形式通告的、到达**目的前缀**的完全路径（便于检测路径环）。路由器收到相邻 AS 的路由通告，在向下一个 AS 发送该路由之前，**将自己的标识及 AS 号加入到完全路径中。边界路由器的选路表中，每个表项包含**目的前缀（允许聚合）、下一跳路由器以及到达目的前缀的 AS 序列。**以 AS 枚举形式通报完全路径：**AS2 的 BGP speaker 通报：128.96, 192.4.153, 192.4.32, 192.4.3 可经由<AS2>到达；AS1 的 BGP speaker 收到后通报：128.96, 192.4.153, 192.4.32, 192.4.3 可经路径<AS1, AS2>到达。**BGP speaker 内部的选路信息库**由三部分组成：Adj-RIBs-In：每个 Adj-RIBs-In 对应一个 BGP 对等方，保存从该 BGP 对等方收到的选路信息；LOC-RIB：已被该 BGP speaker 计算出来的最佳路由；Adj-RIBs-Out：每个 Adj-RIBs-Out 对应一个 BGP 对等方，存放准备向该 BGP 对等方通告的选路信息。**BGP 进程的处理过程：**接收从各个 BGP 对等方发来的更新报文，更新与之相对应的 Adj-RIB-In（添加、替换或删除路由）；输入策略引擎对 Adj-RIB-In 中每条新的路由进行入境过滤（过滤规则由网络管理员定义），结果可能是：丢弃，按原样接受，接受但修改某些属性（如偏好度）；对于每一个目的前缀，从所有可达的路径中按照 BGP 指定的决策顺序确定一条最佳路由，装入 LOC-RIB；输出策略引擎根据出境过滤规则（由管理员定义），计算要通告给每一个 BGP 对等方的路由更新，放入对应的 Adj-RIB-Out 中（路由聚合也在这个阶段完成）；BGP 进程利用 Adj-RIB-Out，向每个 BGP 对等方发送路由更新报文。**Intra-AS 和 Inter-AS 选路协议：**Intra-AS 选路协议:用于在 AS 内部交换选路信息，如 OSPF、RIP，使用某个路由测度（代价）选择到目的节点的最优路径；Inter-AS 选路协议:用于在不同的 AS 之间交换选路信息，如 BGP，主要依据策略而不是路由测度去寻找可达路径（不追求最佳路径）。**为什么会有不同的 AS 间和 AS 内部路由选择协议：**对该问题的答案触及了 AS 内与 AS 间的路由选择目标之间差别的本质：策略（AS 间在意）、规模（AS 间在意，而 AS 内可以进一步划分）、性能（AS 内关心）。**广播和多播路由选择：广播路由选择(broadcast routing):**网络层提供了从一种源结点到网络中的所有其他结点交付分组的服务；**多播路由选择(multicast routing)**使单个源结点能够向其他网络结点的一个子集发送分组的副本。**广播路由选择算法：**用 **N 次单播**实现广播（在源节点上复制分组）：低效：相同的分组在某些链路上可能重复传输，需要其它机制支持：源节点需知道所有目的节点的地址。**理想的广播选路：**源节点不需知道其它节点的地址，只需将分组的目的地址设置为广播地址，路由器负责转发到全网（在网络中复制分组）；网络中产生的分组拷贝最少。**在网络中复制分组：洪泛（flooding）：**节点收到广播分组后，向所有邻居节点（分组到来的链路除外）发送该分组的拷贝；缺点：在有环的网络中，广播分组在网络中无休止地循环，浪费资源。**受控洪泛：**目标：每个路由器仅转发它之前未转发过的广播分组；**两种方法：**节点记录已经转发过的分组 ID，不重复转发分组（OSPF 使用此方法：源地址+分组 ID）；**反向路径转发:**利用节点内部的单播转发表，仅转发从本节点->源节点的最短路径的反向路径上到来的广播分组（该方法使用最多）。**反向路径转发 RPF 基本思想：**当广播分组到达路由器时，路由器检查分组的源地址与输入端口；用分组的源地址查找单播路由表，找到去往该源地址的输出端口；若分组的输入端口与去往该地址的输出端口相同，则扩散该分组，否则丢弃分组；**优点：**算法合理，易于实现且开销不大。**生成树方法：**使用生成树发广播分组：路由器知道自己的哪几个端口在生成树上；当从一个端口收到广播分组后，只在属于生成树的其它端口上转发该分组；**基于生成树的广播不会产生冗余的分组拷贝。生成树的构造：基于核心的方法：**选择一个节点作为核心（也称汇聚点）；其它节点向核心发送单播的加入报文：路由器利用单播转发表向核心转发加入报文，记录报文的输入端口及输出端口，这些端口就是位于生成树上的端口；当加入报文到达生成树上的一个节点时，报文经过的路径被添加到生成树上。**实践中的广播算法：应用层：**Gnutella，TCP，寿命 TTL，范围受限的洪泛；**OSPF 序号控制洪泛**（32bit 序号+16bit 年龄）**多播：如何标识多播通信的接收者：**因特网为这组接收者分配一个标识（多播组标识），使用 D 类地址作为多播组标识；**如何设置多播分组的接收者：**将分组的目的地址设置为其接收者的多播组地址；**如何将接收者的 IP 地址与多播组地址关联起来：**接收者的 IP 地址与多播组地址无关，接收者可以在任何时候加入或离开一个组，**多播组管理协议（IGMP）**允许主机向本地路由器申请加入或离开一个组；**如何将多播分组交付给每一个接收者：**多播选路协议协调多播路由器建立到达所有接收者的路径树。**多播组管理 IGMP 协议**运行在主机与边缘路由器之间：主机利用 IGMP 协议向边缘路由器请求加入一个组、或离开一个组；边缘路由器利用 IGMP 协议向主机询问组成员关系；边缘路由器通过 IGMP 协议可以了解到，在它的某个端口连接的网络上存在着哪些组的成员。由于 IGMP 的交互范围被局限在主机与其相连的路由器之间，需要另一种协议来协调遍及因特网内的多播路由器（包括相连的路由器），以便多播数据报能路由到其最终目的地。后一个功能是由**网络层多播路由选择算法**完成的。**因特网中的网络层多播是由两个互补的组件组成的：**IGMP 与多播路由选择协议。**IGMP 报文格式与类型：**报文承载在 **IP 数据包中，IP 协议号为 2，TTL = 1**(IGMP 只在本地工作)；**报文类型：**查询（一般查询，特殊查询），成员报告，退出报告（可选的），软状态：使用了成员报告报文：当无主机响应一个具有给定组地址的成员报告报文时，该路由器就推断出已没有主机在这个多播组了。**软状态协议要比硬状态协议的控制简单一些**后者不仅要求状态显式地增加和删除，还要求有机制能够负责当实体提早结束或失败时清除状态。**加入一个组：**每个主机维护一张应用进程与多播组的对应表；主机上的一个应用进程想要加入一个多播组时，向主机发送请求；若这是一个新组，主机向其边缘路由器发送一个成员关系报告。**退出一个组：**主机发现某个多播组为空时，从表中清除该组，向路由器发送退出报告；路由器收到退出报告后：在该网络接口上发送针对该多播组的特殊查询报文（询问该端口上还有这个组的成员吗？）若在规定的时间内没有收到该多播组的成员关系报告，从表中清除该组。**监视成员关系：**路由器周期性发送通用查询报文（组地址置为 0），主机发送成员关系报告作为响应。**封装 IGMP 报文的 IP 包使用多播地址作为目的地址：**查询报文：多播地址 224.0.0.1，接收者是该子网上的所有节点；退出报告：多播地址 224.0.0.2，接收者是该子网上的所有路由器；成员关系报告：被报告的多播地址。**多播路由选择算法：目标：**为每个组建立多播转发树（到达该组所有成员的路径树），每个组成员应当只收到多播分组的一个拷贝，非本组成员不应收到多播分组，从源节点到每一个组成员节点的路径应当是最佳的（最短路径）。**建立多播树的两种方法：基于源的树:**源节点建立一棵到达多播组所有成员的最短路径树，源节点 S 和组 G 的每一种组合<S,G>构成一棵树，多播路由器必须有每一个<S,G>树的信息，根据多播分组的<S,G>确定使用哪棵多播树，**优点：**总是使用最佳路径转发多播分组，**缺点：**路由器需要维护大量的多播树信息。**组共享树:**每个多播组使用一棵树，树根为该多播组的核心，源节点先将多播分组发送给核心，核心再在多播树上发送，**优点：**对于每个组，多播路由器只需维护一棵多播树，**缺点：**多播分组使用的路径可能不是最佳的。**基于源的树：最短路径树：MOSPF（多播 OSPF）**通过**扩展 OSPF 协议**实现最短路径多播选路：扩展链路状态，使之包含链路上的组成员关系；**基本思想：**所有参与多播的主机在局域网上定期通报其所属的多播组（IGMP），路由器将每条直连链路上对应的多播组集合作为链路状态在网上广播（MOSPF），当路由器第一次遇到某个<S,G>多播分组时，计算从源节点 S 到多播组 G 的最短路径多播树（按需计算）。**基于源的树：距离矢量多播选路：**扩展 RIP 协议实现多播选路的困难：除边缘路由器外，其它路由器不知道多播组（及其成员）的存在；**DVMRP 采用广播+剪枝的解决办法：**RPF 广播：确保多播分组到达每一个局域网；路径剪枝：路由器删除不包含组成员的路径分支；参与多播的主机定期在局域网上通报所属的多播组，局域网上的路由器记录这些信息（IGMP）；当路由器收到发往组 G 的多播分组，但它并没有从局域网上监听到组 G 的报告时，向上游路由器发送一个剪枝报文，上游路由器停止通过这个接口发送该组的多播分组；如果一个路由器从它的每个下游路由器都收到剪枝报文，路由器向其上游路由器转发剪枝报文；该过程递归进行，直至所有的无关分支都被删除，最终得到一棵<S,G>树。**组共享树：基于核心的树：**指定一个路由器作为组 G 的核心，所有路由器知道该核心所属的组及单播 IP 地址（需要其它机制），多播路由器向核心发送单播加入报文，当报文到达核心或已在树上的节点时，报文经过的路径加入到树中。**组共享树的构造过程：**希望加入多播组 G 的路由器 S 向组 G 的核心发送单播的加入报文，收到加入报文的路由器按照单播选路表向核心转发加入报文，并在多播转发表中创建一条共享树记录<*, G>：报文到达的接口：标记为转发 G 的多播分组的接口；向核心进一步转发报文的接口：标记为允许接收 G 的多播分组的唯一接口；当加入报文到达树上的某个节点或核心时，报文经过的路径被添加到树上。**如何利用组共享树发送多播分组？**当源节点想要发送多播分组时：源节点将多播分组发送给核心，核心在多播树上发送；**多播分组如何到达核心？**多播分组的目的地址为 G，从源节点到核心的路径上，可能有路由器不在多播树上（不知道如何转发）；**建立隧道：**源节点将多播分组封装到一个单播分组中，单播分组的目的地址为核心地址。**广泛使用的因特网多播选路协议是 PIM：**不依赖于网络中所使用的单播选路协议。**PIM 有两种工作模式：**稠密模式：只许多或大多数路由器涉及多播选路过程，使用广播+剪枝方式建立多播树；稀疏模式：只有很小一部分路由器涉及多播选路过程，采用组共享树方式；当源节点流量很高时切换到源树。**多播分组穿越单播网络：**因特网中只有一小部分路由器是多播路由器，多播分组在从一个多播路由器传递到另一个多播路由器时，通常需要穿越单播网络；在多播路由器之间建立隧道：把多播分组封装在单播分组中传输。因特网中的多播路由器以及这些多播路由器之间的隧道，构成了**因特网多播骨干网**。

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：将数据报从一个结点传输到相邻的下一个结点，**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。**链路层概述：结点：**运行链路层协议的任何设备。**链路：**连接相邻节点的通信信道。**帧：**链路层分组称为链路层帧。**链路层提供的服务：链路层服务：组帧（基本服务）**从原始比特流中提取出完整的帧；**链路接入（广播链路需要）**在广播信道上协调各个结点的发送行为；**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（不使用重传）；**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；**流量控制:**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；**半双工和全双工：**半双工通信时，提供收/发转换。**链路层在何处实现：路由器中**链路层在**线卡（line card）**中实现，**主机链路层主体部分**在**网络适配器（网卡）**中实现。线卡/网络适配器连接传输媒体，还实现物理层的功能。**链路层由硬件和软件实现：**网卡中的控制器芯片：组帧、链路接入、检错、可靠交付、流量控制等；主机上的链路层软件：与网络层接口，激活控制器硬件、响应控制器中断，处理差错条件和将数据报向上传递给网络层。**网络适配器之间的通信：发送侧:**将数据报封装到帧中，生成校验比特，执行可靠传输和流量控制；**接收侧：**提取帧，检测传输错误；执行可靠传输和流量控制；解封装数据报，交给上层协议。**差错检测和纠正技术：传输出错的类型：单个错：**由随机的信道热噪声引起，一次只影响 1 位；**突发错：**由瞬间的脉冲噪声引起，一次影响许多位，使用**突发长度**表示**突发错影响的最大数据位数**。**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：**能够检测错误并自行纠正错误的编码。**码字（codeword）：**由 m 比特的数据加上 r 比特的冗余位（校验位）构成。**有效编码集：**由 2^m 个符合编码规则的码字组成。**检错：**若收到的码字为无效码字，判定出现传输错误。**海明距离（Hamming Distance）：**两个码字的对应位取值不同的位数。**纠错：**将收到的无效码字纠正到距其最近的有效码字。**检错码与纠错码的能力都是有限的！编码集的海明距离：**编码集中任意两个有效码字的海明距离的最小值。**检错能力：为检测出所有 d 比特错误，编码集的海明距离至少应为 d+1。纠错能力：为纠正所有 d 比特错误，编码集的海明距离至少应为 2d+1。差错检测的实施：**发送端对要保护的数据 D（包括帧头字段）生成校验位 **EDC（差错检测和纠正比特）**，添加在帧头中；接收端对收到的数据 D'计算校验位 EDC'，根据 EDC'判定是否有错。**奇偶校验：单比特奇偶校验：**可检测单比特错，检错率为 50%，编码集海明距离为 2。**二维奇偶校验：**可检测 2 比特错和纠正单比特错，编码集海明距离为 3，有利于检测突发错。**前向纠错(FEC)：**接收方检测和纠正差错的能力；**优点：**减少所需的发送方重发的次数，允许在接收方立即纠正差错，避免了不得不等待的往返时延，而这些时延是发送方收到 NAK 分组并向接收方重传分组所需要的。**因特网检验和(Internet checksum)：**数据的字节作为 16 比特的整数对待并求和（溢出回卷），这个和的反码形成了携带在报文段首部的因特网检验和；TCP 和 UDP 对所有字段（包括首部和数据字段）计算因特网检验和；**优势：**检验和方法需要相对小的分组开销，与 CRC 相比提供相对弱的差错保护。**为什么运输层使用检验和而链路层使用 CRC：**运输层通常是在主机中作为用户操作系统的一部分用软件实现的。因为运输层差错检测用软件实现，采用简单而快速如检验和这样的差错检测方案是重要的。在另一方面，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。**循环冗余检验（CRC）：CRC** 是一种**多项式编码**，它将一个位串看成是某个一元多项式的系数，如 1011 看成是一元多项式 X3 + X + 1 的系数。**信息多项式 M(x)：**由 m 个信息比特为系数构成的多项式。**冗余多项式 R(x)：**由 r 个冗余比特为系数构成的多项式。**码多项式 T(x)：**在 m 个信息比特后加上 r 个冗余比特构成的码字所对应的多项式，表达式为 $T(x) = x^r * M(x) + R(x)$。**生成多项式 G(x)：**双方确定用来计算 R(x)的一个 **r+1 比特**多项式。**编码方法：**$R(x) = x^r * M(x) ÷ G(x)$ **的余式（减法运算定义为异或操作）**。**检验方法：**若 T(x) ÷ G(x)的余式为 0，判定传输正确。CRC 的检错能力极强，可用硬件实现，是应用最广泛的检错码。**CRC 举例：取 G(X) = X^3 + 1，对信息比特 101110 计算 CRC 码。101110000 ÷ 1001 的余式为 R=011 (CRC code)，码字：101110011；取 G(X) = X^3 + 1，接收端收到比特串 1001001，问是否有错？解答：1001001÷1001 的余式为 001（不为 0），有传输错误。多路访问链路和协议：链路的两种类型：点到点链路：**仅连接了一个发送方和一个接收方的链路，一条全双工链路可以看成是由两条单工链路构成的；**广播链路：**连接了许多节点的一条共享链路，任何一个节点发送的数据可被链路上的其它节点接收到。**冲突（collision）：**在广播链路上，若两个或多个节点同时发送，发送的信号会发生干扰，导致接收失败。**多址接入协议（多路访问协议）：**规定节点共享信道（谁可以发送）的方法，多址接入协议也称**媒体接入控制（MAC）协议**。**理想的多址接入协议：**在速率为 R bps 的广播信道上，1. 当只有一个节点发送时，它能以速率 R 发送（信道利用率高）2. 当有 M 个节点发送时，每个节点应能以 R/M 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:不需要一个特殊的节点来协调发送（健壮性好），不需要时钟和时隙同步（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，每个节点分配一个子信道，不会发生冲突，关注公平性，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，有发送的节点轮流发送，不会出现冲突，信道利用率是以上两种方法的折衷，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**时间片（一个时间帧 N 个时隙）**，每个时间片可以发送一个分组，节点只能在分配给自己的时间片内发送，若无数据发送，其时间片轮空。**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（R/N 带宽），若节点不发送，其子频带空闲。**TDM 和 FDM 的优劣：**消除碰撞而且非常公平，但结点被限制于 R/N bps 的平均速率，必须总是等待它在传输序列中的轮次。**CDMA（码分多址）：**将每个比特时间进一步划分为 m 个微时隙（称 chip），每个节点被分配一个惟一的 m 比特码序列（称 chip code）；**发送方编码：**发送“1”=发送 chip code，发送“0”=发送 chip code 的反码；**信号叠加：**多个节点发送的信号在信道中线性相加；**接收方解码：**用发送方的 chip code 与信道中收到的混合信号计算内积，恢复出原数据；**前提条件：**任意两个 chip code 必须是相互正交的；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；**随机接入 MAC 协议**规定如何检测冲突，以及如何从冲突中恢复。**随机接入 MAC 协议的例子：发送前不监听信道：**ALOHA 家族；**发送前监听信道：**CSMA 家族。**时分（Slotted）ALOHA：**假设：所有帧长度相同；时间被划分为等长的时隙，每个时隙传一帧；节点只在时隙开始时发送（节点的时钟同步（知道时隙何时开始））；所有节点可在时隙结束前检测到是否有冲突发生。**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 P 重传，直至发送成功。**优点：**单个活跃节点可以信道速率连续发送，分布式：节点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：**当网络中存在大量活跃节点且每个节点总有大量的帧待发送时，长期运行过程中**成功时隙**所占的比例。**时分 Aloha 的效率：**假设：有 N 个活跃节点，每个节点在每个时隙开始时以概率 p 发送；给定节点在一个时隙中发送成功的概率 = p(1-p)N-1；给定时隙中有节点发送成功的概率 = Np(1-p)N-1；最大效率：找到令 Np(1-p)N-1 最大的概率 p*；代入 Np*(1-p*)N-1，并令 N 趋向于无穷，得到：**最大效率 = 1/e = 0.37。纯 ALOHA：基本思想：**取消同步时，任何节点有数据发送就可以立即发送，节点通过监听信道判断本次传输是否成功，若不成功，立即以概率 P 重传，以概率（1-P）等待一个帧时后再决定。（**帧时：**发送一帧的时间，假设帧长度相同）。**发生冲突的情形：**在时刻 t0 发送的帧与在 **[t0 - 1,t0 + 1]** 时段内发送的其它帧冲突。**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[t0-1,t0]内发送)*P(无其它节点在[t0, t0+1]内发送)=p(1-p)N-1*(1-p)N-1=p(1-p)2(N-1)；求出令节点发送成功概率 Np(1-p)2(N-1)最大的 p*，并令 N→infty:**最大效率 = 1/(2e) = 0.18。载波侦听多址接入（CSMA）：两个重要的规则：**发送前监听信道（carrier sensing **载波侦听**），信道空闲发送整个帧，信道忙则等待至少一段时间没有传输再发送（推迟发送）；**碰撞检测**，即当一个传输结点在传输时一直在侦听此信道，如果它检测到另一个结点正在传输干扰帧，它就停止传输，在重复“侦听－当空闲时传输”循环之前等待一段随机时间。这两个规则包含在**载波侦听多路访问(CSMA)**和**具有碰撞检测的 CSMA(CSMA/CD)**协议族中。**冲突仍可能发生：**由于存在端到端信道传播时延，节点可能没有监听到正在发送的帧；即使忽略传输延迟，当两个（或多个）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其性能方面起关键作用。**CSMA/CD (Collision Detection)：**若在发送的过程中检测到冲突，怎么办？继续发送余下的部分（浪费带宽）或停止发送余下的部分。CSMA/CD 的**基本思想：**在发送的过程中检测冲突（发生冲突时信号较强）；检测到冲突后，立即停止发送剩余的部分；立即启动冲突解决的过程。**以太网采用 CSMA/CD 协议：1.**NIC 从网络层接收数据报，构造以太帧；**2.**若 NIC 监听到信道空闲，立即发送帧；若信道忙，坚持监听直至发现信道空闲，然后发送帧；**3.**若 NIC 发送完整个帧而没有检测到冲突，认为发送成功！；**4.**若 NIC 在传输过程中检测到冲突，立即停止发送帧，并发送一个阻塞信号（加强冲突）；**5.**中止传输后等待一个随机时间量（指数回退）然后返回步骤 2。**我们希望时间间隔这该样：**当碰撞结点数量较少时，时间间隔较短；当碰撞结点数量较大时，时间间隔较长。**二进制指数后退算法：**当传输一个给定帧时，在该帧经历了一连串的 n 次碰撞后，结点随机地从 { 0, 1, 2, 3,..., 2^n-1 } 中选择一个 K 值，对于以太网，一个结点等待的实际时间是 K*512 比特时间，n 能够取的最大值在 10 以内。**指数回退的目的是根据**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。每次适配器准备传输一个新的帧时，它要运行 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结点能够立刻插入一次成功的传输。**CSMA/CD 效率：Tprop** = 以太网中任意两个节点之间传播延迟的最大值；**Ttrans** = 最长帧的传输时间；**efficiency = 1 / (1 + 5*Tprop/Ttrans)。在以下情况下，以太网的效率趋近于 1：**Tprop 趋近于 0，或 Ttrans 趋向于无穷。**结论：**应控制以太网的规模。**轮流 MAC 协议：**多路访问协议的两个理想特性是：1. 当只有一个结点活跃时，该活跃结点具有 R bps 的吞吐量；2. 当有 M 个结点活跃时，每个活跃结点的吞吐量接近 R/M bps。**ALOHA 和 CSMA 协议具备第一个特性，但不具备第二个特性。轮询协议：**结点之一被指定为主结点，主结点以循环的方式轮询每个结点，主结点首先向结点 1 发送一个报文，告诉它能够传输的帧的最多数量，在结点 1 传输了某些帧后，主结点告诉结点 2 它能够传输的帧的最多数量（主结点能够通过观察在信道上是否缺乏信号，来决定一个结点何时完成了帧的发送）。**优势：**消除了碰撞和空时隙，使得轮询取得高得多的效率；第一个缺点是引入了轮询时延，即通知一个结点“它可以传输”所需的时间，例如，如果只有一个结点是活跃的，那么这个结点将以小于 R bps 的速率传输，因为每次活跃结点发送了它最多数量的帧时，主结点必须依次轮询每一个非活跃的结点；第二个缺点是单点失效。**令牌传递协议：**没有主结点，一个称为**令牌(token)**的小的特殊帧在结点之间以某种固定的次序进行交换；当一个结点收到令牌时，仅当它有一些帧要发送时，它才持有这个令牌；否则，它立即向下一个结点转发该令牌；一个结点收到令牌时，若它确实有帧要传输，它发送最大数目的帧数，然后把令牌转发给下一个结点。**优势：**令牌传递是分散的，并有很高的效率；一个结点的故障可能会使整个信道崩溃；如果一个结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；令牌传递延迟；令牌单点失效。**MAC 协议比较：信道划分 MAC 协议:**重负载下高效：没有冲突，节点公平使用信道；轻负载下低效：即使只有一个活跃节点也只能使用 1/N 的带宽。**随机接入 MAC 协议：**轻负载时高效：单个活跃节点可以使用整个信道；重负载时低效：频繁发生冲突，信道使用效率低。**轮流协议（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）。**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），要能支持数据、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（百公里以上），规模和容量可任意扩大。**链路层寻址和 ARP：**每一块网络适配器（网卡）固定分配一个地址，称为**物理地址、硬件地址、链路层地址、MAC 地址**等。**MAC 地址长** 6 个字节；由 IEEE 负责分配，每块适配器的地址是全球唯一的：网卡生产商向 IEEE 购买一块 MAC 地址空间（前 3 字节），生产商确保生产的每一块网卡有不同的 MAC 地址；MAC 地址固化在网卡的 ROM 中；现在用软件改变网卡的 MAC 地址也是可能的。主机或路由器的适配器（即网络接口）具有链路层地址，因此，具有多个网络接口的主机或路由器将具有与之相关联的多个链路层地址；**链路层交换机并不具有与它们的接口（这些接口是与主机和路由器相连的）相关联的链路层地址**，这是因为链路层交换机的任务是在主机与路由器之间承载数据报；交换机透明地执行该项任务，这就是说，主机或路由器不必明确地将帧寻址到其间的交换机。适配器的 MAC 地址具有**扁平结构（这与层次结构相反）**，而且不论适配器到哪里用都不会变化。**目的 MAC 地址**有三种类型：**单播地址：**适配器的 MAC 地址，地址最高比特为 0；**多播地址：**标识一个多播组的逻辑地址，地址最高比特为 1；**广播地址：**全 1。网络适配器仅将发送给本节点的帧交给主机：目的地址为适配器 MAC 地址的单播帧，所有广播帧，指定接收的多播帧（若将适配器设置成**混收模式**，适配器将收到的所有帧交给主机）。**主机和路由器接口除了网络层地址之外还有 MAC 地址，这有如下几个原因：**局域网是为任意网络层协议而设计的， 而不只是用于 IP 和因特网；如果适配器使用网络层地址而不是 MAC 地址的话，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，另一种选择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，然后网络层则能够核对网络层地址是否匹配，这种选择带来的一个问题是，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结点的帧中断。为了使网络体系结构中各层次成为极为独立的构建模块，不同的层次需要有它们自己的寻址方案。**如何实现直接交付？**当发送节点 A、接收节点 B 位于**同一个物理网络**上时，数据报可从 **A 直接交付给 B：**A 的网络层将数据报、以及 B 的物理地址交给数据链路层；数据链路层将数据报封装在一个链路层帧中，帧的目的地址=B 的物理地址，B 的适配器收到帧，根据目的地址判断是发给本机的，取出数据报交给网络层。**发送节点如何获得接收节点的物理地址？地址解析（Address Resolution）：问题：已知 IP 地址，如何得到对应的 MAC 地址？静态映射 IP 地址-MAC 地址的缺点：**主机每次使用的 IP 地址可能不同（DHCP），主机可能更换网卡。**地址解析协议（ARP）**用于动态获得 IP 地址-MAC 地址映射，其**基本思想是：**若节点 A 希望获得节点 B 的 MAC 地址，节点 A 广播 B 的 IP 地址（地址解析请求），节点 B 用自己的 MAC 地址进行响应。**ARP 报文格式：硬件类型：**硬件接口类型；对于以太网，该值为“1”。**协议类型：**高层协议地址类型。对于 IP 地址，该值为 080016。**操作：**ARP 请求为 1，ARP 响应为 2。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 想知道 B 的 MAC 地址：**1.** A 构造一个 ARP 请求，在发送方字段填入自己的 MAC 地址和 IP 地址，在目标字段填入 B 的 IP 地址；**2.** A 将 ARP 请求封装在广播帧中发送；**3.** 每个收到 ARP 请求的节点用目标 IP 地址与自己的 IP 地址比较，地址相符的节点进行响应（B 响应）。**4.** B 构造一个 ARP 响应，交换发送方与目标字段内容，在发送方硬件地址字段填入自己的 MAC 地址，修改操作字段为 2；**5.** B 将 ARP 响应封装在单播帧（目的地址为 A 的 MAC 地址）中发送。**改进 ARP 的措施：ARP 缓存：**每个节点在内存中维护一个地址映射（绑定）表，称 ARP 缓存；在每次发送数据报前先查询 ARP 缓存，若找不到则发送 ARP 请求，并在收到 ARP 响应后将地址映射缓存起来；ARP 缓存中的信息，在超时（一般为 15～20 分钟）后删除。**主动学习：**从 ARP 请求中获取地址绑定信息：每个节点可以收到全部的 ARP 请求报文，可将发送节点的地址映射缓存到自己的 ARP 表中；节点在启动时自动广播自己的地址映射：节点 A 在启动时主动广播一个 ARP 请求，在目标字段内填入自己的 IP 地址，收到 ARP 请求的节点将 A 的地址映射缓存起来，若 A 收到 ARP 响应，报告 IP 地址重复错误。**ARP 是即插即用的**，这就是说，一个 ARP 表是自动建立的——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。然而，一个 ARP 分组具有链路层地址的字段，因而可认为是链路层协议，但它也包含网络层地址，因而也可认为是为网络层协议。**所以，可能最好把 ARP 看成是跨越链路层和网络层边界两边的协议**，即不完全符合我们在第 1 章中学习的简单的分层协议栈。**数据报到达子网之外：**数据报从 A（子网 1）经过 R（路由器）到达 B（子网 2）：A 知道下一跳地址为 111.111.111.110（R-1），R 知道 B 从其端口 R-2 直接可达；A 创建 IP 数据报，src IP = A，dest IP = B；A 利用 ARP 获得下一跳 111.111.111.110 对应的 MAC 地址（R-1）（而非获得 B 的 MAC）；A 创建链路层帧，封装 IP 数据包，src MAC = A，dest MAC = R-1，发送；R 接收帧，取出 IP 数据报，发现目的地址为 B；R 利用 ARP 获得 B 的 MAC 地址；R 创建链路层帧，封装 IP 数据报，src MAC=R-2, dest MAC = B，发送；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，而 ARP 只为在同一个子网上的主机和路由器接口解析 IP 地址。**以太网：**第一个广泛应用的局域网技术，也是目前占主导地位的有线局域网技术；与其它局域网技术相比，技术简单、成本低；为提高速率，以太网技术不断演化和发展。**总线拓扑：共享式以太网：总线（1970s 中期）：**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；**集线器（hub，物理层）（1990s 后期）：**一个物理层中继器，从一个端口进入的物理信号（光，电），放大后立即从其它端口输出；集线器相当于共享电缆。**星型拓扑：交换式以太网：交换机（21 世纪早期）：**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！星型拓扑：**各节点仅与中心节点直接通信，各节点之间不直接通信；不同于基于集线器的星型连接。**以太网帧结构：(按顺序)前同步码：**7 个 10101010 字节，后跟一个 10101011 字节，用于在发送方和接收方之间建立时钟同步。**目的地址（6 字节）+源地址（6 字节）。Type（2 字节）：**指出 Data 所属的高层协议（如 IP、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、运输层报文段的端口字段类似）。**Data：**46～1500 字节，不足 46 字节填充至 46 字节；以太网的最大传输单元(MTU)是 1500 字节，这意味着如果 IP 数据报超过了 1500 字节，则主机必须将该数据报分片；如果 IP 数据报小于 46 字节，数据报必须被填充到 46 字节，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（4 字节，循环冗余检测）：**对 dest addr.、src addr.、type 和 data 四个字段计算得到的 CRC 码。**所有的以太网技术都向网络层提供无连接服务**，意味着没有握手。**以太网技术都向网络层提供不可靠服务**，对该帧执行 CRC 校验，但是当该帧通过 CRC 校验时它既不发送确认帧；而当该帧没有通过 CRC 校验时它也不发送否定确认帧，当某帧没有通过 CRC 校验，适配器 B 只是丢弃该帧。（在链路层）缺乏可靠的传输有助于使得以太网简单和便宜。但是它也意味着传递到网络层的数据报流能够有间隙。**为什么有最小帧长的要求？**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：节点检测冲突需要时间，假设信号在相距最远的两个适配器之间的往返延迟为 2τ，则帧的发送时间不应小于 2τ，即帧的最小长度≧链路速率×2τ；为什么最小帧长为 64 字节（不包括前导码）：根据早期以太网的最大直径（2500 米）和数据速率（10Mbps）计算得到。**802.3 以太网标准：链路层 & 物理层：**历史上出现过许多不同的以太网技术：链路层相同：MAC 协议，帧格式，帧处理；物理层不同：传输媒体：光纤，同轴电缆，双绞线；数据速率：如 10Mbps，100Mbps，1Gbps；物理层编码方式不同。所有这些以太网技术由 IEEE 802.3 工作组标准化，形成 IEEE 802.3 标准族。**10Mbps 以太网（早期以太网）：**10Base-5：**基带**同轴电缆（粗），每段电缆最大长度 **500** 米；10Base-2：基带同轴电缆（细），每段电缆最大长度约 **200** 米；10Base-T：3 类**双绞线**和集线器，双绞线最大长度 100 米；10Base-F：多模**光纤**和集线器，光纤最大长度 2000 米。**100Mbps 以太网（快速以太网）：**仅能使用光纤/双绞线，以及集线器/交换机：100Base-TX（可使用集线器或交换机）：5 类双绞线（2 对），不超过 100 米；100Base-T4（可使用集线器或交换机）：3 类双绞线（4 对），不超过 100 米；100Base-FX（只能使用交换机）：多模光纤（2 条），不超过 2000 米。**千兆、万兆以太网：**只用交换机，并增加了对流量控制的支持：1000Base-SX：多模光纤，不超过 550 米；1000Base-LX：单模或多模光纤，不超过 5000 米；1000Base-CX（很少用）：2 对屏蔽双绞线，不超过 25 米；1000Base-T：4 对 5 类双绞线，不超过 100 米；10GBase-T：只使用光纤，长距离用单模光纤，短距离用多模光纤。**DIX 以太帧与 802.3 帧：**最早提出的以太帧称为 DIX（DEC-Intel-Xerox）以太帧：type：指出处理 data 域的协议实体；符合 IEEE 802.3 标准的帧（802.3 帧）：length：替代 DIX 帧中的 type 域，指出 data 的长度；这两种帧格式都可使用，当 type/length 的值大于 1500 时解释为 type，否则解释为 length。**讨论：共享式以太网和交换式以太网：共享式以太网：**集线器的所有端口位于同一个冲突域，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：交换机的每个端口为一个冲突域，多对端口可以同时通信，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：[illegible]

（一个中央基站）；每个无线接口（终端及 AP）均有一个全局唯一的 MAC 地址。**802.11 信道与关联：**802.11 将通信频段划分成若干信道，每个 BSS 分配一个信道：管理员安装 AP 时，为 AP 分配一个*服务集标识符（SSID）*，并选择 AP 使用的信道；相邻 AP 使用的信道可能相互干扰。**主机必须与一个 AP 关联：**扫描 11 个信道，监听各个 AP 发送的信标帧（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 获得 AP 所在子网中的一个 IP 地址。**802.11 主动/被动扫描：被动扫描：**主机扫描信道和监听 AP 发送的信标帧，主机选择一个 AP 发送关联请求帧，AP 向主机发送关联响应帧；**主动扫描：**主机广播探测请求帧，AP 发送探测响应帧，主机从收到的探测响应中选择一个 AP 发送关联请求，AP 发送关联响应帧。**802.11 MAC 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（接收信号的强度远小于发送信号的强度），不能检测出所有的冲突（隐藏节点）。**目标：**避免冲突 CSMA/C(ollision)A(voidance)。**以太网和 802.11 都使用载波侦听随机接入，但这两种 MAC 协议有重要的区别：**802.11 使用碰撞避免而非碰撞检测；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。**802.11 的操作模式：PCF（Point Coordination Function）模式：**该模式只能用于有基础设施（基站）的无线网络，由基站控制单元内的所有通信活动；**轮询：**基站依次询问单元中的节点，被询问到的节点可以发送它们的帧，不会有冲突发生；**新节点注册：**新加入的节点可以注册一个恒定速率的轮询服务，声明自己希望得到的带宽；PCF 的实现是可选的。**DCF（Distributed Coordination Function）模式：**可用于有基础设施的无线网络和无基础设施的无线网络，所有实现必须支持 DCF 模式；所有节点（AP 和无线终端）使用 CSMA/CA 协议竞争信道。**CSMA/CA 支持两种机制：**信道预约机制（可选），无信道预约的机制。**使用信道预约机制的 CSMA/CA：**假设 A 欲向 AP 发送一个数据帧：A 向 AP 发送一个 RTS 帧，帧中给出随后要发送的数据帧及确认帧需要的总时间；AP 收到后回复一个 CTS 帧，帧中给出同样的时间；A 收到 CTS 帧后开始发送；AP 收到帧后，发送一个 ACK 帧进行确认；（A 附近）收到 RTS 帧及（AP 附近）收到 CTS 帧的节点均沉默指定的时间，让出信道让 A 和 AP 完成发送；若 A 和 B 同时发送 RTS 帧，产生冲突，不成功的发送方随机等待一段时间后重试。**帧间距机制：**802.11 允许 DCF 和 PCF 在一个单元内共存，这是通过帧间距机制实现的：**SIFS：**允许正处于会话中的节点优先发送，如收到 RTS 的节点发送一个 CTS，收到数据帧的节点允许发送一个 ACK 帧；**PIFS：**如果在 SIFS 后没有节点发送，在 PIFS 之后 PCF 模式的基站可以发送一个信标帧或一个轮询帧。**DIFS：**如果 PIFS 后没有基站发送，DIFS 之后任何节点可以竞争信道。**EIFS：**如果以上间隔都没有发送，EIFS 之后收到坏帧或未知帧的节点可以发送一个错误报告帧。**不使用信道预约机制的 CSMA/CA：**当节点有帧要发送时，侦听信道：1）若一开始就侦听到信道空闲，等待 DIFS 时间后发送帧 2）否则，选取一个随机回退值，在侦听到信道空闲时递减该值；在此过程中若侦听到信道忙，冻结计数值 3）当计数值减为 0 时，发送整个帧并等待确认 4）若收到确认帧，表明帧发送成功，若还有新的帧要发送，从第 2 步开始 CSMA/CA；若未收到确认，重新进入第 2 步中的回退阶段，并从一个更大的范围内选取随机回退值；如果有 k 个节点等待发送，它们随机选取的回退值确定了它们的发送顺序。**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；**原因：**冲突对无线网络损害很大，要尽可能避免。**802.11 帧格式：四个地址字段：**Address 1：帧的目的 MAC 地址；Address 2：帧的源 MAC 地址；Address 3：连接 AP 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：**无线终端 H1 向路由器 R1 发送帧，它的 AP 已知：H1 构造一个 address 1 = AP MAC，address 2 = H1 MAC，address 3 = R1 MAC，将该帧发给 AP；AP 将这个 802.11 帧转换为 802.3 帧（有线），后者的 dest addr = R1 MAC，source addr = H1 MAC。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是不可见的。**其他字段 duration of reserved transmission time (RTS/CTS)，frame seq # (for reliable ARQ)，frame type (RTS, CTS, ACK, data)。802.11 终端在子网内移动：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：**当 H1 检测到来自 AP1 的信号逐渐减弱时，开始扫描新的信标帧；当 H1 收到来自 AP2、信号更强的信标帧时，先解除与 AP1 的关联，然后关联到 AP2。发生切换时，交换机（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 先进功能：速率适应：**当主机移动或信噪比变化时，基站和主机动态改变传输速率（物理层调制技术）；**功率管理：节点设置功率管理比特，告知 AP 它将进入休眠状态：**节点进入休眠，并在下一个信标帧之前醒来；在节点休眠期间，AP 缓存发往该节点的帧；AP 在发送的信标帧中包含一个移动节点列表，这些节点有帧缓存在 AP 中；列表中的节点向 AP 请求帧，其余节点重新进入休眠。**终端在 IP 子网间移动：**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；即使对方获知了终端的新地址，应用必须重新建立连接，因为通信的端点（套接字）变了。**归属网络：**移动节点的永久"居所"(e.g., 128.119.40/24)。**永久地址：**移动节点在归属网络中的地址，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）。**归属代理：**当移动节点在外地时，为移动节点执行移动管理功能的实体。**外地网络：**移动节点当前所在的网络(e.g., 79.129.13/24)。**外地代理：**外地网络上为移动节点执行移动管理功能的实体。**转交地址：**移动节点在外地网络上的地址(e.g., 79,129.13.2)。**通信者：**希望与移动节点通信的节点。**移动节点注册：**移动节点进入外地网络后，向外地代理注册；移动节点通过外地代理向归属代理注册；**最终结果：**外地代理知道移动节点在本地网络上；归属代理知道移动节点的转交地址，记录到地址绑定表中。**间接选路到移动节点：**通信者在数据包中使用移动节点的永久地址；归属代理截获数据包，转发给外地代理；归属代理截获数据包，转发给外地代理；移动节点直接将响应发送给通信者。**间接选路：三角选路问题：**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（从而移动节点的位置对于通信者是透明的）；转交地址——归属代理用来向移动节点转发数据报；**三角选路：**通信者-归属网络-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：终端在外地网络间移动：**假设节点移动到另一个网络：向新的外地代理注册；新的外地代理向归属代理注册；归属代理更新移动节点的转交地址；归属代理使用新的转交地址向移动节点转发包；节点移动及变换外地网络等**对通信者都是透明的：**正在进行的通信可以保持！**直接选路到移动节点：**通信者向归属代理请求，并获知移动节点的转交地址（此步以后不必再做）；通信者将包发送给外地代理；外地代理将包转发给移动节点；移动节点直接向通信者发送。**直接选路: comments：**克服了三角选路的问题；对通信者不透明：通信者需要知道移动节点的转交地址，通信者（包括固定节点）需要增加对移动通信的支持。**Mobile IP（RFC 3344）：具有许多我们已经看到的特性，如：**归属代理，外地代理，永久地址，转交地址，移动节点注册；**标准化了三个部分：**代理发现，移动节点注册，数据报间接选路。**代理发现：**愿意充当归属代理或外地代理的路由器定期在网络上发送代理通告，宣布自己的存在及 IP 地址；愿意充当外地代理的路由器在代理通告中会提供一个或多个转交地址（通常使用自己的 IP 地址作为转交地址）；移动节点通过接收和分析代理通告，判断自己是否处于外地网络以及是否切换了网络；如果发现在外地网络上，移动节点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。**移动主机注册：**移动节点向外地代理发送一个注册请求，给出自己的永久地址、转交地址、归属代理地址以及认证信息等；外地代理记录相关信息，向归属代理转发注册请求；归属代理处理注册请求，若认证通过，将移动节点的永久地址及转交地址保存在绑定表中，发回一个注册响应；外地代理收到有效的注册响应后，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；当移动节点回到归属网络时，要向归属代理注销。**数据报间接选路：**数据包首先被归属代理得到；归属代理查找地址绑定表，获得移动节点当前的转交地址；归属代理将数据包发送到转交地址；外地代理将数据包转发给移动节点。**归属代理如何得到数据报？若通信者不在归属网络上：**数据包首先到达移动节点归属网络上的路由器；路由器查表得知可以直接交付，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；利用得到的 MAC 地址，将数据报封装到链路层帧中发送；**若通信者在归属网络上：**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。**数据报如何能被归属代理得到？**链路层帧的目的地址必须是归属代理的 MAC 地址；也就是说，移动节点的永久地址应当映射到归属代理的 MAC 地址。**ARP 代理：**归属代理为位于外地网络的移动主机发送 ARP 响应，用自己的 MAC 地址进行响应；也就是说，将移动主机的永久地址映射到归属代理的 MAC 地址。**免费 ARP：**当接收到移动主机的注册请求后，归属代理主动发送 ARP 请求，刷新其它节点的 ARP 缓存。**数据报如何到达转交地址？归属代理如何将数据报发送到转交地址？**归属代理收到的数据报，目的地址为移动节点的永久地址，而移动节点的转交地址位于外地网络。*如何将目的地址在归属网络的数据报送达外地网络？***方法：**修改目的地址=转交地址 （×）（转交地址是外地代理的 IP 地址，而数据报的最终目的地应是移动节点）；**使用隧道（√）（隧道技术的又一个应用例子）。归属代理通过隧道转发数据包：归属代理向外地代理发送的包：**Src IP = 归属代理 IP，Dst IP = 转交地址，里面封装着通信者发送的包（这个包的 dest 是永久地址）；**外地代理向移动节点发送的包：**通信者发送的原始包。**外地代理如何转发数据包到移动节点？**外地代理解封装收到的数据包，得到原始数据报；*外地代理如何获得移动节点的 MAC 地址？在移动节点注册阶段，外地代理获知了移动节点的永久地址和 MAC 地址，记录在其转发表中；外地代理根据目的 IP 地址查找转发表，得到移动节点的 MAC 地址*；外地代理利用移动节点的 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=移动节点永久地址，DestIP=通信者 IP 地址，SrcMAC=移动节点 MAC，DestMAC=外地代理 MAC；外地代理按照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 是外地代理的地址。**无线和移动对上层协议的影响：无线链路带来的问题：**误码率、丢包率、延迟增大。**节点移动带来的问题：**丢包、延迟增大。**逻辑上，没什么影响：**为上层协议提供的仍然是尽力而为的服务，因此 TCP 和 UDP 也可以运行在无线网络上。**性能上，有很大影响：**丢包率高，传输延迟增大；TCP 将丢包（长延迟也当作丢包）解释为拥塞，不必要地减小拥塞窗口，导致应用吞吐率很低；无线链路、有线/无线混合链路上的 TCP 拥塞控制是一个研究问题。

计算机网络中的安全：

什么是网络安全：网络安全的通用定义：网络安全是指网络系统的硬件、软件及其系统中的数据受到保护，不受偶然的或者恶意的原因而遭到破坏、更改、泄露，系统连续可靠地运行，网络服务不中断。**安全通信**具有下列所需要的特性：• **机密性：**报文内容的机密性，通信活动的机密性；• **报文完整性：**报文来自真实的源，且传输过程中未被修改。• **端点鉴别：**发送者和接收者能够证实对方的身份；• **运行安全性：**网络不受攻击，网络服务可用。**安全攻击的类型：被动攻击：**试图从系统中获取信息，但不对系统产生影响；**两种类型：偷听：**监听并记录网络中传输的内容；**流量分析：**从通信频度、报文长度等流量模式推断通信的性质。**主动攻击：**试图改变系统资源或影响系统的操作；**四种类型：伪装：**一个实体假冒另一个实体；**重放：**从网络中被动地获取一个数据单元，经过一段时间后重新发送到网络中；**报文修改：**改变报文的部分内容、推迟发送报文或改变报文的发送顺序；**拒绝服务：**阻止通信设施的正常使用或管理。**常见的安全机制：加密：**使用数学算法对数据进行变换，使其不易理解；**鉴别：**通过报文交换确信一个实体的身份，以防假冒；**数据完整性：**用于保护数据单元或数据单元流的完整性，以防报文修改；**数字签名：**附加在一个数据单元后面的数据，用来证明数据单元的起源及完整性，以防伪造及抵赖；**流量填充：**在数据流间隙中插入比特，以挫败流量分析的企图；**访问控制：**通过授权机制限制用户对资源的访问，防止越权。**密码学的原则：明文（plaintext）：**欲加密的原始数据；**密文（ciphertext）：**明文经加密算法作用后的输出；**密钥（key）：**加密和解密时需要使用的参数；**密码分析（cryptanalysis）：**破译密文；**密码学（cryptology）：**设计密码和破译密码的技术统称为密码学。**按照加密密钥与解密密钥是否相同，加密算法分为：对称**加密算法：加密密钥与解密密钥相同；**非对称**加密算法：加密密钥与解密密钥不同。**按照明文被处理的方式，加密算法分为：块密码（分组密码）：**每次处理一个明文块，生成一个密文块；**流密码：**处理连续输入的明文流，并生成连续输出的密文流。**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；例子：凯撒密码，单字母表替换，多字母表替换；密钥：字母映射表。**换位密码：**保留明文字母不变，但改变字母的位置；例子：列换位密码。**密码的安全性：传统**加密方法的安全性建立在算法保密的基础上，现**代**加密方法也使用替换和换位两种基本手段，但现代密码学的基本原则是：加密与解密的算法是公开的，只有密钥是需要隐藏的。一个加密算法被称为是**计算安全**的，如果由该算法产生的**密文满足以下两个条件之一：**破译密文的代价超过信息本身的价值；破译密文所需的时间超过信息的有效生命期。现代密码学中，密码的安全性是通过算法的复杂性和密钥的长度来保证的。**针对加密系统的密码分析攻击：惟密文攻击：**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的情况）；**已知明文攻击：**密码分析者除了有截获的密文外，还有一些已知的"明文-密文对"来帮助破译密码，以得出密钥；**选择明文攻击：**密码分析者可以任意选择一定数量的明文，用被攻击的加密算法加密，得到相应的密文，以利于将来更有效地破解由同样加密算法及相关密钥加密的信息。**一个安全的加密系统必须能抵御选择明文攻击。**全表快密码难以实现。**DES（Data Encryption Standard）：**块加密算法，每次以 64 比特的明文块作为输入，输出 64 比特的密文块；DES 是基于迭代的算法，每一轮迭代执行相同的替换和换位操作，但使用不同的密钥；DES 使用一个 56 比特的主密钥，每一轮迭代使用的子密钥（48 比特）由主密钥产生；一种对称加密算法，加密和解密使用相同的函数，两者的不同只是子密钥的次序刚好相反；缺点：密钥长度不够长，迭代次数不够多。**DES 的计算过程：**首先进行一次初始换位（permute），然后进行 16 轮相同的迭代，每轮迭代使用一个不同的子密钥（由主密钥生成），最后再进行一次换位（与初始换位相反）。**3DES** 使用两个密钥进行**三轮 DES 计算：**第一轮令 DES 设备工作于加密模式，使用密钥 K1 对明文进行变换；第二轮令 DES 设备工作于解码模式，使用密钥 K2 对第一轮的输出进行变换；第三轮令 DES 设备工作于加密模式，用密钥 K1 对第二轮的输出进行变换，输出密文。**有关 3DES 的三个问题：**为什么使用两个密钥而不是三个密钥？112 比特的密钥已经足够长。为什么不使用两重 DES（EE 模式）而是三重 DES？考虑采用 EE 模式的两重 DES，且攻击者已经拥有了一个匹配的明文--密文对（P1, C1），即有 C1 = EK2 (EK1 (P1))，令 X = EK1(P1) = DK2(C1)，攻击者分别计算 EK1(P1)和 DK2(C1)，并寻找使它们相等的 K1 和 K2，则穷尽整个密钥空间只需 256 的攻击量而不是 2112。（中途攻击）为什么是 EDE 而不是 EEE？为了与单次 DES 兼容。3DES 用户解密单次 DES 用户加密的数据，只需令 K1 = K2 就行了。**AES（Advanced Encryption Standard）：**每次处理 128 比特明文块，输出 128 比特密文块；密钥长度可以是 128、192 或 256 比特；如果使用强力方法破解，假设破解 DES 需要 1 秒，则破解 AES（128 比特密钥）需要 149 万亿年！**密码块链接（Cipher Block Chaining）：**若每个明文块被独立加密，相同的明文块生成相同的密文块，容易被重放攻击利用。**密码块链接（CBC）：**发送方生成一个随机的初始向量 c(0)，用明文发送给接收者；每一个明文块加密前，先与前一个密文块进行异或，然后再加密：第一个明文块与 c(0)异或，相同的明文块几乎不可能得到相同的密文块，并且不因引入随机性而产生额外的大量带宽（只额外传送 C0），而后果是需要在协议中提供机制分发初始向量（IV）C0。**公开密钥加密：对称加密算法：**要求发送者和接收者使用同一个密钥；存在密钥传递问题：发送方选择了一个密钥后，如何将密钥安全地传递给接收方？**非对称加密算法：**发送者和接收者不共享密钥，发送者使用加密密钥，接收者使用解密密钥，不存在密钥传递问题：加密密钥是公开的，解密密钥是私有的。**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；当 Alice 希望向 Bob 发送一个加密信息时：Alice 从公开的文件中查到 Bob 的加密密钥，用 Bob 的加密密钥加密信息，发送给 Bob；Bob 用自己的解密密钥解密信息。**公开密钥和私有密钥：**公开密钥：加密密钥，由发送者使用；私有密钥：解密密钥，由接收者使用。**要求：**KB-(KB+(m)) = m，给定公钥 KB+，不可能计算出私钥 KB-。**公开密钥算法应满足的条件：**从计算上说，生成一对加密密钥和解密密钥是容易的，已知加密密钥，从明文计算出密文是容易的，已知解密密钥，从密文计算出明文是容易的，从加密密钥推出解密密钥是不可能的，从加密密钥和密文计算出原始明文是不可能的。**公开密钥算法两个问题：**入侵者知道该公钥和加密算法，可以据此发起选择明文攻击；加密密钥是公开的，任何人都可能向 Bob 发送一个已加密的报文，在单一共享密钥情况下，发送方知道共享秘密密钥的事实就已经向接收方隐含地证实了发送方的身份，然而在公钥体制中，这点就行不通了，因为任何一个人都可向 Bob 发送使用 Bob 的公开可用密钥加密的报文，这就需要用数字签名把发送方和报文绑定起来。**RSA 算法：生成密钥：**选择两个大素数 p 和 q （典型值为大于 10^100）；计算 n = p*q 和 z = (p-1)*(q-1)；选择一个与 z 互质的数，令其为 d；找到一个 e 使满足 e*d = 1 (mod z)；公开密钥为(e, n)，私有密钥为 (d, n)。**RSA 算法：加密和解密：加密方法：**将明文看成是一个比特串，将其划分成一个个数据块 M，且有 0≤M < n；对每个数据块 M，计算 C = M^e (mod n)，C 即为 M 的密文；**解密方法：**对每个密文块 C，计算 M = C^d (mod n)，M 即为要求的明文。**RSA：另一个重要的特性：**先用公钥再用私钥和先用私钥再用公钥结果相同。**RSA 的特点：**优点：安全性好：RSA 的安全性建立在难以对大数据取因子的基础上，这是目前数学家尚未解决的难题；使用方便：免除了传递密钥的麻烦。缺点：计算开销大，速度慢。**RSA 的应用：**RSA 一般用来加密少量数据，如用于鉴别、数字签名或发送一次性会话密钥等。**报文完整性和数字签名：报文完整性（又称报文鉴别）：**用于验证一个报文是否可信的技术。**一个报文是可信的，如果**它来自声称的源并且没有被修改。**报文鉴别涉及两个方面：**起源鉴别：报文是否来自声称的源；完整性检查：报文是否被修改过。**方法一：对整个报文加密：**如果发送方和接收方有一个共享的密钥，可以通过加密报文来提供报文鉴别：发送方用共享的密钥加密整个报文，发送给接收方；如果接收方能够正确解密收到的报文，则报文必是可信的。缺点：混淆了机密性和报文鉴别两个概念，有时我们只想知道报文是否可信，而报文本身并不需要保密；加密整个报文会带来不必要的计算开销。**将报文鉴别与数据机密性分开：设想：**发送者用明文发送报文，并在报文后附上一个标签，允许接收者利用这个标签来鉴别报文的真伪；**用于鉴别报文的标签必须满足两个条件：**能够验证报文的完整性（是否被修改），不能被伪造。**数据完整性和报文摘要：报文摘要（数字指纹）：**将一个散列函数作用到一个任意长的报文 m 上，生成一个固定长度的散列值 H(m)，这个散列值称为该报文的报文摘要（message digest），也称数字指纹。**使用报文摘要验证报文的完整性：**发送者对发送的报文计算一个报文摘要，作为标签和报文一起发给接收者；接收者对收到的报文也计算一个报文摘要，和收到的标签（发送方计算的报文摘要）进行比较。**如何保证报文摘要不被修改？报文鉴别之方法二：**发送方计算报文摘要，然后用与接收方共享的密钥加密报文摘要，形成报文鉴别标签（也称**报文鉴别码**）；接收方用共享的密钥解密报文鉴别码，得到发送方计算的报文摘要，与自己计算的报文摘要进行比较；缺点：需要使用加密算法。**密码散列函数：为什么要开发一个不需要加密算法的报文鉴别技术？**加密软件通常运行得很慢，即使只加密少量的数据；加密硬件的代价是不能忽略的；加密算法可能受专利保护（如 RSA），因而使用代价很高；加密算法可能受到出口控制（如 DES），因此有些组织可能无法得到加密算法。**使用密码散列函数（cryptographic hash function）生成报文鉴别码：**使用密码散列函数计算报文摘要时需要包含一个密钥，但它并不用来做加密运算；发送方用双方共享的一个秘密密钥 KS 添加到报文 m 之前，然后计算报文摘要 H (KS || m)形成报文鉴别码。**散列函数 H 应满足的特性：**H 能够作用于任意长度的数据块，并生成固定长度的输出；对于任意给定的数据块 x，H(x)很容易计算；对于任意给定的值 h，要找到一个 x 满足 H(x)=h，在计算上是不可能的（单向性）：该特性对于使用密码散列函数的报文鉴别很重要，如果根据 H(KS||m)=h 可以找到一个 x，使得 H(x)=h，那么根据 x 和 m 可以推出 KS；对于任意给定的数据块 x，要找到一个 y≠x 并满足 H(y)=H(x)，在计算上是不可能的：该特性对于使用加密算法的报文鉴别很重要，如果能找到一个不同于 x 的数据块 y，使得 H(y)=H(x)，那么就可以用 y 替换 x 而不被接收方察觉；要找到一对（x, y）满足 H(y) = H(x)，在计算上是不可能的。（抵抗生日攻击）满足前四个特性的散列函数称为**弱散列**函数，满足所有五个特性的散列函数称为**强散列**函数。**散列函数标准：**目前使用最多的两种散列函数：MD5：散列码长度为 128 比特；SHA-1：美国联邦政府的标准，散列码长度为 160 比特；目前获得最多支持的密码散列函数方案为 HMAC，已应用到许多安全协议中。**数字签名：一个可以替代手写签名的数字签名必须满足以下三个条件：**接收方通过文档中的数字签名能够鉴别发送方的身份（起源鉴别），发送方过后不能否认发送过签名的文档（防抵赖），接收方不可能伪造被签名文档的内容。MAC 无法胜任这项工作，因为有两个人拥有它。**用私钥加密报文摘要：**发送方先计算报文摘要，然后用自己的私钥加密报文摘要形成数字签名，数字签名附加在报文后面一起发送；接收方拷贝一份数字签名，妥善保存，以备将来需要时使用；接收方用发送方的公钥得到原始的报文摘要，对收到的报文计算摘要，如果两者相符，表明报文是真实的。**数字签名与 MAC 进行比较：**数字签名和 MAC 都以一个报文（或一个文档）开始。为了从该报文中生成一个 MAC，我们为该报文附加一个鉴别密钥，然后取得该结果的散列；注意到在生成 MAC 过程中既不涉及公开密钥加密，也不涉及对称密钥加密；为了生成一个数字签名，我们首先取得该报文的散列，然后用我们的私钥加密该报文（使用公钥密码）；因此，数字签名是一种"技术含量更高的"技术，因为它需要一个如后面描述的、具有认证中心支撑的公钥基础设施(PKI)。**如何可靠地获取公钥？**考虑下面的例子：Bob 将公钥 EB 发布在自己的主页上；Alice 获取 Bob 主页的请求被 Trudy 截获，Trudy 将假冒的 Bob 主页发送给 Alice，主页中的公钥是 Trudy 的公钥 ET；Alice 使用 Trudy 的公钥加密会话密钥，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。**问题：**当 Alice 从公开的途径得到 Bob 的公钥后，Alice 如何确认她得到的就是 Bob 的公钥，而不是其他人的公钥？**公钥证书：**为使公钥密码体系有实际应用，每个实体必须能够确认它得到的公钥确实来自声称的实体；解决方案是引入证书机制：使用证书（certificate）来证明某个主体（principal）拥有某个公钥，证书由一个可信任的第三方机构颁发，该机构称为认证权威 CA（certification authority）；证书包含主体的公钥和 CA 的签名，任何人无法伪造或篡改证书的内容；当一个主体获得其公钥证书后，可将证书放在任何一个可公开访问的地方。**证书的获取：**Bob 向 CA 注册其公钥：Bob 向 CA 提供身份证明；CA 验证了 Bob 的身份后创建证书，绑定 Bob 及其公钥；证书包含 Bob 的公钥，并有 CA 的签名。**证书的验证：**当 Alice 需要 Bob 的公钥时：获取 Bob 的证书；使用 CA 的公钥验证 Bob 的证书，得到 Bob 的公钥。**X.509 证书：**目前最常用的证书标准；建立在公钥算法和数字签名的基础上：CA 对证书内容先进行 SHA-1 散列，然后用 CA 的私钥对报文摘要加密，形成数字签名；为验证公钥证书的真实性：验证方用 CA 的公钥解开证书的签名，得到证书内容的报文摘要；对收到的证书内容计算报文摘要，并与解密得到的报文摘要进行比较，两者相同表明这是合法的公钥证书。**谁可以运行 CA？世界上有几个 CA？**使用一个 CA 签发全世界所有的证书?流量压力，单点失效。由一个组织运行多个 CA?密钥泄露，信任问题。**分布式公钥基础设施（Public Key Infrastructure，PKI）：**提供公钥加密和数字签名服务的系统或平台；包含不同组织运行的 CA，每个 CA 拥有自己的私钥，负责为一部分用户签发证书；用户自己决定使用哪一个 CA。**信任锚与信任链：信任锚（trust anchor）：**信任的起点，系统中的所有实体都以根 CA 的公钥作为它们的信任锚，信任锚必须通过安全的物理途径获取；**信任链（chain of trust）：**从叶结点到根 CA 的证书序列；**根 CA 的选择：**有许多根 CA，每个根 CA 都有自己的一个分级结构，所有根 CA 间可以进行交叉认证；用户自行决定信任哪个根 CA；实际上，许多根 CA 的公钥被预装在浏览器上，这些根 CA 由浏览器厂商认证并嵌入到软件中，随软件一起发布。**证书的撤销：**每个证书都有有效期，过期后证书自动失效；CA 也可以显式地撤销证书，这要求 CA 定期地发布证书撤销列表（Certificate Revocation List，CRL），表中给出已经撤销的证书序列号；每个用户在使用一个证书前都要去获取 CRL，检查该证书是否在 CRL 中。**证书存放在哪里？**使用 DNS 作为**证书目录，**该方案的标准为 DNSSEC；使用专门的目录服务器存放证书，该方案的标准为 LDAP；证书撤销列表通常与证书存放在一起，CA 定期地将 CRL 推进目录服务器，由目录服务器负责将 CRL 中列出的证书清除掉。**端点鉴别：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。鉴别协议通常在两个通信实体运行其他协议（例如，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，各方才继续下面的工作。**鉴别协议 ap1.0：**直接发送一个报文。**鉴别协议 ap2.0：**有一个总是用于通信的周知网络地址（IP）。Trudy 用 Alice 的 IP 地址创建一个数据包（IP 地址欺骗）。**鉴别协议 ap3.0：**Alice 向 Bob 发送口令证明自己，口令是鉴别者和被鉴别者之间共享的秘密。Trudy 监听到 Alice 发送的明文口令，过后发送给 Bob。**鉴别协议 ap3.1：**Alice 将口令加密，发送给 Bob。Trudy 截获数据包，过后发送给 Bob（重放攻击）**鉴别协议 ap4.0：目标：**避免重放攻击。**失败的情况是因为** Bob 不能区分 Alice 的初始鉴别报文和后来入侵者回放的 Alice 的初始鉴别报文所致；也就是说，Bob 无法判断 Alice 是否还活跃（即当前是否还在连接的另一端），或他接收到的报文是否就是前面鉴别 Alice 时录制的回放。**不重数(nonce)：**在一个协议的生存期中只使用一次的数，也就是说，一旦某协议使用了一个不重数，就永远不会再使用那个数字了。**协议 ap4.0 以如下方式使用一个不重数：1)** Alice 向 Bob 发送报文"我是 Alice"；**2)** Bob 选择一个不重数 R，然后把这个值发送给 Alice；**3)** Alice 使用她与 Bob 共享的对称秘密密钥 KA-B 来加密这个不重数，然后把加密的不重数 KA-B(R)发回给 Bob，与在协议 ap3.1 中一样，由于 Alice 知道 KA-B 并用它加密一个值，就使得 Bob 知道收到的报文是由 Alice 产生的。这个不重数用于确定 Alice 是活跃的；**4)** Bob 解密接收到的报文，如果解密得到的不重数等于他发送给 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：**"I am Alice"→R→KA-(R)→"send me your public key"→KA+。**Bob 计算：KA+(KA-(R)) = R**，只有 Alice 拥有这个私钥，因而一定是 Alice！**X.509 定义了三种鉴别程序，供不同的应用选择：**单向鉴别：涉及一个用户到另一个用户的一次报文传输（接收方鉴别发送方）；双向鉴别：通信双方相互鉴别；三向鉴别：通信双方相互鉴别，并提供报文同步机制。**单向鉴别（one-way authentication）：**A→B：tA ||rA|| IDB || Data || Kb+(Ka-b) || signatureA；**说明：**tA：时间戳，由报文的产生时间和到期时间组成；rA：A 随机选择的一个不重数，供接收者检测重放攻击；IDB：B 的标识，指示报文的接收者；Data：报文中包含的数据信息；Ka-b ：若 Data 需要保密，则 Ka-b 为 A 加密 Data 使用的对称密钥；Kb+ ：B 的公开密钥，用于加密对称密钥 Ka-b；signatureA ：A 的数字签名，对 tA、rA、IDB 和 Data 的明文生成；**接收方鉴别：**B 用自己的私钥解出 Ka-b，用 Ka-b 解密 Data，计算前面 4 个部分的报文摘要；B 用 A 的公钥从签名中得到原始的报文摘要，进行比较。**双向鉴别（two-way authentication）：**A→B：tA || rA || IDB || Data || Kb+(Ka-b) || signatureA；B→A：tB || rB || IDA || rA|| Data || Ka+(Kb-a) || signatureB。**三向鉴别（three-way authentication）：**A→B：tA || rA || IDB || Data || Kb+ (Ka-b) || signatureA；B→A：tB || rB || IDA || rA|| Data || Ka+(Kb-a) || signatureB；A→B：rB || signatureA。**安全电子邮件：为什么要在因特网的多个层次上提供安全性功能呢？仅在网络层提供安全性功能并加以实施还不足够吗？**首先，尽管可以通过加密数据报中的所有数据（即所有的运输层报文段），以及通过鉴别所有数据报的源 IP 地址，在网络层能够提供"地毯式覆盖"安全性，但是却并不能提供用户级的安全性。例如，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。第二，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）通常较为容易，而等待在网络层上广泛地部署安全性，可能还需要未来若干年才能解决。**安全电子邮件：电子邮件安全**包括**真实性**和**机密性**两个方面，同时最为期望的安全特性还有**"发送方鉴别""报文完整性"和"接收方鉴别"**。**提供机密性的方式：对称密钥算法？**仅有 Alice 和 Bob 具有该密钥的副本，这使得分发对称密钥非常困难。**公开密钥密码？**效率相对低下，尤其对于长报文更是如此。**为了**克服效率问题，我们利用了**会话密钥**，**具体来说：1).** Alice 选择一个随机对称会话密钥 Ks；**2).** 用这个对称密钥加密她的报文 m；**3).** 用 Bob 的公钥 KB+加密这个对称密钥；**4).** 级联该加密的报文和加密的对称密钥以形成一个"包"；**5).** 向 Bob 的电子邮件地址发送这个包。**当 Bob 接收到这个包时：**他使用其私钥 KB-得到对称密钥 Ks；使用这个对称密钥 Ks 解密报文 m。**只关心发送方鉴别和报文完整性：**使用数字签名和报文摘要。**具体说来：**Alice 对她要发送的报文 m 应用一个散列函数 H(例如 MD5),从而得到一个报文摘要；用她的私钥 KA-对散列函数的结果进行签名，从而得到一个数字签名；把初始报文（未加密）和该数字签名级联起来生成一个包；向 Bob 的电子邮件地址发送这个包。**当 Bob 接收到这个包时：**他将 Alice 的公钥 KB+应用到被签名的报文摘要上；将该操作的结果与他自己对该报的散列 H 进行比较。**设计一个提供机密性、发送方鉴别和报文完整性的电子邮件系统：**过程结合起来而实现：Alice 首先生成一个预备包，它与图 8-20 中的包完全相同，其中包含她的初始报文和该报文数字签名过的散列；然后 Alice 把这个预备包看作一个报文，再用图 8-19 中的发送方的步骤发送这个新报文，即生成一个新包发给 Bob。**当 Bob 接收到这个包后，**他首先应用图 8-19 中他这一侧的步骤，然后再应用图 8-20 中他这一侧的步骤。注意到在这一方案中，Alice 两次使用了公开密钥密码：一次用了她的私钥，另一次用了 Bob 的公钥。同样，Bob 也两次使用了公开密钥密码：一次用了他的私钥，一次用了 Alice 的公钥。**PGP：目前最流行的两个安全电子邮件协议：PGP：**一个开放源码的安全电子邮件软件包，提供对邮件的保密、鉴别、数字签名和压缩服务，PGP 较多地用于个人电子邮件安全。（因特网安全电子邮件的事实标准）；**S/MIME：**基于公钥加密技术对 MIME 所做的安全扩展。S/MIME 较可能作为一种工业标准被商业组织或一些机构使用。**PGP 提供五种服务：**鉴别，机密性，压缩，兼容电子邮件，分段；**鉴别：**PGP 使用基于公开密钥算法的数字签名提供鉴别服务，生成可供鉴别的电子邮件：发送方创建电子邮件（报文），用 SHA-1 计算邮件的报文摘要，然后用发送者的私钥加密报文摘要，形成数字签名，将数字签名附在报文的前面，与报文一起发送：Sgn || Data。**机密性：**使用对称密钥算法保护邮件的机密性：发送方将选择的一次性会话密钥用接收方的公钥加密，与报文一起发送给接收方；仅使用机密性服务的过程：发送方（A）生成一个报文和一个随机的 128 比特数（一次性会话密钥），先用会话密钥加密报文，再用接收方（B）的公钥加密会话密钥，将加密后的会话密钥放在报文前面，与报文一起发送：KB+(KA-B) || KA-B(Data)。**压缩：**缺省地，PGP 在完成签名之后、在加密报文之前对报文进行压缩，压缩算法采用 ZIP：KB+(KA-B) || KA-B (Zip (Sgn || Data))；在压缩报文之前计算数字签名，是为了方便日后对报文的验证；在加密报文之前进行压缩，一方面可以减少要加密的数据量，另一方面压缩后的消息冗余很少，增加密码分析的困难。**兼容电子邮件：**PGP 使用 Base64 编码将二进制数据流转换成可打印 ASCII 文本，以解决邮件的传输问题；PGP 可被配置为仅对报文中的某些部分（如签名部分）进行 Base64 编码转换。**同时使用以上四种服务的过程：**发送方先对明文报文计算签名，将签名放在报文前面；签名与明文一起被压缩；用会话密钥对压缩后的数据块进行加密；用接收方的公钥加密会话密钥，放在报文的前面；将整个数据块转换成 Base64 编码格式：EncodeBase64(KB+(KA-B) || KA-B (Zip (Sgn || Data)))。**邮件分段：**许多电子邮件系统能够接收的最大报文长度不超过 50,000 字节；PGP 在完成对报文的全部处理后，自动将超过长度的报文分成小块传输，会话密钥和签名只在第一个片段中出现；接收端去掉每个片段的头部，然后将所有的片段重组装成一个数据块。**使 TCP 连接安全：SSL：**向基于 **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）。**安全服务：**服务器鉴别，数据加密，客户鉴别（可选）。**SSL 建立在 TCP 之上，依靠 TCP 提供可靠的端到端连接。**SSL 是**涉及到两个层次的一组协议：**SSL 记录协议：为各种高层协议（如 HTTP）提供基本的安全服务；其它三个高层协议：用于 SSL 交换管理。**宏观描述：SSL 握手协议：**允许服务器和客户之间相互鉴别，并协商加密算法、MAC 算法及密钥等。**握手协议由客户和服务器之间的一系列报文交换组成：**浏览器向服务器发送建立 SSL 会话的请求报文，说明可支持的 SSL 协议最高版本、支持的加密算法（按优先级从高到低排列）和压缩方法等，以及浏览器选择的一个随机数 Rc；服务器从浏览器给出的选择中确定合适的 SSL 版本号、加密算法和压缩方法，与服务器选择的一个随机数 Rs 一起发送给浏览器；服务器向浏览器发送它的公钥证书（和必要的证书链）以及其它信息；浏览器检查签发证书的 CA 是否在浏览器的可信 CA 列表中，若不在向用户警告该问题；如果在则使用该 CA 的公钥验证证书，得到服务器的公钥；如果客户也需要被鉴别（收到服务器的证书请求），则浏览器向服务器发送它的公钥证书；浏览器生成一个 48 字节的随机数，称预密钥，用服务器的公钥加密后发送给服务器；客户和服务器各自从预密钥、Rc 和 Rs 中计算加密数据需要的会话密钥，以及计算 MAC 需要的密钥；浏览器向服务器发送一个报文，通知它后面的报文都用这个会话密钥加密，然后发送一个用协商的算法及密钥加密的报文，指示握手协议的浏览器部分完成；服务器向浏览器发送一个报文，通知它后面的报文都用这个会话密钥加密，然后发送一个用协商的算法及密钥加密的报文，指示握手协议的服务器部分完成。**SSL 记录协议：**SSL 记录协议为 SSL 连接提供**两种服务：机密性：**通过加密 SSL 载荷实现；**完整性：**通过报文鉴别码保护。**SSL 记录协议的操作过程：**从上层接收一个要传输的应用报文，将报文划分成长度不超过 214 字节的数据块；（可选）对数据块进行压缩；使用密码散列函数对数据块计算报文鉴别码；使用对称密钥算法对（压缩的）数据块及报文鉴别码进行加密，加密算法可以是 DES、3DES、IDEA、RC 等；在处理完的数据块前加上 SSL 头，包括内容类型、SSL 版本号、压缩数据块的长度等。**阻止插入、删除和代替报文段：**使用序号；Bob 维护一个序号计数器，计数器开始为 0，Bob 每发送的一个 SSL 记录它都增加 1；Bob 并不实际在记录中包括一个序号，但当他计算 MAC 时，他把该序号包括在 MAC 的计算中；所以，该 MAC 现在是数据加 MAC 密钥 MB 加当前序号的散列；Alice 跟踪 Bob 的序号，通过在 MAC 的计算中包括适当的序号，使她验证一条记录的数据完整性。**网络层安全性：IPsec 和虚拟专用网：IPv4 在设计时没有考虑安全性：**缺少对通信双方身份的鉴别，容易遭受地址欺骗攻击；缺少对网络中数据的完整性和机密性的保护，数据很容易被窃听、修改甚至劫持。**IP 安全协议（IPSec）：**IETF 以 RFC 形式公布的一组安全协议集；**目标是**把安全特征集成到 IP 层，以便对因特网中的安全业务提供低层的支持。**IPsec 和虚拟专用网：专用网：**通过电信专线将分散在各地的计算机（网络）连接而成的网络；安全性好，但代价高。**虚拟专用网（Virtual Private Network）：**建立在公用网上的一个覆盖网络（overlay），在逻辑上与其它流量隔离；数据在发送到公用网之前进行加密。**VPN 的实现：VPN 的典型结构：**在每个局域网上设置一个安全网关，在每一对安全网关间创建一条穿过因特网的隧道，在隧道中使用 IPSec；**VPN 的优点：**可以在一对局域网间提供完整性控制及机密性服务，甚至对流量分析也有相当的抵御能力；对因特网中的路由器及用户软件是透明的，只要系统管理员设置好安全网关就可以了。传进公用网的 IPsec 被 IPv4 封装，同时拥有 IPv4 和 IPsec 首部。**安全关联：IPSec 提供了一个用于集成多种安全服务、加密算法及安全控制粒度的安全体系框架：IPSec 提供的安全服务包括：**访问控制、无连接完整性、数据起源认证、抗重放攻击、机密性等；IPSec 的安全机制独立于算法，因此在选择和改变算法时不会影响其它部分的实现；IPSec 提供多种安全控制粒度，包括：一条 TCP 连接上的通信，一对主机间的通信，一对安全网关之间的所有通信；用户可以为数据通信选择合适的安全服务、算法、协议和控制粒度。从技术上说，**IPSec 主要包括两个部分：IPSec 安全协议：**包括 **AH（不提供机密性）和 ESP** 两个安全协议，定义了用于安全通信的 IP 扩展头和字段，以提供机密性、完整性和源鉴别服务；**密钥管理协议：**定义了通信实体间进行身份鉴别、协商加密算法以及生成共享会话密钥的方法。*将以上两部分绑定在一起的是称为***安全关联（SA）***的抽象。***SA 是：**通信对等实体之间对某些要素的协定，如使用的安全协议、协议的操作模式、使用的密码算法、密钥及密钥的生存期等。**SA 是两个通信端点间的一个单工连接**，由一个安全参数索引（SPI）**唯一标识**，如果在两个方向上都需要安全通信，则需要建立两个 SA。SPI 携带在数据包中，由数据包的处理进程用来查找密钥及相关信息。**SA 可以建立在**一对主机之间、一台主机与一个安全网关之间、或一对安全网关之间。**IPsec 数据报：IPSec 的使用模式：传输模式：**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：**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 头之间，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：**传输模式比隧道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进行流量分析。**鉴别头部（Authentication Header）协议：SPI：**32 比特的数，和目的 IP 地址、安全协议结合起来，唯一标识数据报的 SA；**SeqNum：**对 SA 上发送的数据包进行编号，供接收端检测重放攻击。一个 SA 上的序号不能重用，因此在传输的数据包数量达到 232 之前，必须协商一个新的 SA 和新的密钥；**Authentication Data：**包含报文鉴别码的可变长度域，所有 AH 实现必须支持 HMAC-MD5-96 和 HMAC-SHA-1-96。**AH 协议提供**无连接完整性、数据起源认证和抗重放攻击，但不提供机密性服务：HMAC 覆盖数据包的载荷部分，因而可提供无连接完整性服务；HMAC 覆盖原始 IP 头中的不变域（传输模式）或整个原始 IP 头（隧道模式），因而可提供数据起源认证；AH 头中有序号，且被 HMAC 覆盖，因而可抵抗重放攻击。**封装安全载荷（Encapsulating Security Payload）：ESP 数据包大致分为以下几个部分：**ESP 头：包含 SPI 和 SeqNum；载荷：原始数据包中被加密部分的密文；ESP 尾：包括填充（需要的话）、填充长度和下一个头，ESP 尾也要被加密；鉴别数据：覆盖 ESP 头、载荷和 ESP 尾的报文鉴别码。**隧道模式路由 R1 使用下列方法将这个"普通 1Pv4 数据报"转换成一个 IPsec 数据报：**• 在初始 IPv4 数据报（它包括初始首部字段）后面附上一个"ESP 尾部"字段；• 使用算法和巾 SA 规定的密钥加密该结果；• 在这个加密量的前面附加上一个称为"ESP 首部"的字段；得到的包称为"enchilada"；• 使用算法和由 SA 规定的密钥生成一个覆盖整个 enchilada 的鉴别 MAC；• 该 MAC 附加到 enchilada 的后面形成载荷；• 最后，生成一个具有所有经典 1Pv4 首部字段（通常共 20 字节长）的全新 IP 首部，该新首部附加到载荷之前。（运输模式在第一部分缺少了初始 IP 首部）。**ESP 协议提供的安全服务：ESP 协议提供**数据机密性、无连接完整性、抗重放攻击、数据起源鉴别和有限的数据流机密性服务：原始数据包的载荷部分被加密，因而可提供数据机密性；HMAC 覆盖数据包载荷部分，可提供无连接完整性服务；ESP 头中有序号，且被 HMAC 覆盖，可抵抗重放攻击；ESP 隧道模式中，原始 IP 头也被 HMAC 覆盖，因而 ESP 隧道模式可提供数据起源鉴别；ESP 隧道模式中，原始 IP 头也被加密，路由器只能看到外层 IP 头，因而 ESP 隧道模式可提供数据流机密性服务。**AH 协议和 ESP 协议的安全性比较：**ESP 隧道模式的安全性强于 ESP 传输模式；数据机密性服务：只有 ESP 提供，AH 不提供；鉴别服务：ESP 隧道模式的鉴别服务，安全性强于 AH，ESP 传输模式的鉴别服务，安全性不如 AH。**使无线 LAN 安全：802.11 WEP（Wired Equivalent Privacy 有线等效保密）：**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。**802.11i：**具有更强安全机制的 802.11 版本，提供较强的加密机制及鉴别机制，提供密钥分发机制。**有限等效保密：主机鉴别过程：**无线主机向 AP 请求鉴别；AP 向主机发送一个 128 比特的不重数；主机使用与 AP 共享的对称密钥加密不重数，发送给 AP；AP 解密不重数，若与 AP 发送给主机的不重数相同，完成主机鉴别。利用主机与基站共享密钥这个事实鉴别主机。**WEP 数据加密：**主机与 AP 共享一个 40 比特的对称密钥 KS（半永久）；对于每个帧，发送方生成一个 24 比特的初始向量 IV，添加到 KS 后面，形成一个 64 比特的密钥 (KS, IV)；(KS, IV) 用于生成一个密钥流 {kiIV | i=1,2,···}；第 i 个密钥 kiIV 用来加密帧中的第 i 个字节 di：ci = di XOR kiIV；IV 和加密后的字节 ci 放在帧中传输；接收方使用相同的 (KS, IV) 生成相同的密钥流，执行解密运算：di = ci XOR kiIV。**WEP 加密的安全漏洞：**每个 KS 只有 224 个 (KS, IV) 可用：IV 会被重复使用；IV 用明文传输：攻击者可以观察到 IV 的重用。**攻击：**Trudy（可能通过欺骗方式）让 Alice 加密他选择的明文(d1 d2 d3 d4 ···)；Trudy 能够获得 Alice 加密的密文: ci = di XOR kiIV；Trudy 知道 ci 和 di，就可以计算出 kiIV：di XOR ci = kiIV；Trudy 得到了加密所用的密钥流 k1IV k2IV k3IV·····；当过后观察到 IV 被重用时，Trudy 就可以破解密文了！**802.11i 增强的安全性：**可以使用各种（较强的）加密算法；提供了密钥分发机制；使用专门的鉴别服务器（而不是 AP）来提供鉴别服务。**802.11i 的操作：1) 发现。**在发现阶段，AP 通告它的存在以及它能够向无线客户结点提供的鉴别和加密形式，客户则请求它希望的特定鉴别和加密形式。**2)相互鉴别和主密钥(MK)生成。**鉴别发生在无线客户和鉴别服务器之间。在这个阶段，接入点基本是起中继的作用，在客户和鉴别服务器之间转发报文。**3)成对主密钥(Pairwise Master Key，PMK) 生成。**MK 是一个仅为客户和鉴别服务器所知的共享密钥，它们都使用 MK 来生成一个次密钥，即成对主密钥(PMK)。鉴别服务器则向 AP 发送该 PMK。客户和 AP 现在具有一个共享的密钥。**4)临时密钥(Temporal Key，TK) 生成。**使用 PMK，无线客户和 AP 现在能够生成附加的、将用于通信的密钥。其中的关键是临时密钥，TK 将被用于执行经无线链路向任意远程主机发送数据的链路级的加密。**运行安全性：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）；**目的是**保护内部网络免受来自外部网络的攻击。**防火墙的类型：**包过滤防火墙，状态检测防火墙，应用网关。**包过滤防火墙：**内部网络通过有包过滤功能的路由器连接到因特网上；路由器对数据包进行逐包过滤，基于以下字段决定转发包还是丢弃包：源 IP 地址、目的 IP 地址；TCP/UDP 源端口号、目的端口号；ICMP 报文类型；TCP SYN 标志和 ACK 标志。**包过滤策略的例子：**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；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 130.207.244.203 的端口 80；防止因特网广播吞噬网络带宽→除 DNS 包和路由器广播包，丢弃其它进入的 UDP 包；防止网络拓扑被探测（traceroute）→丢弃所有外出的 ICMP TTL expired 包；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，这个策略去除了所有从外部发起的所有 TCP 连接，但是允许内部发起 TCP 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（动作，匹配条件）；对于每个进出的包，从上到下地匹配规则。**包过滤防火墙孤立地过滤每个包，仍会允许一些异常的包进入：**例如，允许 dest port = 80，ACK=1 的包进入，哪怕并没有相应的连接存在。**状态检测防火墙**可以跟踪 TCP 连接的状态:跟踪连接的建立（SYN）和关闭（FIN）等状态，判断收到的包**是否有意义**；扩展 ACL，指示在允许放行一个包前需检查连接的状态。**应用网关：**应用网关除了检查网络层及传输层协议头，还检查应用层数据。**例如：**允许特定的内部用户使用 telnet 登录外部主机；所有 telnet 用户必须连接到应用网关；对于授权用户，应用网关建立与目的主机的 telnet 会话，并在 2 个连接之间中继数据；包过滤防火墙**阻塞所有**不源自应用网关的 telnet 连接。**防火墙的局限性：**无法抵御 IP 欺骗攻击：路由器无法知道包是否来自声称的源；应用网关处理开销大，速度慢；每个被代理的应用都需要一个应用网关；应用网关对于用户不透明：客户软件必须设置应用网关的 IP 地址；对于 UDP 包，过滤器或者全部允许，或者全部禁止；和外界的通信强度与网络安全等级是一对矛盾；许多受到高度保护的站点仍然遭到攻击。**入侵检测系统：**防火墙不检查数据包之间的**关联。IDS: intrusion detection system：**深度数据包检查：查看包内容（如检查包中是否包含已知的病毒特征、攻击特征等）；检查多个包之间的关联性：端口扫描、DoS 攻击。**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 不仅需要做深度分组检查，而且必须要将每个过往的分组与数以万计的"特征(signature)"进行比较；这可能导致极大的处理量；将 IDS 传感器进一步向下游放置，每个传感器仅看到该机构流量的一部分，维护能够更容易。**基于特征的 IDS 的一些限制**：它们要求根据以前的攻击知识来产生一个准确的特征，换言之，对不得不记录的新攻击完全缺乏判断力；另一个缺点是，即使与一个特征匹配，它也可能不是一个攻击的结果，因此产生了一个虚假告警；最后，因为每个分组必须与范围广泛的特征集合相比较，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。**基于异常的 IDS** 最大的特点是它们不依赖现有攻占的以前知识；在另一方面，区分正常流量和统计异常流量是一个极具挑战性的问题。