

网络层

向上提供简单灵活的、无连接的、不可靠、尽最大努力交付的数据报服务。
网络在发送分组时不需要先建立连接。不提供服务质量的承诺。

网际协议IP 是TCP/IP体系的最主要协议，配套有 地址解析协议ARP、网际控制报文协议ICMP、网际组管理协议IGMP

IP地址

给连接在网络的主机或路由器分配一个唯一32的位地址，每8位为一组

基本分类IP

IP地址是一个分层架构，第一段是标志主机/路由器所连接网络的网络号net-id，第二段是标志主机/路由器的主机号host-id。

|网络号 |主机号 |，根据划分长度不同分为 A、B、C、D、E 类地址。

A: 0~126; B: 128~191; C: 192~223; D: 224~239; 其余归为E留作备用。

一些特殊的IP地址：

网络号	主机号	源地址使用	目的地址使用	代表的意义
0	0	可以	不可	在本网络上的本主机
0	host-id	可以	不可	在本网络上的某台主机host-id
全1	全1	不可	可以	只在本网络上进行广播
net-id	全1	不可	可以	对net-id上的所有主机进行广播
127	非全0或全1的任何数	可以	可以	用于本地软件环回测试

广播地址：主机位全为1；网络地址：主机位全为0。
此二者无法分配给任何主机。

网络类别	最大可指派的网络数	第一个可指派的网络号	最后一个可指派的网络号	每个网络中最大主机数
A	126 (2 ⁷ -2)	1	126	16777214
B	16383 (2 ¹⁴ -2)	128.1	191.255	65534
C	2097151 (2 ²⁴ -2)	192.0.1	223.255.255	254

划分子网

从两级IP划分为三级IP地址。从主机号host-id借用若干单位作为子网号subnet-id（主机号也相应减少若干位），不改变IP地址原来的网络号net-id。

发送到某个单位指定主机的IP数据报，由路由器接收后再按目的网络号和子网号找到目的子网，并交付目的主机。

子网划分是在分类IP的基础上完成的。规定子网号和主机号不能为全0或全1。子网掩码决定了子网号，即使网络地址写法相同，但子网掩码不同也表示了不同网络。

子网掩码subnetMask:

引入子网掩码来表示IP中的子网部分。

规则：子网掩码长度=32位；左边部分一连串1对应网络号和子网号；右边部分一连串0对应主机号。

默认子网掩码：A 255.0.0.0，B 255.255.0.0，C 255.255.255.0。

(IP地址) AND (子网掩码) = 网络地址

不同的子网掩码得出相同的网络地址，但不同掩码的效果是不同的。

无分类编址CIDR

消除了传统ABC类地址以及子网划分，可更有效地分配IPv4的地址空间。使用各种长度的“网络前缀network-prefix”来代替分类地址中的网络号和子网号。从划分子网的三级回到了两级。

使用“斜线记法slashNotation”，在IP地址后面加一个斜线“/”写上网络前缀所占位数（此数值对应三级编址子网掩码中1的个数）。

eg. 128.14.32.0/20共有 2^{12} 个地址

路由聚合:

一个CIDR地址块可以表示很多地址，让路由表中的一个项目可以表示很多个原来传统分类地址的路由。路由聚合也称为构成超网supernetting。

路由聚合相当于取公共前缀。

私有IP

本地地址（专用地址、私有地址）——仅在某组织内部使用的IP地址，可由组织自行分配，不需要想互联网管理机构申请。

全球地址——全球唯一IP地址，必须向物联网管理结构申请。

三个专用的地址块：A类：10.0.0.0/8，24位块；B类：172.16.0.0/12，20位块；C类：192.168.0.0/16，16位块。

使用专用IP地址的网络称为专用互联网或本地互联网，仅在某个组织内部使用。

网络地址转换NAT：使用本地地址的主机在和外界通信时，都要在NAT路由器上将本地地址转换成全球IP地址，才可以使用互联网

IPv6

将地址从IPv4的32位扩大到128位。

由 40字节的基本首部 和 不超过65535字节的有效负荷 组成。有效载荷允许有零个或多个首部扩展，后面是数据部分。

|基本首部|有效载荷|，

每个16位值用16进制值表示，各值之间用冒号分隔。16进制记法中允许省略前面的0（eg. 0000写为0）

IPv6数据报地址：

地址类型	二进制前缀
未指明地址	::/128
环回地址	::1/128
多播地址	FF00::/8
本地链路单播地址	FE80::/10
全球单播地址	(除上述四种外的其它前缀)

IPv4到IPv6的过渡

逐步严谨，向后兼容（IPv6能够接受和转发IPv4分组，并能够成为IPv4分组选择路由）。有两种向IPv6过渡策略：

- 双协议栈（同时装有IPv4和IPv6协议栈），
- 隧道技术（把IPv6封装成IPv4数据报）。

ARP协议

网络层使用IP地址，数据链路层使用MAC。它们需要去做对应的映射。
ARP的作用，是从网络层使用的IP地址，解析出在数据链路层使用的硬件地址。

RARP是实现从MAC地址到IP地址的映射。每个主机都有ARP高速缓存(ARP Cache)，存放网络上各主机和路由器的IP地址到硬件地址的映射表。

广播和单播硬件地址：设A向B发送，建立请求和响应的过程（解析）

- 1. ARP广播：源IP是A，目的IP是B；源MAC是A，目的MAC是广播
- 2. ARP响应(单播)：源IP是B，目的IP是A；源MAC是B，目的MAC是A

如果A和B不在同一个网络，则需要通过ARP找到位于本网络的某个路由器硬件地址，然后把分组发给该路由器，并让其分组转发给下一个网络。剩下的工作由下一个网络完成。

ICMP协议

有效地转发IP数据报和高交付成功的机会，允许主机或路由器报告差错情况和提供有关异常情况的报告。报文格式如下：

| 类型0~7 | 代码8~15 | 检验和16~31 | (前四个字节都一样，取决于ICMP报文类型)
| ICMP数据部分(长度取决于类型) |

ICMP差错报告报文

终点不可达，时间超过，参数问题，改变路由(重定向Redirect)，源抑制

不发送ICMP差错报文的情况：

对ICMP差错报告报文不再发送；对第一个分片的数据报片的所有有序数据报片都不发送；对具有多播地址的数据报都不发送；对具有特殊地址（127.0.0.0或0.0.0.0等）的数据报都不发送。

ICMP询问报文

请求和回答报文，事件请求和回答报文。

ICMP的应用

PING用来测试两台主机的连通性和联通质量。

Tracerout用于追踪一个分组从原点到终点的路径（windows命令为tracert）

IP首部格式

IP数据报由首部和数据两部分组成（首部采用固定时是20字节，所有IP数据报都要有。数据部分长度可变，1-40字节不等）。

当数据量很大时允许分片：标志flag占三位，取前两位有意义。最低位是MF，=1表示后面还有分片，=0表示最后一个分片；中间位是DF，=0时允许分片。

生存时间TTL，可通过路由器数的最大值（每经过一个路由器将会-1，为0时丢弃，并向源告知超时）。

IP组播

- IP多播：实现更好的一对多通信。对外只需发送一份，让多播路由器复制后发送给多个成员。尽最大努力交付
 - 软件多播：在网络层实现多播
 - 硬件多播：TCP/IP协议使用到的以太网地址块范围是01-00-5E-00-00-00 ~ 01-00-5E-7F-FF-FF。
 - IGMP协议：利用网络组管理协议把网络上的路由器协同工作，一遍以最小代价完成多播

路由选择

当两台非直连计算机需要经过几个网络通信时，路由器提供路由选择协议来开辟一个最佳网状连接路径。路径必须需是正确且完整，计算上简单，能适应通信量和网络拓扑变化（要有自适应性，稳定性，公平性，最佳性）。

静态路由选择策略：简单，开销小，不能适应网络状态变化。

动态路由选择：自适应，实现复杂，开销大。

自治系统AS

在单一技术下管理的一组路由器，使用AS内部路由选择协议和共同的度量来确定分组在该AS内的路由，同时使用一种AS之间的路由选择协议用于确定分组在AS之间的路由。

虽然一个AS会使用多种内部路由选择协议和度量，但一种AS对另一种AS表现出的的是一个单一和一致的路由选择策略。有两类主要协议：

内部网关协议IGP：自治系统内使用的路由选择协议，使用最多（例如RIP和OSPF）。

外部网关协议EGP：将一个自治系统的路由选择信息传递到另一个自治系统。BGP-4使用最多。

协议	层次	下层	范围	原理
RIP	应用层	UDP	AS内	距离向量

协议	层次	下层	范围	原理
OSPF	网络层	IP	AS内	链路状态
BGP	应用层	TCP	AS间	路径向量

内部网关协议IGP

RIP

分布式的、基于距离向量的路由协议。要求每个路由器都要维护从自己到其它每一个目的网络的距离记录。由此形成一个路由表：

|目的网络 |下一跳 |距离 |。

一个路由器下的网络到另一个直连网络的距离定义为1，从一个路由器到非直连网络的定义为所经过路由器数+1——也被称作跳数。

RIP认为一个好的路由即是经过的路由器数目最少——即为“距离”，允许一条路径最多只包含15个处理器。“距离”到达最大值16时相当于不可达。所以RIP只适用于小型网络。

刚开始工作时，只知道直连网络距离为1，且路由表为空。之后每个路由器也只和数目有限的相邻路由器交换，并更新路由信息。经过若干更新后所有路由器都会知道到达本自治系统中任何网络的最短距离，也就是“收敛”。

RIP仅和相邻路由器按固定时间间隔交换信息，交换的是本路由器的全部信息（路由表）。使用的是Bellman-Ford算法。RIP让互联网中的所有路由器都和自己相邻的路由器不断交换信息，最终每个路由器都知道相互的距离关系。

RIP好消息传得快，坏消息传的慢。当网络出现故障，要经过较长时间（数分钟）才能将此信息传递到所有路由器。

OSPF

基于Dijkstra算法找最短路径，采用分布式链路状态协议。

使用洪泛法向本治系统中所有路由器发送信息。发送的信息是与本路由器相邻的所有路由器的链路状态，链路状态说明本路由器都和哪些路由器相邻，以及该链路的“度量”。只有链路状态变化时才会发送信息。

适用于大型动态网络。

链路状态数据库：所有路由器建立起的全网拓扑结构图。

OSPF使用层次结构区域划分。在上层的称作主干区域，标识符规定为0.0.0.0，作用是连通下层区域。

OSPF的五种分组（确定可达性，达到数据库同步，新情况同步）：问候Hello分组；数据库描述databaseDescription分组；链路状态请求linkStateRequest分组；链路状态更新linkStateUpdate分组；链路状态确认linkStateAcknowledgment分组。

外部网关协议BGP-4

实现在AS之间交换路由信息的协议，简写BGP。力求寻找一条能够到达目的网络且比较好的路由（不能兜圈子），而且并非要寻找一条最佳路由。

BGP发言人：一个AS至少选择一个路由器作为该AS的BGP发言人。两个发言人通过一个共享网络连接在一起，而BGP发言人往往是BGP边界路由器。发言人交换路径向量。
交换过程要建立TCP连接，在此连接上交换BGP报文来建立BGP会话并交换路由信息。

BGP的四种报文：打开OPEN，更新UPDATE，保活KEEPALIVE，通知NOTIFICATION。

数据转发流程

IP层转发分组流程

按主机错在网络地址制作路由表，那么每个路由器中的路由表就只包含四个项目。在路由表中对每一条路由，最主要的是(目的网络地址，下一跳地址)。

特定主机路由：为特定目的主机指明一个路由，方便网络管理人员控制和测试网络。
默认路由：不会进行任何形式的检查（0.0.0.0/0），适用于只有一个路由器和互联网连接的小型网络。

查找路由表：根据目的网络地址就能确定下一跳路由器，IP数据报最终一定可以找到目的主机所在的网络上的路由器，只有到达最后一个路由器时才试图向目的主机进行直接交付。

使用子网分组转发

划分子网的情况下从IP地址不能唯一地得出网络，因为网络地址取决于那个网络所采用的子网掩码且数据报首部不包含。因此分组转发算法需要做改动。

最长前缀匹配

又被称为最长匹配or最佳匹配。

使用CIDR时，路由表的每个表项由“网路前缀”和“下一跳地址”组成。查找路由表时可能会得到不止一个匹配结果。
应当从匹配结果中选择具有最长网络前缀的路由：最长前缀匹配longest-profixMatching。网络前缀越长，地址块越小，路由越具体。

路由器

一种典型的具有多输入多输出端口的网络层设备，关键设备，主要作用：联通不同网络，路由选择，分组转发。

互联网设备总结：

设备名	中继器	集线器	网桥	交换机	路由器
冲突域	×	×	√	√	√
广播域	×	×	×	×	√
层次	物理层	物理层	数据链路层	数据链路层	网络层
工作原理	信号再生	信号再生	根据MAC地址转发	根据MAC地址转发	根据IP地址转发

SDN

软件定义网络software-DefinedNetworking是一种网络架构和管理方法，通过将网络控制平面controlPlane与数据转发平面dataPlane分离，实现网络的灵活性、可编程性和集中化管理。SDN将网络硬件设备的控制功能集中到一个中央控制器中。

其核心思想是将网络的控制逻辑从网络设备中抽离出来，使网络设备成为可编程的数据转发平面。网络管理员可以通过中央控制器对整个网络进行集中管理和编程。

SDN的优势包括简化网络管理、提高网络灵活性、降低网络成本和加速网络创新等，使网络能够更好地适应不断变化的需求和应用场景。

北向API面向用户，南向API面向资源。

移动网络

1. 归属代理：一个移动节点归属网上的路由器至少有一个接口在归属网上。当移动节点离开归属网时，通过IP通道把数据报传给移动节点，并且负责维护移动节点当前的位置信息
2. 外区代理：移动节点当前所在网络上的路由器，向已登记的移动节点提供选路服务
3. 归属地址：识别端到端连接的静态地址。不论移动节点连接到网络何处，其归属地址保持不变
4. 转交地址：隧道重点地址。可能是外区代理转变地址，也可能是主流本地的转变地址。外区代表转交地址是外区代理的一个地址，移动节点利用其登记
5. 位置登记：移动节点必须向其所在位置进行登记
6. 代理发现：移动节点必须首先找到一个移动代理，便于随时随地与其他节点通信
7. 隧道技术：移动节点在外区网时，归属代理需将其原始数据包转发给已登记的外区代理

补充

- 在ISO/OSI参考模型中，网络层的主要功能是路由选择、拥塞控制、网络互联
- 常用的十进制转二进制：128 = 1000 000；192 = 1100 000；224 = 1110 0000；240 = 1111 0000；248 = 1111 1000；252 = 1111 1100；254 = 1111 1110；255 = 1111 1111
- .../30是点对点，只有2个有效位，通常作为PPP协议的两端
- 网络优先：当每个网络中主机数相等或等大时
- 主机优先：每个网络中主机数不等，优先分配大的网络
- 局域网的分配可以带入哈夫曼算法
- IPv6零压缩只能用一次
- 直接为ICMP提供服务协议的是IP
- IP数据报报头中，报头长度字段以32比特为计数单位，总长度字段以8比特为计数单位
- 互联网中，IP数据报传输的路径经由的主机和中途路由器，源主机和中途路由器通常不知道完整路径