

TCP/IP 详解卷 1：协议 - 3

阅读目的：了解链路层基本组成、设计历史以及链路层使用到的协议（PPP 等）。

阅读时间：5 小时

阅读概况：第 3 章

第 3 章 链路层

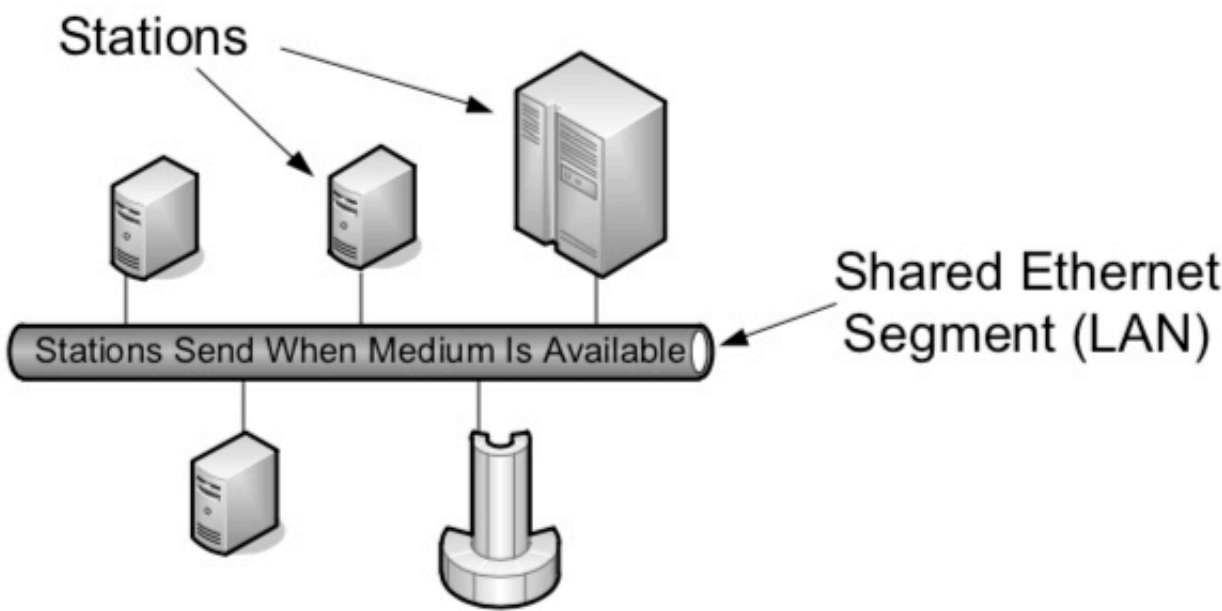
1. 以太网标准

TCP/IP 协议族中设计链路层的目的是为 IP 模块发送和接收 IP 数据报。它可用于携带一些支持 IP 的辅助性协议（例如 ARP）。TCP/IP 支持多种不同的链路层，它依赖于使用的网络硬件类型：

- 有线局域网。例如以太网，城域网(MAN)（有线电视和 DSL 连接）；
- 有线语音网络。例如支持调制解调器的电话线；
- 无线网络，例如 Wi-Fi (无线局域网)；
- 基于蜂窝技术的各种无线数据服务，例如 HSPA、EV-DO、LTE 和 WiMAX。

以太网和 IEEE 802 局域网/城域网标准

以太网这个术语通常指一套标准，第一个常见格式的以太网目前被称为“10Mb/s 以太网”或“共享以太网”，它被 IEEE 采纳(轻微修改)为 **802.3 标准**。这种网络的结构通常如图所示：



由于多个站共享同一网络，该标准需要在每个以太网接口实现一种分布式算法，以控制一个站发送自己的数据。这种特定方法称为 **带冲突(或称碰撞)检测的载波侦听多路访问 (CSMA/CD)**，它协调哪些计算机可访问共享的介质(电缆)，同时不需要其他特殊协议或同步。

采用 CSMA/CD，一个站(例如计算机)首先检测目前网络上正在发送的信号，并在网络空闲时发送自己的帧。这是协议中的“载波侦听”部分。如果其他站碰巧同时发送，发生重叠的电信号被检测为一次“碰撞”。

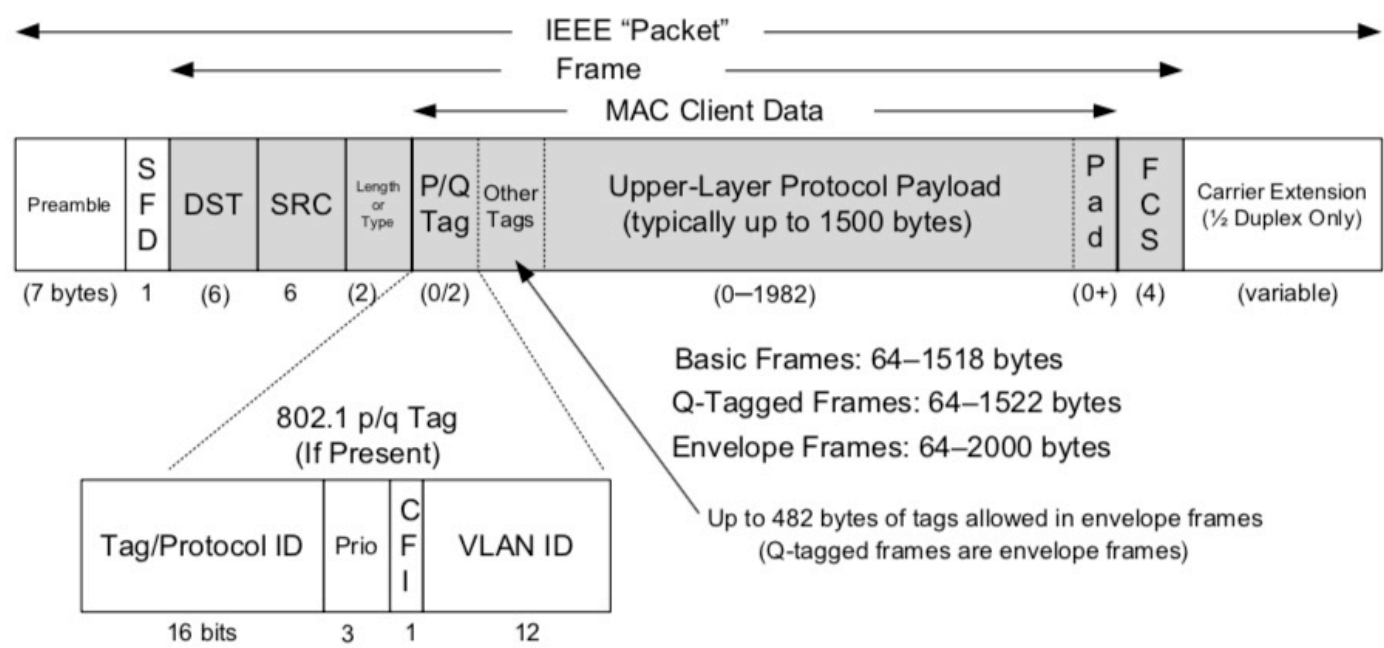
如果发生碰撞，每个站会等待一个随机时间，然后再次尝试发送。这个时间量的选择依据一个统一的概率分布，随后每个碰撞被检测到的时间长度加倍。最终，每个站会得到机会发送，或者在尝试一定次数(传统以太网为 16)后超时。采用 CSMA/CD 在任何给定的时间内，网络中只能有一个帧传输。

如 CSMA/CD 这样的访问方法更正式的名称为 **介质访问控制(MAC)协议**。MAC 协议有很多类型，有些基于每个站尝试独立使用网络(例如 CSMA/CD 的基于竞争的协议)，有些基于预先安排的协调(例如依据为每个站分配的时段发送)。

由 IEEE 标准化为一种 CSMA/CD 网络，称为 **802.3**。在 IEEE 标准中，带 802 前缀的标准定义了局域网和城域网的工作过程。当前最流行的 802 标准包括 802.3(以太网)和 802.11 (WLAN/Wi-Fi)。

2. 以太网帧格式

所有的以太网(802.3)帧都基于一个共同的格式。在原有规范的基础上帧格式已被改进以支持额外功能。



以太网帧开始是一个 **前导字段**，接收器电路用它确定一个帧的到达时间，并确定编码位(称为时钟恢复)之间的时间量。由于以太网是一个异步的局域网(即每个以太网接口卡中不保持精确的时钟同步)，从一个接口到另一个接口的编码位之间的间隔可能不同。

前导是一个公认的模式(典型值为 `0xAA`)，在发现帧起始分隔符(SFD)时，接收器使用它“恢复时钟”。SFD的固定值为 `0xAB` 。

这个基本的帧格式包括 48 位(6字节)的 **目的地址(DST)** 和 **源地址(SRC)** 字段。

这些地址有时也采用其他名称，例如“MAC地址”、“链路层地址”、“802 地址”、“硬件地址”或“物理地址”。

以太网帧的目的地址也允许寻址到多个站点(称为“广播”或“组播”)。广播功能用于 ARP 协议，组播功能用于 ICMPv6 协议，以实现网络层地址和链路层地址之间的转换。

源地址后面紧跟着一个 **类型字段**，或一个 **长度字段**。在多数情况下，它用于确定头部后面的数据类型。TCP/IP 网络使用的常见值包括 IPv4 (`0x0800`)、IPv6 (`0x86DD`) 和 ARP (`0x0806`)。 `0x8100` 表示一个 Q 标签帧(可携带一个“虚拟局域网”或 802.1q 标准的 VLANID)。

最初的 IEEE (802.3)规范将长度/类型字段作为长度字段而不是类型字段使用。因此，这个字段被重载(可用于多个目的)。关键是看字段值。目前，如果字段值大于或等于 1536，则该字段表示类型，它是由标准分配的超过 1536 的值。如果字段值等于或小于 1500，则该字段表示长度。 [ETHERTYPES] 给出了类型的完整列表。

帧的大小

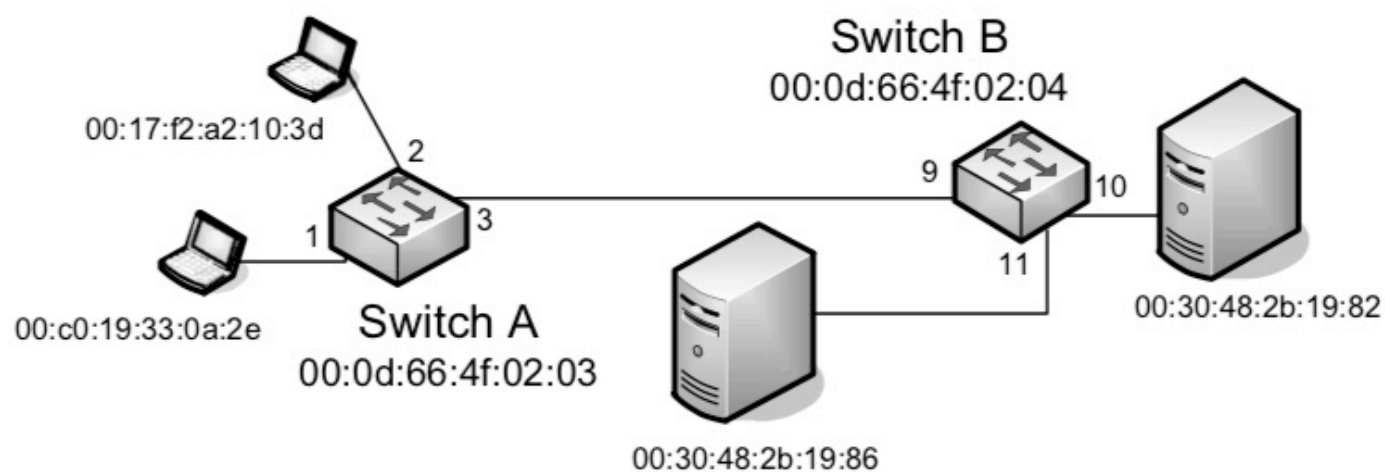
以太网帧有最小和最大尺寸。最小的帧是 **64 字节**，要求数据区(有效载荷)长度(无标签)最小为 **48 字节**。当有效载荷较小时，填充字节(值为 0)被添加到有效载荷尾部，以确保达到最小长度。

传统以太网的最大帧长度是 1518 字节(包括 4 字节 CRC 和 14 字节头部)。

选择这个值出于一种折中:如果一个帧中包括一个错误(接收到不正确的 CRC 校验)，只需重发 1.5KB 以修复该问题。另一方面， MTU 大小限制为 1500 字节。为了发送一个更大的消息，则需要多个帧(例如对于 TCP/IP 网络常用的较大尺寸 64KB，需要至少 44 个帧)。

3. 网桥和交换机

IEEE 802.1d 标准规定了网桥的操作，**交换机本质上是高性能的网桥**。网桥或交换机用于连接多个物理的链路层网络(例如一对物理的以太网段)或成组的站。最基本的设置涉及连接两个交换机来形成一个扩展的局域网，如图所示：



当第一次打开一个交换机(网桥)时，它的数据库是空的，因此它不知道除自己之外的任何站的位置。当它每次接收到一个目的地不是自己的帧时，它为除该帧到达的端口之外的所有端口做一个备份，并向所有端口发送这个帧的备份。

如果交换机(网桥)未学习到站的位置，每个帧将会被交付到每个网段，这样会导致不必要的开销。学习能力可以显著降低开销，它是交换机和网桥的一个基本功能。

Station	Port
00:17:f2:a2:10:3d	2
00:c0:19:33:0a:2e	1
00:0d:66:4f:02:03	
00:0d:66:4f:02:04	3
00:30:48:2b:19:82	3
00:30:48:2b:19:86	3

Switch A's Database

Station	Port
00:17:f2:a2:10:3d	9
00:c0:19:33:0a:2e	9
00:0d:66:4f:02:03	9
00:0d:66:4f:02:04	
00:30:48:2b:19:82	10
00:30:48:2b:19:86	11

Switch B's Database

Figure 3-9 Filtering databases on switches A and B from Figure 3-8 are created over time (“learned”) by observing the source address on frames seen on switch ports.

4. 生成树协议

*单独整理

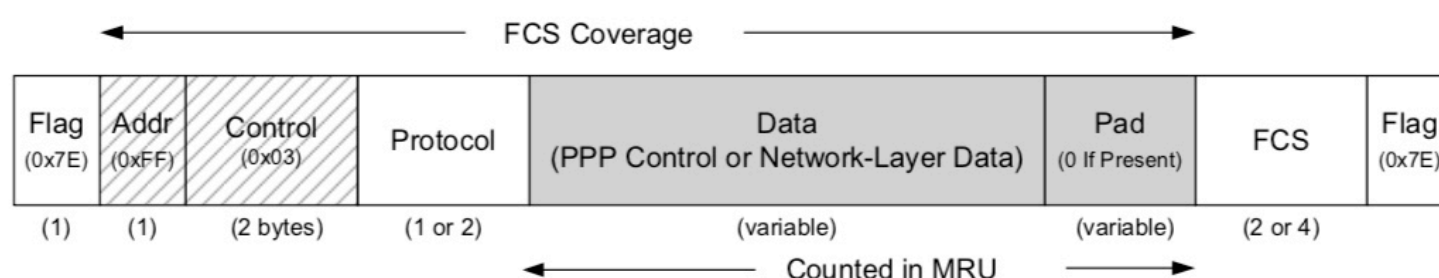
5. 点到点协议

点到点协议（PPP）是一种在串行链路上传输 IP 数据报的流行方式，从低速的拨号调制解调器到高速的光链路都被广泛使用，也可分配 Internet 系统的参数。

PPP 实际上是一个协议集合，而不是一个单一的协议。它支持建立链接的基本方法 -- **链路控制协议(Link Control Protocol, LCP)**，以及一系列 NCP 协议，在 LCP 建立了基本链路之后，用于为各种协议(包括 IPv4、IPv6 和非 IP 协议)建立网络层链路。一些相关标准涉及对 PPP 的压缩和加密控制，以及在链接建立后的一些认证方法。

链路控制协议

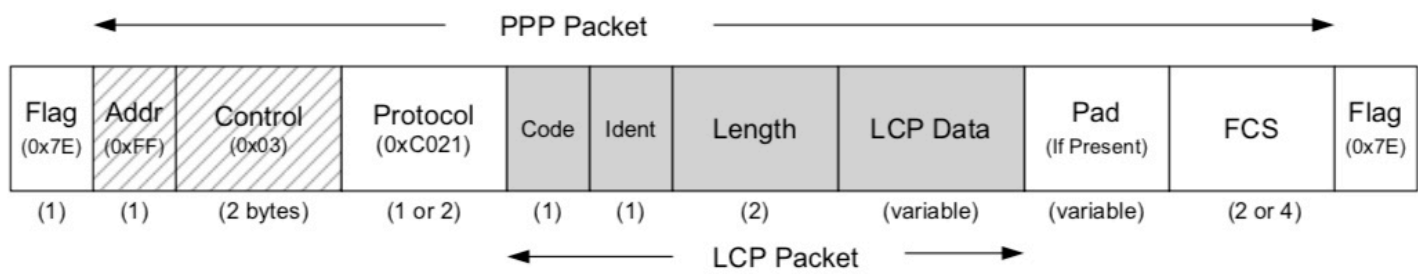
PPP 的 LCP 用于在点到点链路上建立和维护低层的双方通信路径。因此 PPP 操作只需关注一条链路的两端，它不需要像以太网和 Wi-Fi 的 MAC 层协议那样处理共享资源访问的问题。



通常，PPP 被配置为不重传。

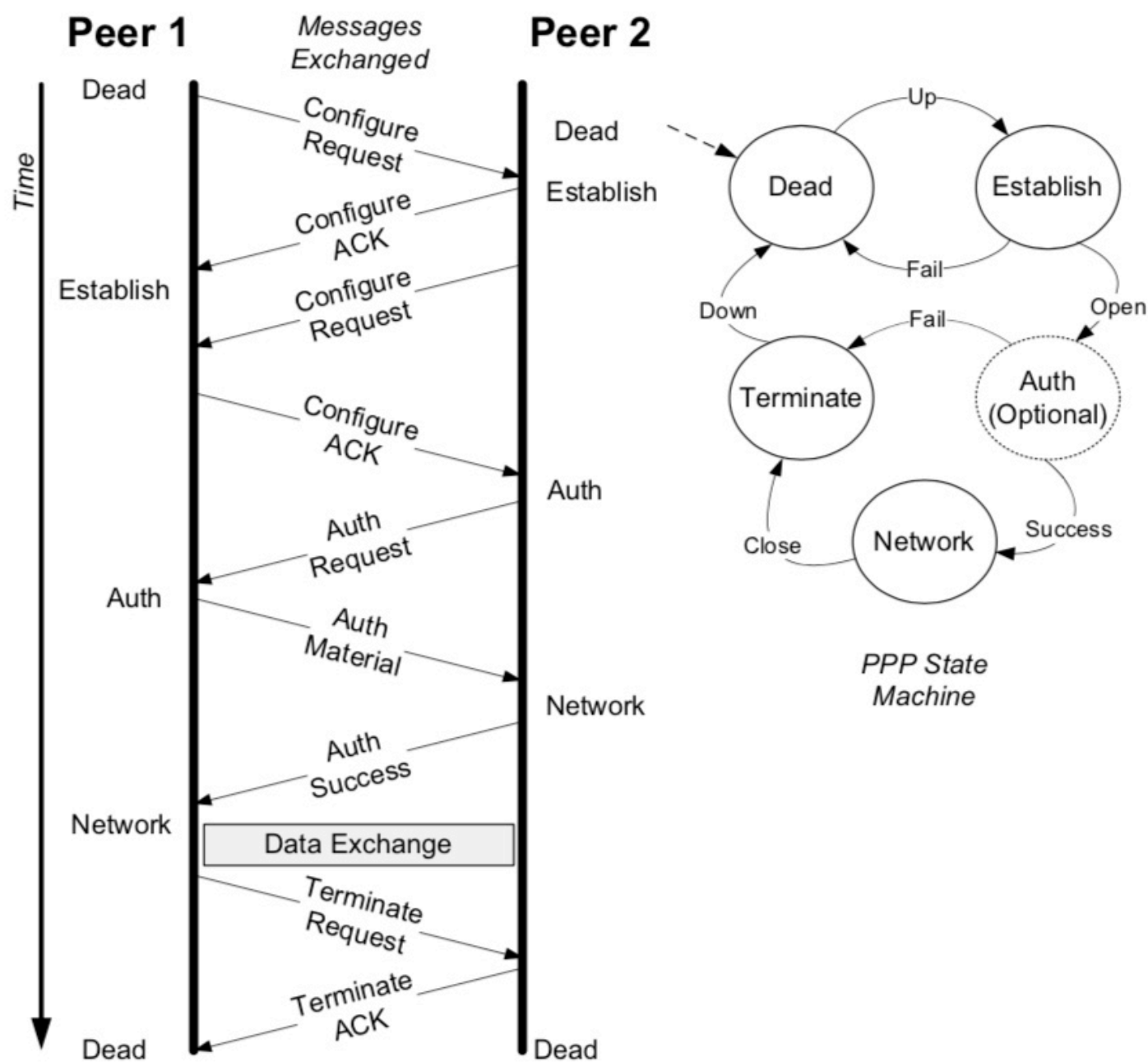
LCP 操作

LCP 在基本 PPP 分组之上进行了简单的封装。如图所示。



LCP 的主要工作是使一条点到点链路达到最低要求。**配置** 消息使链路两端开始基本配置过程，并建立商定的选项。**终止** 消息用于在完成后清除一条链路。

一旦底层协议表明一个关联变为活跃(例如调制解调器检测到载波)，则认为这个链路已被建立。链路质量测试包含 **链路质量报告** 和 **确认交换**，它也可以在此期间完成。如果链接需要认证(这是很常见的)，例如当拨号到一个 ISP 时，可能需要一些额外的信息交换，以认证链路上的一方或双方的身份。当底层协议或硬件表明一个关联已停止(例如载波消失)，或发送一个链路终止请求，并从对方接收到一个终止响应，则认为这个链路已被终止。



6. 环回

但在很多情况下，客户机可能希望使用 Internet 协议(例如 TCP/IP)与同一计算机上的服务器通信。可通过网络层的 **环回(回送)** -- 通常使用一个虚拟的环回网络接口来实现。

它就像一个真正的网络接口，但实际上是一个由操作系统提供的专用软件，可通过 TCP/IP 与同一主机的其他部分通信。

类 UNIX 系统(包括Linux)为环回接口分配的 IPv4 地址为 127.0.0.1 (IPv6 地址为::1)，为它分配的名称为 localhost。发送到环回接口的 IP 数据报不会出现在任何网络中。在 Linux 中，环回接口被称为 lo。

7. MTU 和 Path MTU

在很多链路层网络(例如以太网)中，携带高层协议 PDU 的帧大小是有限制的。以太网有效载荷的字节数通常被限制为 1500，PPP 通常采用相同大小以保持与以太网兼容。链路层的这种特征被称为 **最大传输单元 (MTU)**。大多数的分组网络(例如以太网)都有固定的上限。

如果 IP 需要发送一个数据报，并且这个数据报比链路层 MTU 大，则 IP 通过分片将数据报分解成较小的部分，使每个分片都小于 MTU。

当同一网络中的两台主机之间通信时，本地链路的 MTU 在会话期间对数据报大小有直接影响。当两台主机之间跨越多个网络通信时，每条链路可能有不同大小的 MTU。**在包含所有链路的整个网络路径上，最小的 MTU 称为 Path MTU。**

任何两台主机之间的 Path MTU 不会永远不变，这取决于当时使用的路径。如果网络中的路由器或链路故障，MTU 可能改变。另外，路径通常不对称(主机 A 到 B 路径可能不是 B 到 A 的反向路径)，Path MTU 不需要在两个方向上相同。