IP 网际协议

1. 协议名称

中文名：网际协议

英文名：Internet Protocol(缩写：IP)

2. 协议作用

路由转发、“存储，交换，转发”、拥塞控制、呼叫准入

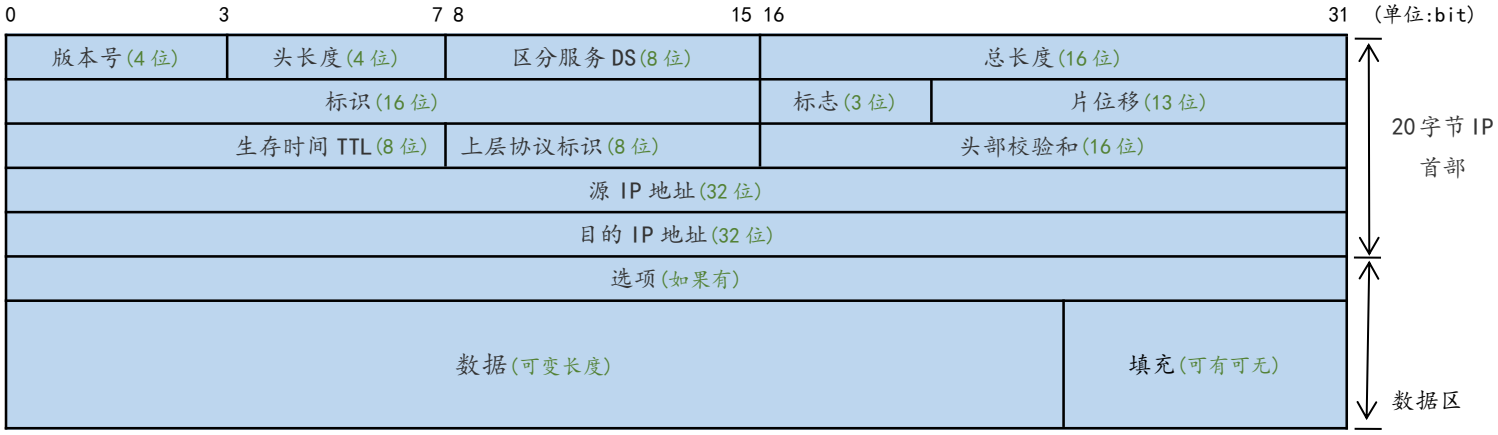
3. 协议工作在 TCP/IP 四层模型哪个层次？

网络层。

4. 协议性质

不可靠、无连接 （TCP/IP 四层网络模型中网络层的性质）

5. 协议格式

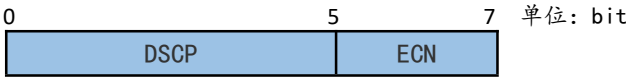


5.1 版本号(Version)：占 4bit，表示 IP 协议版本。 ($0100_2=4_{10}$ 表示 IPv4)，($0110_2=6_{10}$ 表示 IPv6)。

5.2 头长度(Head Length)：占 4bit，表示整个数据报首部的长度。该长度是以 4 字节为一个计数单位，普通 IP 数据报（没有任何选项）该字段的值是 $0101_2 = 5_{10}$ 即 5x4=20 字节。4bit 最大表示 15_{10} 所以 IP 数据报首部最大 15x4=60 字节。

5.3 区分服务(Differentiated Services)：占 8bit，规定本数据报的处理方式，用来获得更好的服务。这个字段在旧标准中叫做服务类型(Type of Service),但实际上一直没有被使用过，1998 年 IETF 把这个字段改名为区分服务 DS(Differentiated Services)。只有在使用区分服务时,这个字段才起作用。

区分服务将 8bit 分成两部分，DSCP(Differentiated Services Codepoint，差分服务代码点)占 6bit，ECN(Explicit Congestion Notification，显示拥塞通知)占 2bit。



DSCP：取值范围为 0-63，0 优先级最低，63 优先级最高。

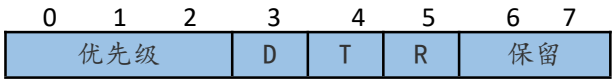
ECN：支持 ECN 的主机发送数据报时将 ECN 设置为 01 或者 10。对于支持 ECN 的主机发送的包，如果路径上的路由器支持 ECN 并且经历拥塞，它将 ECN 域设置为 11。如果该数值已经被设置为 11，那么下游路径上的路由器不会修改该值。

00：发送主机不支持 ECN

01 或者 10：发送主机支持 ECN

11：路由器正在经历拥塞

服务类型(Type of Service)将8bit分成5个子域：



单位：bit

(1) **优先级(Priority)**：表示范围0-7，数越大，表示该数据报优先权越高。网络中路由器可以使用优先权进行拥塞控制，如当网络发生拥塞时可以根据数据报的优先权来决定数据报的取舍。

0 尽力服务数据

1 中优先级数据

2 高优先级数据

3 呼叫信号

4 视频会议

5 语音

6 预留

7 预留

(2) **短延迟位D(Delay)**：该位置1时，数据报请求以短延迟信道传输，0表示正常延时。

(3) **高吞吐量位T(Throughput)**：该位置1时，数据报请求以高吞吐量信道传输，0表示普通。

(4) **高可靠位R(Reliability)**：该位置1时，数据报请求以高可靠性信道传输，0表示普通。

(5) **保留位**。

目前在Internet中使用的TCP/IP协议大多数情况下网络并未对TOS进行处理，但在实际编程时，有专门的函数来设置该字段的各域。一些重要的网际应用协议中都设置了建议使用的TOS值：

应用程序	短延迟位D	高吞吐量T	高可靠性位	低成本位	十六进制值	特性
Telnet	1	0	0	0	0x10	短延迟
FTP 控制	1	0	0	0	0x10	短迟延
FTP 数据	0	1	0	0	0x08	高吞吐量
TFTP	1	0	0	0	0x10	短迟延
SMTP 命令	1	0	0	0	0x10	短迟延
SMTP 数据	0	1	0	0	0x08	高吞吐量
DNS UDP 查询	1	0	0	0	0x10	短迟延
DNS TCP 查询	0	0	0	0	0x00	普通
DNS 区域传输	0	1	0	0	0x08	高吞吐量
ICMP 差错	0	0	0	0	0x00	普通
ICMP 查询	0	0	0	0	0x00	普通
SNMP	0	0	1	0	0x04	高可靠性
IGP	0	0	1	0	0x04	高可靠性
NNTP	0	0	0	1	0x02	低成本

5.4 总长度(Total Length)：占16bit，总长度字段是指整个IP数据报的长度(首部+数据)，以字节为单位。利用头部长度(Head Length)和总长度(Total Length)就可以计算出IP数据报中数据内容的起始位置和长度。由于该字段长度为16位二进制数，因此理论上IP数据报最长可达2^{16}共65536个字节（事实上受物理网络的限制，要比这个数值小很多）。

5.5 标识(Identification)：占16bit，标识字段唯一地标识主机发送的每一份数据报。通常每发送一份数据它的值就会加1。同时方便接收方进行分片重组(每份数据报可能被分片)。

5.6 标志(Flags)：占 3bit,是否分片、帮助分片重组。

保留段位(Reserved bit)： 占 1bit,保留未使用。

不分片段位(Don't fragment)： 占 1bit,取“0”： 允许分片，取“1”： 不能分片。

更多段位(More fragment)： 占 1bit,取“0”:数据包后面没有包，该包为最后的包、
取“1”:数据包后面有更多的包。

5.7 片位移(Fragement offset)： 占 13bit,与更多段位(More fragment)组合，帮助接收方组合分片的报文，以 8 个字节为单位。
允许最多（2^{13}-1）×8 =65528 个字节，这将超过 65,535 个字节的最大 IP 报文长度与包括报头长度（ 65528+20=65548 个字节）。

5.6 生存时间(TTL:Time to Live)：占 8bit，表示数据报可以在网络中传输的最长时间。实际应用中把生存时间字段设置成了数据报可以经过的最大路由器数。TTL 的初始值由源主机设置（通常为 32、64、128 或 256），经过路由器或三层交换机 TTL 值减 1。当该字段为 0 时，数据报就丢弃，并发送 ICMP 报文通知源主机，因此可以防止进入一个循环回路时，数据报无休止地传输下去。

5.7 上层协议标识(Protocol)：占 8bit，表示 IP 协议的上层协议。

二进制	十进制	协议	说明
00000000	0	无	保留
00000001	1	ICMP	网际控制报文协议
00000010	2	IGMP	网际组管理协议
00000011	3	GGP	网关-网关协议
00000100	4	无	未分配
00000101	5	ST	流
00000110	6	TCP	传输控制协议
00001000	8	EGP	外部网关协议
00001001	9	IGP	内部网关协议
00001011	11	NVP	网络声音协议
00001111	17	UDP	用户数据报协议
01011000	88	EIGRP	增强内部网关路由协议
01011001	89	OSPF	开放式最短路径优先(动态路由协议)

5.8 头部校验和(Header checksum)：占 16bit，用于协议头数据有效性的校验，可以保证 IP 报头区在传输时的正确性和完整性。头部检验和字段是根据 IP 协议头计算出的检验和，它不对头部后面的数据进行计算。

校验原理：

发送端首先将检验和字段置 0，然后对头部中每 16 位二进制数进行反码求和的运算，并将结果记录在校验和字段中。 由于接收方在计算过程中包含了发送方放在头部的校验和，因此，如果头部在传输过程中没有发生任何差错，那么接收方计算的结果应该是全 1。

5.9 源地址(Source)：占 32bit，表示发送端 IP 地址。

5.10 目的地址(Destination)：占 32bit，表示接收端 IP 地址。

5.11 选项(Option)：IP 选项一般用不到，IP 选项主要用于控制和测试两大目的。作为选项，用户可以使用也可以不使用 IP 选项，但作为 IP 协议的组成部分，所有实现 IP 协议的设备能处理 IP 选项。在使用选项的过程中 ，有可能造成数据包头部不是 32bit 的整数倍，那么则需要“0”填充。增加首部的可变部分是为了增加 IP 数据报的功能,但这同时也使得 IP 数据报的首部长度成为可变的.这就增加了每一个路由器处理数据报的开销,实际上这些选项很少被使用。新

的 IP 版本 IPv6 就将 IP 数据报的首部长度做成固定的。

最后一个字段是任选项，是数据报中的一个可变长的可选信息。目前，这些任选项定义如下：

A. 安全和处理限制（用于军事领域，详细内容参见 RFC 1108[Kent 1991]）

B. 记录路径（让每个路由器都记下它的 IP 地址，见 7.3 节）

C. 时间戳（让每个路由器都记下它的 IP 地址和时间，见 7.4 节）

D. 宽松的源站选路（为数据报指定一系列必须经过的 IP 地址，见 8.5 节）

E. 严格的源站选路（与宽松的源站选路类似，但是要求只能经过指定的这些地址，不能经过其他的址）。

这些选项很少被使用，并非所有的主机和路由器都支持这些选项.

5.12 数据(Data)：可变长度。IP 数据报总长度字段占 16bit，并以字节为单位，故 IP 数据报最大长度 2^{16} 65536 字节，IP 首部 20 字节。理论上“数据(Data)”字段可以有 65536 减去 20 ，65516 字节。

5.13 填充(Padding)：可有可无。使用全 0 的填充字段将不足的数据部分补齐成为 4 字节的整数倍。

6.分片、分组、IP 数据报

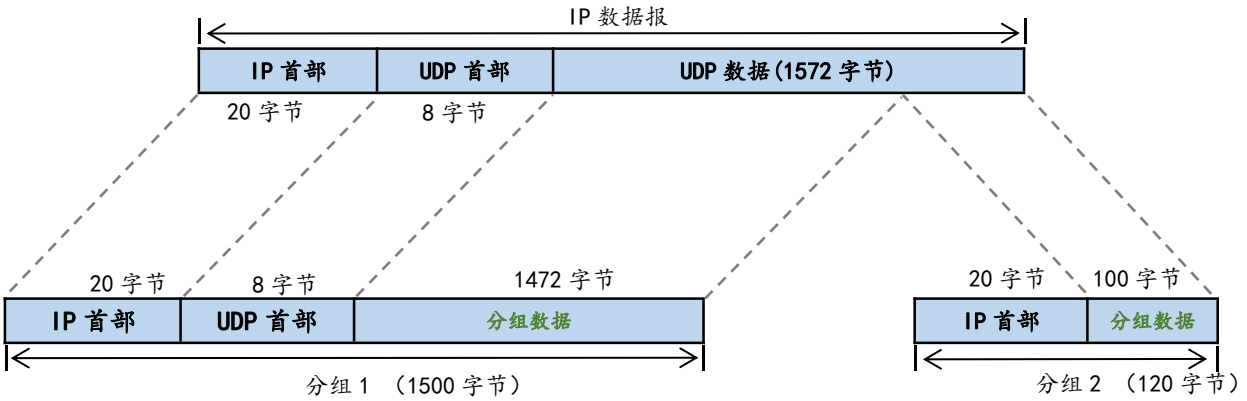
分片：物理网络层一般要限制每次发送数据帧的最大长度。任何时候 IP 层接收到一份要发送的 IP 数据报时，它要判断向本地哪个接口发送数据（选路），并查询该接口获得其 MTU。IP 把 MTU 与 IP 数据报长度进行比较，如果需要则进行分片。分片可以发生在原始发送端主机上，也可以发生在中间路由器上。

当 IP 数据报被分片后，每一片都成为一个分组，具有自己独立的 IP 首部，并在选择路由时与其他分组独立。这样，当数据报的这些片到达目的端时有可能会失序，但是在 IP 首部中有足够的信息让接收端能正确组装这些数据报片。

需要注意的是，在分片的数据中，传输层的首部只会出现在第一个分片中。因为传输层的数据格式对 IP 层是透明的，传输层的首部只有在传输层才会有它的作用，IP 层不知道也不需要保证在每个分片中都有传输层首部。所以，在网络上传输的数据包是有可能没有传输层首部的。

IP 数据报：是指 IP 层端到端的传输单元（在分片之前和重新组装之后）。

分组：是指在 IP 层和数据链路层之间传送的数据单元。一个分组可以是一个完整的 IP 数据报，也可以是 IP 数据报的一个分片。



Tips:IPv6 使用路径 MTU 发现机制，路由器不再分片。

7. **如何分片：**

IP 首部	数据(5000 字节)

20 字节

MTU=1500Byte(MTU 不一定固定就是 1500byte，此处以 1500byte 为例讲解)

分组	总字节数	头字节	数据字节	Identification	DF	MF	片位移	片位移计算
1	1500	20	1480	0xc62f	0	1	0	**0**
2	1500	20	1480	0xc62f	0	1	185	(0+1480/8)=**185**
3	1500	20	1480	0xc62f	0	1	370	(185+1480/8)=**370**
3	580	20	560	0xc62f	0	0	555	(370+1480/8)=**555**

8. **如何分片重组：**

根据片位移的大小顺序组合接收到的数据分组，重组为完整的数据包。