

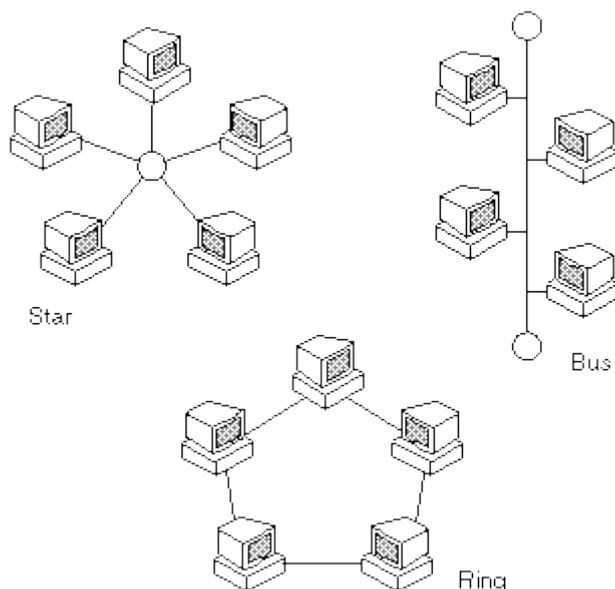
以太网速超越极限：3Mbps，100Gbps

1970年代，虽然盛行着电视这种又落后又简陋的媒体，但还不至于苍白得完全没有现代通讯系统的色彩。的确，当时美国国防部高级研究计划署开发的最先进的阿帕网（The Advanced Research Projects Agency Network，世界上第一个运营的封包交换网络，全球互联网的始祖），其调制解调器速度也仅为50 Kbps，且体积足有冰箱般巨大；至于流传甚广的贝尔103型调制解调器，其传输速度也只有可怜的每秒300比特。幸好随着计算机总量的增加，长途数字通信已经相当普及了。终端机也可以通过相对较短的几根简单线路或者更复杂的多点系统，与大型计算机或微型计算机连接。



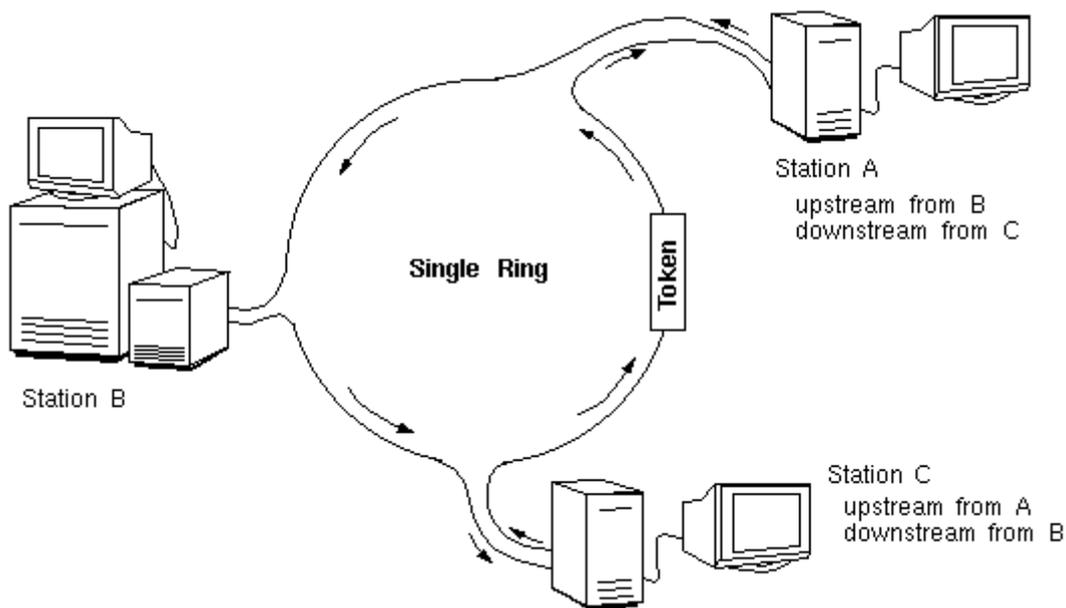
这些都是陈芝麻烂谷子了。七十年代真正的新鲜玩意儿是局域网（the local area network，LAN）。它是如何实现将所有计算机连接起来的呢？

局域网的亮点在于，它可以连接不止两个计算机系统。因此，一根简单的单线双向传输线是不够用的。理论上说，星型、环型、总线拓扑架构都可以将成千上万台计算机连成局域网。星型总线拓扑架构将每台计算机与好几个中心节点连接，这显然可以胜任这一要求。总线拓扑结构则是由唯一一根长电缆组成，计算机沿着电缆排布。而在环型拓扑结构里，一根电缆将一台台计算机串起来，最后首尾相接，形成完整的环型结构。



Pic: 拓扑结构 (Topology)

而在实际操作上，事情远没有理论简单。令牌环网 (Token Ring)，是一种使用星型拓扑结构的局域网技术。但这从线缆的连接方式是看不出来的，因为所有计算机都被连接到中央监控系统了（类似于今天以太网的交换机）。但是这些线路确实组成了一个环型结构。令牌环网使用了一个复杂的令牌传递系统，以此决定是哪一台计算机、在什么时候可以发送数据封包。在环型结构中流通着一个令牌，持有令牌的计算机就要负责发出信息。令牌总线使用了物理总线拓扑结构，但同时也有一个令牌传递方案，用来判决每台计算机对总线的访问权限。由于其复杂性，令牌网络在很多种故障类型面前都显得弱不禁风。但这种网络形式确实具有“运行效果可以预测”的优势，网络的运行效果可以事先精确地计算得知这一点，在某些应用层面具有重要意义。



Pic: 令牌环网 (Token Ring)

但在局域网标准化战争中笑到最后却是以太网。游说标准化组织中的官员，同时倡导自己智能、简洁、低实施成本的设计，是以太网的杀手锏。以太网坚持不懈地发掘、吸收更高速率的网络传输协议，并将它们的技术优势据为己用，令这场网络标准化之争日益失去悬念。几十年后，以太网终于一统江湖。

如果你曾经留意过你家电脑屁股后面突出的网线，好奇过以太网是从哪里开始、跑了多远、怎么运作的，不必再抓耳挠腮了。且听我慢慢道来。

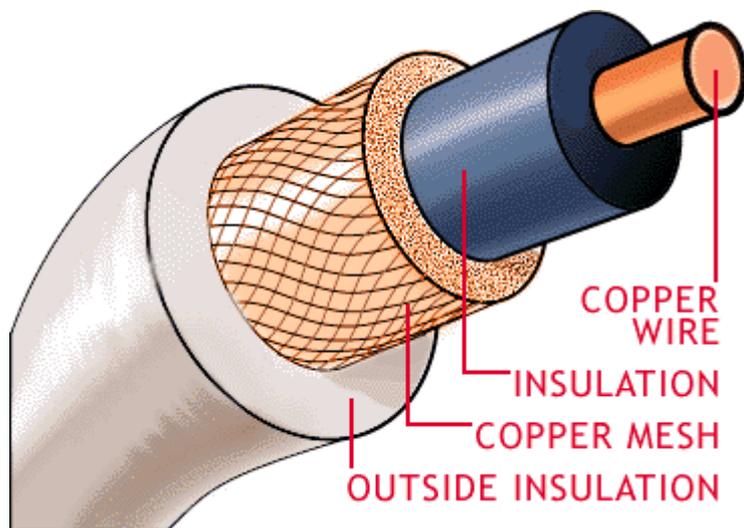


施乐 PARC 为您呈现

以太网是由鲍勃·麦特考菲 (Bob Metcalfe) 和其他施乐公司 (Xerox) 帕罗奥图研究中心 (Palo Alto Research Center, PARC) 的员工在 1970 年代中期发明的。PARC 的实验用以太网网速达 3Mbps，“这个数据传输速度比计算机访

问主记忆体要慢，但已经够用了”。由于可观的网速，数据封包不会在以太网接口被延迟。“以太网”这个名字来自“光以太”，一个曾经被认为是电磁波传递介质的虚构概念（正如声波的传递介质是空气）。

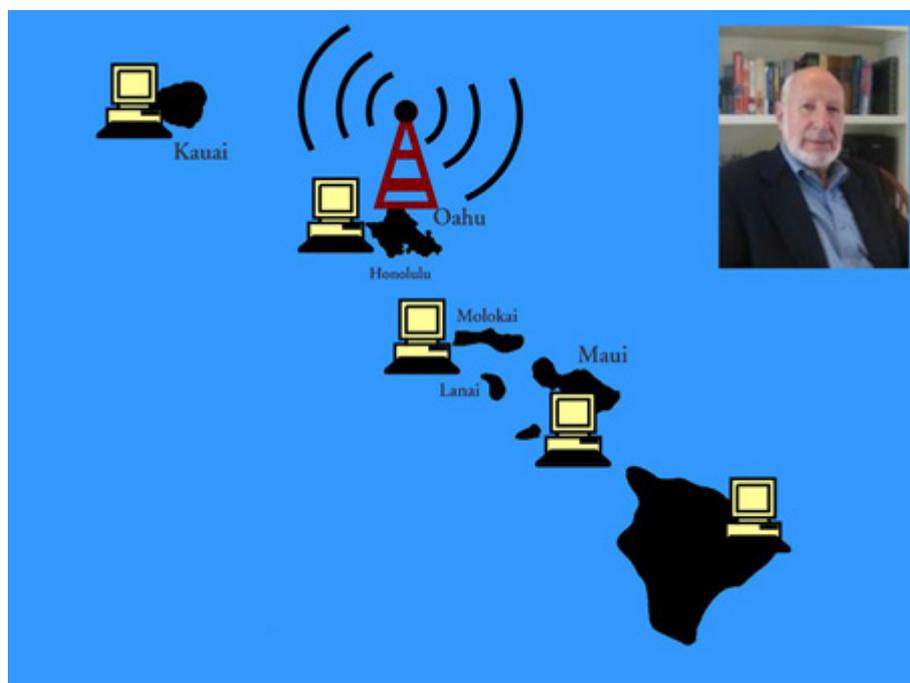
通过在粗同轴电缆中以广播方式发送数据包，以太网把网线布线当作广播用的“光以太”。计算机通过贯穿熔覆和外导体的小孔与内导体接触并相连，与以太网电缆建立起连接。同轴电缆不允许延伸旁路，其两端配备了终端电阻，可以校正电缆的电气性能，因此信号可以沿着电缆一直传输而不会被反射。所有计算机都能看到所有途经的数据包，但以太网接口会忽略不是发送到本地计算机或本地播散地址的数据包，因此软件只需处理发送至本机的数据包就可以了。



Pic: 同轴电缆 (coaxial line)

其他的局域网技术都采用了分散式的机制来分配与共享通信介质的访问，这当中当然不包括以太网。PARC 研发的分布式控制机制实在太高明了，令我想形容为“简直是精神病人操控的精神病院才能想得出来”，虽然这显然有失公允。不过我相信大型机和微机生产商们应该都会觉得我的比喻还不算太离谱。

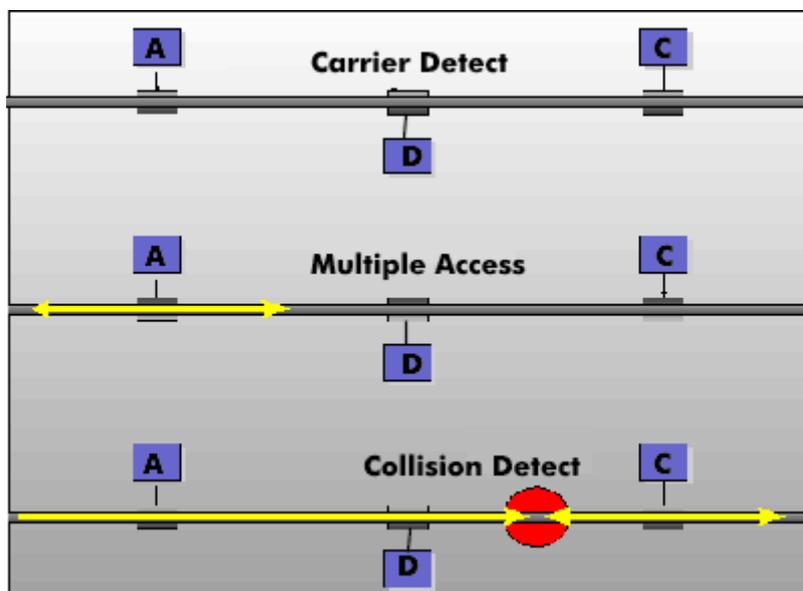
以太网的介质访问控制程序 (mediaaccess control, MAC) 叫“带有碰撞检测的载波侦听多路访问” (CSMA/CD)。这一程序是建立在 ALOHA 网络之上的。ALOHA 网络是 1970 年代早期搭建在几个夏威夷小岛之间的电台网络，几个发射站的远距离无线发射机都使用相同的频律。每个发射站都能随心所欲地传输信息。显然，当两个发射站同时发射时，会发生干涉效应，双方发射的信息都会丢失。



Pic: ALOHAnet

为了解决这个问题，网络的中央位置在正确接收到数据包后要进行一次确认。如果发送方没有收到来自中央位置的确认信息，就隔一段时间后再次发送相同的数据包。当两个发射站同时发射产生了数据碰撞时，这个重发制度能保证双方的数据最终都能发射出去。

以太网在 ALOHA 网络的基础上作了若干改进。首先是“载波感测”。以太网站先检查“以太”是否空载，如果检测到信号了就排队等候自己的发送机会。其次是“多路访问”和“碰撞检测”。在共享介质中发射数据前，以太网发射站通过比较准备发送的数据和目前正在电缆中传输的数据，以免产生干涉效应。如果两者相同，就必然会产生数据碰撞。这种情况下，发送就失败了。为了确保发送干涉信息的源头也能知晓发生了数据碰撞，发射站在检测到碰撞后，要发一个是 32 比特的倍数的拥塞信号。



Pic: CSMA/CD

现在双方都知道它们的数据发送失败了，因此它们要通过一个“指数型延迟存取程序”来重新发送数据。一方面，为了不浪费宝贵的带宽，重新发送自然是越快越好；另一方面，若双方都马上重发，必然会再次引起数据碰撞，重发就没意义了。因此，每个以太网发射站都事先通过“某个整数*传输 512 比特数据所需的时间”求得一个最大延迟时间。当一个数据封包被成功发送后，最大延迟时间就被设定为 1。数据碰撞发生时，就将最大延迟时间翻倍，直至涨到 1024。之后，以太网系统会在最大延迟时间值内，随机选择一个数作为实际延迟时间。

比方说，发生第一次数据碰撞后，最大延迟时间变成 2，可选的实际延迟时间就成了 0 和 1。显然，有 50%的可能性两个发射站都选择 0 或者都选择 1，导致另一次数据碰撞。如果这样，最大延迟时间就变成了 4，这两个发射站第三次发生数据碰撞的几率降到了 25%。如果人品太差连续发生了 16 次数据碰撞，以太网系统就认命了，乖乖放弃这个数据包。

数据碰撞对网络运行效果造成的冲击一度令人们充满恐惧、不安和怀疑。但实际上，数据碰撞很快就能被检测到，且发生碰撞的数据传输马上就中断了，所以整个过程都不会浪费太多时间。CSMA/CD 制式的以太网在不超负荷的情况下，运行效果其实是相当好的：鲍勃·麦特考菲和大卫·博格斯（David Boggs）在 1976 年发布的描述实验用 3Mbps 以太网的文件里指出，即使 256 台计算机都在不断地发送大于等于 500 字节的数据包，仍有超过 95% 的网络容量在成功地传输着数据。可见这系统实在高明。

标准化之路

直到 1970 年代末期，以太网的所有权都属于施乐。但比起控制着一块小蛋糕，施乐更愿意在一块大蛋糕里切上一块。所以施乐与数字设备（Digital）和英特尔（Intel）合作，合体后的 DIX（三公司名称缩写）大财团创造了（或者至少说是共同投资了）10Mbps 的以太网规格。之后还迅速修复了部分 bug，推出了 DIX 以太网规格的 2.0 版本。

美国电机与电子工程协会（Institute of Electrical and Electronics Engineers，IEEE）随后也加入了这场游戏。最终，为免帮某些厂商卖广告的嫌疑，IEEE 在尽力避免使用“以太网”这一名称的前提下，推出了如今被当作以太网官方标准的 802.3 标准。（除了在以太网这个名字的范畴有所出入之外，DIX2.0 和 IEEE802.3 是完全兼容的。）



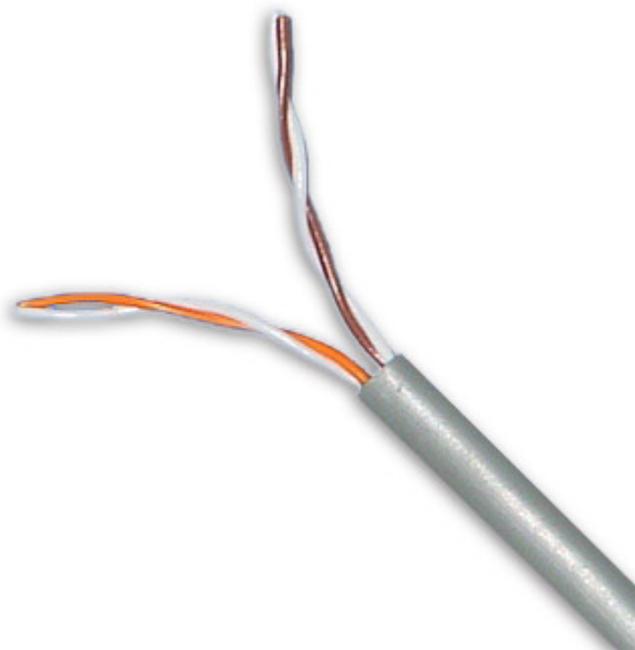
pic: IEEE

尽管实际效果在刚开始时尚算可以，但工程师们不得不承认，整栋大厦的用户只能用同一根电缆连接的效率实在有限。简单地从粗同轴电缆上产生旁路是不可能的，这样数据信号会受到不良影响。解决方案是使用中继器。它们能重复产生信号，并实现了连接两个以上的以太网电缆或以太网段。

还有另一个麻烦，足足 9.5 毫米粗的同轴电缆用起来也不甚方便。譬如说，我就好幾次见过电信公司的家伙在布线时为了令穿墙而过的电缆弯折往下走而不得不出动到大锤，还委实耗费了好一会儿功夫。此外还有人跟我说过，他喜欢将一整段笔直的同轴电缆放在车子里：“如果条子在你的车里翻到一根棒球棒，会告你私藏武器；而一根同轴电缆在干架时跟球棒一样好使，但条子从来不会因为一根电缆我麻烦。”

尽管在威吓流氓这种歪门邪道上的效果差些，但说到正儿八经的用途，细同轴电缆可比粗同轴电缆好用多了。细同轴电缆只有粗同轴电缆一半粗，跟电视天线电缆很像。若一台新的计算机想接入粗同轴电缆连接成的局域网，必须像吸血鬼一样吸附在网线的小孔上。细同轴电缆则摆脱了这种形式。细同轴电缆末端是 BNC 接头，计算机可以通过 T 型接头与之连接。但细同轴电缆构成的以太网段也有一大缺点，一旦电缆在任何位置出现受损情况，就整个局域网都无法运行了。新电脑接入网络时自然会出现这种情况。而且由于电缆必须穿过每一台计算机，全网中断的意外更是时有发生。所以必须寻求更好的网络架建方法。

1980 年代末出台了一种新的规格，这种规格将以太网架建在非屏蔽双绞线电缆上：用地球人的话来说就是架建在电话线上。以太网使用的非屏蔽绞线电缆（UTP），由四对细同轴电缆绞接而成。这种电缆可以用单一铜线或细铜线束构成（前者电气性能更好，后者更容易加工）。UTP 电缆配备了如今很常见的 RJ45 型塑料卡标式连接器。架建在 UTP 上的 10Mbps 和 100Mbps 以太网则只需用到双绞线：一根发送数据，一根接收数据。



Pic: UTP 双绞线 (UTP with two twisted pairs)

搭建这种规格的局域网多了一个比以往稍显复杂的问题，即每一根 UTP 电缆都是自身的以太网段。因此，为了在两台以上的计算机之间搭建局域网，必须使用多端口中继器（它更常见的名字是集线器）。简单来说，中继器将正常情况下收到的信号，和数据碰撞时产生的拥塞信号重复发送到所有的端口。在以太网环境里，复杂的规则制约了集线器的拓扑结构和使用范围。考虑到如今应该没人再对如何使用集线器搭建一个规模巨大的以太网感兴趣了，具体细节姑且不表。

搭建局域网麻烦不少，到现在还一直纠缠着我们。计算机用 1 号和 2 号插脚（插脚是指在计算机和其它通讯设备中信号接口的接口处分叉的接触处，插脚是阳性连接器的一部分，用来插进阴性连接器）发送信号，3 号和 6 号插脚接收信号。但对于集线器和交换机而言，情况有所不同。这句话的意思是一台计算机通过普通电缆和集线器连接，但需要连接两台计算机或两台集线器的话，就必须动用到交叉线缆，一头把 1、2 号插脚相连，另一头把 3、6 号插脚相连（反之亦然）。好玩的是，由苹果公司领导的开发联盟开发的火线接口（IEEE 1394，别名火线 FireWire 接口，一种高速传送接口）干脆只允许使用交叉线缆，结果这样一刀切反而降低了对用户造成的不便。



Pic: 火线 (FireWire)

无论如何，解决了以上种种复杂难题后，最终的硕果是一套高速而灵活的系统——这套系统快得目前仍在被使用。不过，我们需要更高的网速。

追求速度：快速以太网

尽管现在看来难以置信，但在 1980 年代初，10Mbps 的以太网已经是非常非常快的网速了。但不妨静下心来想想：有哪项超过 30 年的技术现在的电脑还在用？300 波特（每秒钟发送 300 个 bit 数据）的调制解调器？读写周期 500ns 的储存器？菊瓣字轮式打印机？唯有 10Mbps 以太网。到了今天，10Mbps 仍不算慢得完全没法用的网速，它依然是我们电脑上 10/100/1000Mbps 以太网接口的组成部分。

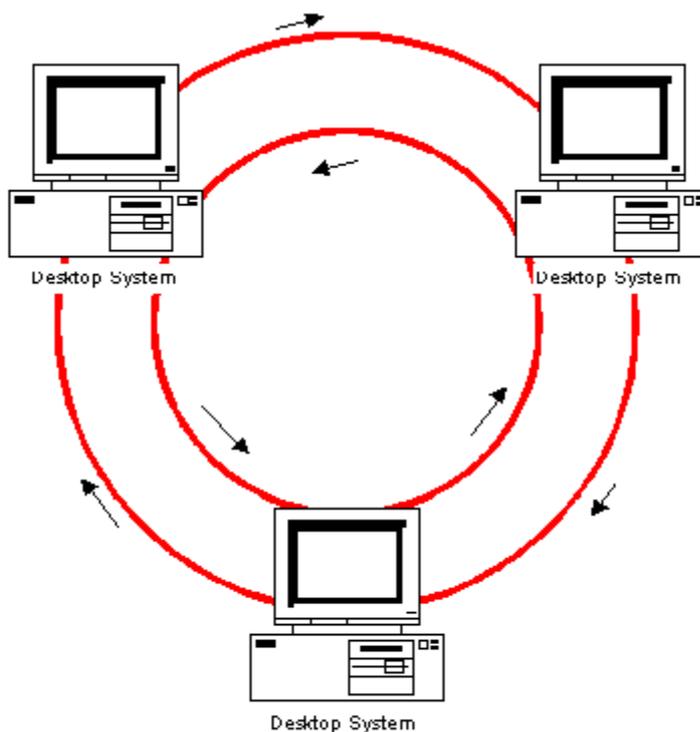
到了 1990 年代初期，以太网不再像十年前那样觉得自己快得无可匹敌了。不妨看看 1977 年迪吉多（Digital Equipment Corporation，DEC）发布的计算机 VAX-11/780。780 配有一些 2MB 的 RAM，主频为 5MHz。它几乎每秒能处理一百万条机器语言指令，1757 条 dhrystones 指令（dhrystones 是 1984 年开发的一个 CPU 性能基准程序，这一名字是对更古老的性能基准程序 Whetstone 的发挥）。目前流行的英特尔 i7 计算机可能有 3GHz 的主频和 3GB 大的 RAM，每秒能执行 1700 万条 dhrystones 指令。如果网速增长得跟处理器速度一样快，今天的 i7 至少会有 10Gbps 的网络接口，也许还会有 100Gbps 的。



Pic: VAX11-780

但事实是网速的增长并没有这么夸张。幸好，到了 1990 年代，有一种新的局域网技术，足足比常规以太网快 10 倍：光纤分布式数据接口（Fiber Distributed Data Interface, FDDI）。

FDDI 使用的是一种速度高达 100Mbps 的环型拓扑结构。当主环在某处发生问题整体崩溃时，FDDI 还有另一个支持自动线路备份的后备环。一个 FDDI 网络至少可以横跨 200 公里。所以 FDDI 可以充当非常高效的跨局域网大容量网络主干。尽管以太网和 FDDI 在很多方面不同，实现双方数据封包格式互相转化还是可以的，所以以太网和 FDDI 可以通过网桥实现连接。



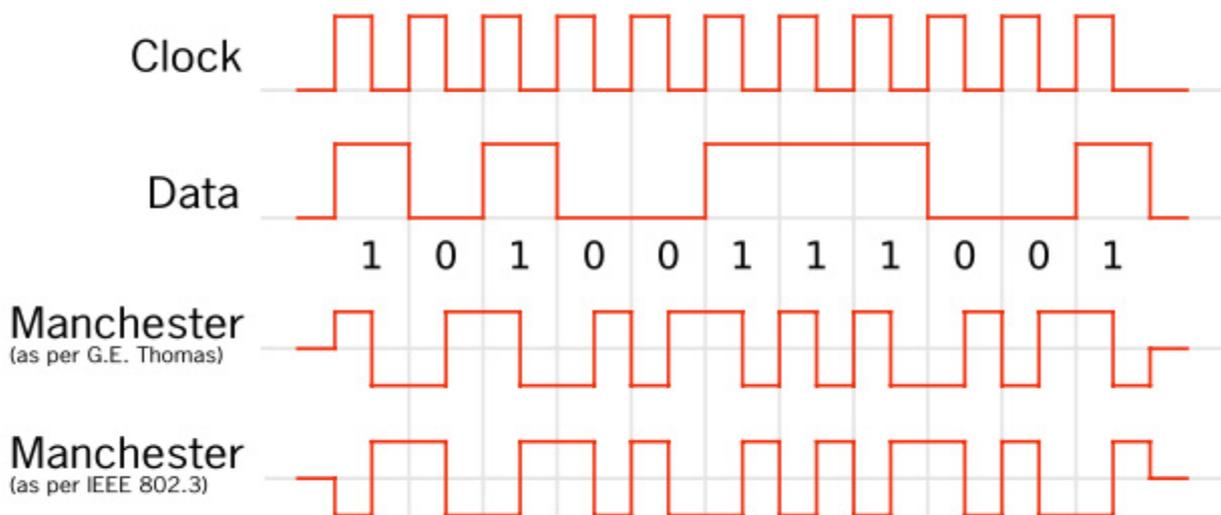
FDDI - all stations functioning

网桥连接在多个局域网段上，可以获悉哪个地址在哪个网段被使用了。之后在必要时，网桥将源网段的数据包重发送到终网段。这意味着，不同于中继器，本地的信息传输（也包括数据碰撞！）会保留在本地。所以网桥将网络分割为独立的冲突域，但所有的数据封包依然在整个网络里到处跑，所以使用了网桥的网络依然是个单一的广播域。

路由器可以将网络划分为多个广播域。路由器在整个网络模型的网络层运行，比以太网高了一级。这意味着路由器在收到数据封包时就会去除以太网的数据头信息，加上低一层的数据头信息（以太网层或者其他层的），然后发送给下一方。

FDDI 在连接以太网段和服务器时能发挥作用，但它和细同轴电缆以太网一样有“牵一发而动全身”的问题，而且成本相当昂贵。随后有人开发了 CDDI，即改用铜线的 FDDI，但这货基本上一无是处。因此 IEEE 后来又开发了快速以太网，即 100Mbps 版本的以太网。

10Mbps 以太网使用了“曼切斯特编码技术”传输数据。曼切斯特编码技术将每一比特的数据转换为电缆上一个或低或高的电压。0 被编码为“低电压-高电压”的跳变波形，而 1 则是“高电压-低电压”的跳变波形。这种做法基本上把传输的数据量翻倍了，但可以避免数据中出现长串 0 或 1 的麻烦：传输介质基本上无法长期维持低电压或高电压的状态，这样信号看起来就跟直流电压一样毫无变动。这还会引起连锁反应。计算机也会发懵：“我刚刚看到那一长串零到底有几个？93 比特还是 94 比特？”曼切斯特编码技术通过在每一比特数据之间插入一个高电压或低电压，避免了以上两种问题。无论同轴电缆还是第三等级 UTP 线缆，都足以应付这些额外所需的带宽。



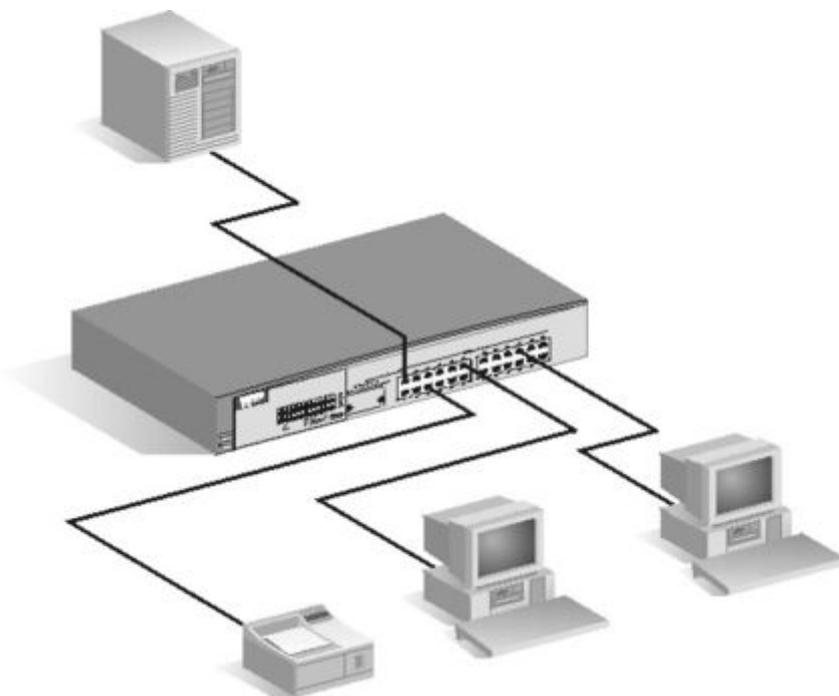
对 100Mbps 以太网而言，实现曼切斯特编码技术自然不成问题。但对 UTP 来说，以这样的速度传输数据就力不从心了。所以，100BASE-TX 以太网从 CDDI 处借鉴了一种名为 4B/5B MLT-3 的编码技术。这种技术将 4 比特的数据转化为 5 比特。这样的话，就能保证每个 5 比特的数据块里肯定有 2 个跳变波形。这种做法也产生了一些特殊信号，例如说没有数据供发送时产生的空白信号。

多阶传输-3 型编码技术则是在做“-1, 0, +1, 0”的循环。如果在一个 4B/5B 数据块里一比特的数据是 1, 那么就跳变到下一个值。如果该比特的数据是 0, 那么信号在这一比特里就维持在之前的电位。这种做法限制了数据频率的峰值, 使之与 UTP 线缆的局限性相匹配。但是, UTP 线缆必须符合要求更严格的 CAT-5 型规格, 而不仅仅是专供 10BASE-T 以太网使用的 CAT-3 型规格。除了 100BASE-TX 规格以外, 还有很多 CAT-5 型 UTP 可用的快速以太网线缆规格, 但只有 100BASE-TX 成为了大众市场的产品。

从网桥到交换机

快速以太网和普通以太网使用相同的 CDMA/CD (带有碰撞检测的载波侦听多路访问技术)。但由于线缆长度和中继器数量的限制, 数据碰撞被检测到的速度不能跟网速一样, 一下子提高 10 倍。很快, 10Mbps、100Mbps 的集线器都相继问世了。他们分别实现了 10Mbps 系统之间的连接和 100Mbps 系统之间的连接。当然, 要是能实现这两种计算机之间的互通就更好了, 所以 10Mbps 和 100Mbps 系统之间通常都会通过网桥连接。

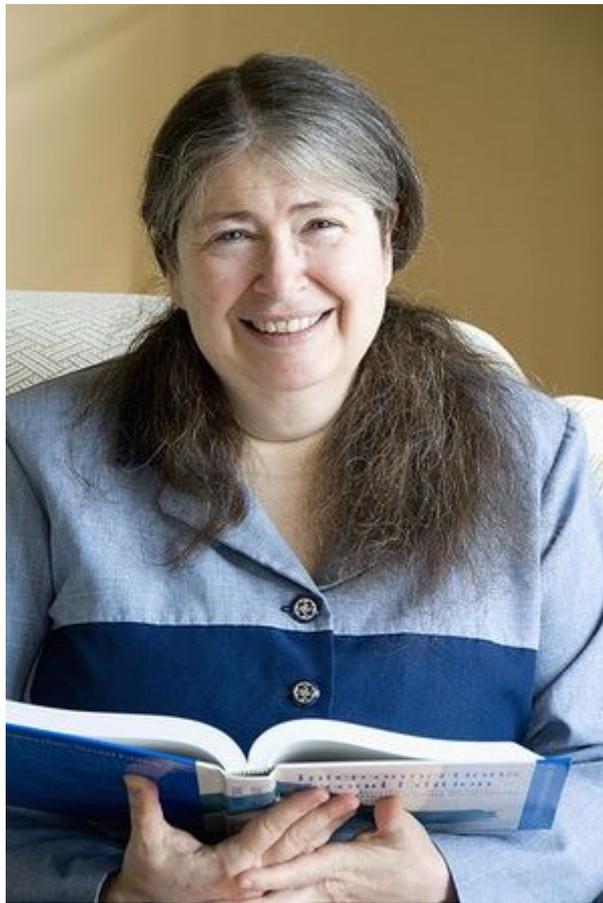
下一步要做的就是所有的网络端口之间搭建网桥了。这些多端口网桥叫交换式集线器或以太网交换机。通过交换机, 如果在端口 1 的计算机要向端口 3 的计算机发送数据、同时端口 2 计算机要向端口 4 计算机发送数据, 这种情况下不会产生数据碰撞。因为数据封包只会通过直达目标地址的端口直接传输。交换机通过观察流经自己的数据封包源地址得悉地址与端口与地址的对应关系。如果数据包没有标明目的地址, 交换机就像广播数据一样, 将之一股脑地扔到所有的端口去。



集线器和交换机有同一个局限，即以太网网络必须是无路由循环的。如果将交换机 A 的端口 1 和交换机 B 的端口 1 相连，并且将交换机 B 的端口 2 和交换机 A 的端口 2 相连，马上就会造成灾难性的后果：数据包在网络里死循环，溢出的广播数据肆虐泛滥，直至网络崩溃。所以在网络中建立后备连接时很有用的。当主连接崩溃时，网络传输依然可以在后备连接里进行。

通过制定侦测以太网中路由循环、清除死循环连接的网络传输协议，交换机的死循环问题总算是解决了。这中做法令实际生效的网络拓扑结构比较接近数学家所谓的“树”：任意两个节点之间只有少于等于一条通路。如果任意两点之间有且只有一条通路，则构成了一棵生成树。即没有网络节点是孤立的。如果有一个活动链接失效了，就再执行一次生成树协议（spanning tree protocol, STP），以生成一棵全新的生成树，保证网络继续运行。

生成树算法由拉迪亚·珀尔曼（Radia Perlman）于 1985 年在迪吉多所创。她用一首打油诗令生成树算法流芳千古：



Algorhyme

《算法小诗》

我遍寻未见

比一棵树更动人的图

拥有无路由循环连通性的树

令数据包可以遍历每个局域网的树

首先 必须选定根节点

凭 ID 选定根节点

然后 探寻源自根节点的最小代价通路

之后 在树上安排通路

我们就这样打造出一棵网络树

最后由网桥支撑起的一棵生成树

拉迪亚·珀尔曼

更上一层楼：10 千兆位以太网

快速以太网在 1995 年被标准化，但仅仅过了 3 年，以太网的下一轮迭代又来临了：10 千兆位以太网。一如既往，网速有飙升至原来的 10 倍。同样一如既往的是，为了力争好表现，新一代以太网也从借鉴了别处的技术。这次的外来和尚是光纤通道（这显然是英国血统的玩意儿）。这种技术原来主要用于储存网络。尽管被广泛地应用在不同类型和长度的纤维上，但 10 千兆位以太网主要坚守着光线通道的血统。

但对于 1000BASE-T 网络家族而言，IEEE 又要参照着对 100BASE-T2 和 100BASE-T4 的处理方式，对 100BASE-TX 这种同样不受快速以太网标准管束的新网络类型出台新一代的标准。因为一方面而言，UTP 线缆要求已经再次升级到 CAT-5e 版了，另一方面 100BASE-T 使用了四对绞线——同一时间，两个走向。

这要求进行一些类似拨号上网调制解调器所做的高级数字信号处理，只是快了大概一万倍。每对绞线通过 4D-PAM5 一次发送 2 比特的数据。4D 指四个数据符（2 比特），PAM5 表示分 5 个层级的脉冲幅度调制。传送信号的速度和快速以太网一样，达到每秒 12.5 亿个数据符。还有一个复杂的位加扰系统，专门用来确保潜在干涉效应之类的特性都已经被优化。

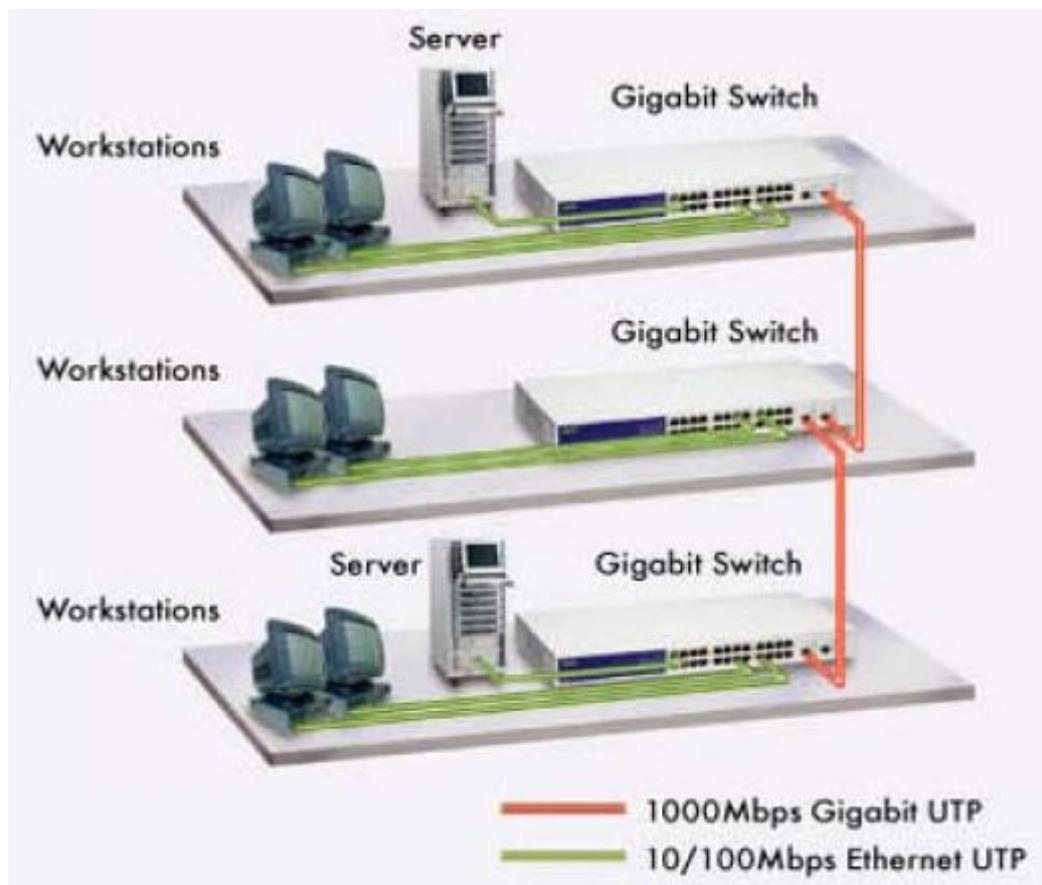
CSMA/CD 机制取决于一个特定条件：数据封包头一个比特的数据在碰撞域内到处游走时，发射站还没发出该数据封包最后一个比特的数据，所以发送方和接收方都会产生“我们在同时传送”的错觉。但随着比特率的提高，数据传输的时间大大缩短，快速以太网碰撞域的物理范围也因此缩小了。而对于 10 千兆位以太网，碰撞域更进一步缩小到了大约 20 米——小得令 CSMA/CD 机制根本没法正常运行。为了避免这种麻烦，10 千兆位以太网添加了“载波扩展”功能，通过添加冗余数据将数据封包增大到 512 字节，这样总长达 200 多米的线缆依然能使用 CSMA/CD 机制。

但据我所知，没有商家实施了上述方案：他们改为采用交换机。如果有交换机或直接连接两台计算机的电缆，CSMA/CD 技术就根本没必要了：双方同时传输数据不会产生任何问题。这种方法叫全双工操作，和传统 CSMA/CD 采用的半双工操作相对。这种衍生的 UTP 电缆支持一种“自动配置协议”，允许两个以太网系统自行决定采用全双工抑或半双工操作。

在自动配置协议广泛流传开来之前，人们有时会手工配置计算机系统，以改用全双工操作，但另一台计算机依然会使用半双工操作。在流量少的情况下，这种冲突产生的问题自然也少。但随着流量增加，操作方式之间的冲突就越来越频繁了。全双工操作的计算机系统会忽略这个问题，受损的数据封包也不会被重发。而近些年来，自动配置协议有效而可靠地为我们解决了上述问题，相信将来也没有任何理由取缔这个可靠的小助手，给我们自找麻烦。

惊人的速度：100 千兆位以太网

如今，在一栋大厦或者一间办公室里搭建局域网的常见方法是用好几台小的交换机（大概是一个配线柜里一台），所有的 UTP 电缆都从这里发出。这些小交换机再和一台更大更高速、充当局域网网络主干的交换机连接。因为不同楼层、不同服务器的计算机用户最终都要汇总到一间机房，所以即使整栋大厦的计算机不至于把 10 千兆位以太网塞爆，但交换机之间的数据传输对带宽的需求依然大得夸张。所以即使到了今天，配备 100 千兆位以太网的计算机也不常见。现在非常需要 100 千兆位以太网作为网络主干技术。相关的标准在 2002 年已经发布了。



在电信行业的世界里，有一种叫 SONET 或 SDH (Synchronous Optical Networking/ Synchronous Digital Hierarchy, 同步光纤网络和同步数字系列) 的技术，通过光纤传播传输大量的语音呼叫和数字形式的数据。SONET 的运行速度达到 155Mbps, 622Mbps, 2.488Gbps... 甚至 9.953Gbps! 这速度简直快得令人无法抗拒。所以有一种形式的 100 千兆位以太网就采用了低层的 SONET/SDH 成帧方法。这种网络叫 WAN PHY (Wide Area Network in physical layer, 物理层广域网)。但也有一种物理层局域网，速度达到 10.3125Gbps。100 千兆位以太网不再支持 CSMA/CD 半双工操作了；在这种速度下，只有全双工操作能运作。

100 千兆位以太网和大部分物理层局域网的衍生形式都使用光纤。要令 100 千兆位以太网在 UTP 电缆上运行的效果跟在光纤上运行的效果一样好，可不是件轻松活儿。对 100 千兆位以太网来说就更困难了；100 千兆位以太网在光纤上哪怕跨越相当长的距离也能运行得很好，因此 100 千兆位以太网很受网络服务提供商欢迎。但要令 100 千兆位以太网在 UTP 电缆上运行还要变一个小戏法——直到 2006 年，10GBASE-T 标准才发布。10GBASE-T 标准要求比 1000BASE-T 更好的电缆——要符合 CAT-6a 标准，碰撞域超过 100 米。CAT-6a 使用比 CAT-5e 更粗的绝缘层，所以新式电缆在老式电缆接口上未必适用。

10GBASE-T 更将快速以太网和 10 千兆位以太网每秒能处理的数据符数从 12.5 亿个提升到 80 亿个，从每个数据符能编码 2 比特提升到每个数据符能编 3.125 比特。10GBASE-T 还加强了回音消除、近端噪音消除功能，新增了另外一些与 UTP 上运行的千兆以太网共同推出的信号处理程序、前向纠错功能，用来修复偶发性的传输错误。

1000 千兆位以太网触手可及

显而易见，在 10 千兆位以太网之后，100 千兆位以太网就是新一代接班人了。然而，要在光纤中以 100Gbps 的速度传输数据委实困难重重，因为在光纤中传递信息时，激光脉冲会变得非常短，所以激光脉冲在传输过程中很难保持波形。因此，IEEE 作了小小的让步，将 40 千兆位以太网作为保留选项，而不是坚持每次升级网速一定要升 10 倍的传统。

目前，已经有大量 100GBASE-*系列的标准了，但它们当中相当部分，要么为了达到 40 或 100Gbps 的速度而必须使用四条并行数据路径，要么只能在短距离之间使用。创建大一统 100GBASE 标准的目标依然前路漫漫。

以太网的未来

令人震惊的是，以太网在问世后居然成功地存活了超过 30 年，速度产生了超过 4 个数量级的狂飙突进。这表示，在元祖级 10Mbps 以太网传输一个比特所花的时间里，100 千兆位以太网可以传输一整个数据封包了（假设它有 1212 字节那么长）。在过去的 30 年里，以太网的所有方面都发生了改变：包括 MAC 程序，比特率编码技术，布线方式……讽刺的是，唯有数据封包的格式固步自封。这可是由于一开始 IEEE 偏心与众不同的 DIX 2.0 标准，而故意忽略、不多做规定的部分啊。

这种向后兼容实际上却是个大问题：在 10Mbps 的网速下，你每秒大概可以发送 14,000 个 46 字节的，或者 830 个 1500 字节的数据封包。但即使在千兆位级的网速下，数据封包最大也只能达到 1500 字节，这就成问题了。实际上现在很多千兆位以太网网卡的传输控制协议/网际协议堆栈（TCP/IP stack）都可以发送或接收大得多的数据包，将之分割或合并后再传送给 CPU，这样可以令 CPU 的日子好过些。因为 CPU 处理数据时是以数据封包数为单位的，而不考虑数据封包的大小。而想以 100 千兆级速度每秒传输多达 14 亿个的 46 字节大小数据封包，根本就是存心搞笑。不幸的是，如果允许传输更大的数据包的话，又会破坏了新旧系统之间的兼容性。所以如今 IEEE 一直在努力作出改变。

如果只扮演互联网这条高速公路的匝道的角色的话，现在局域网可谓无处不在。以太网凭借着它的千姿百态取得了举世赞叹的成功，同时排挤着与之竞争的其他局域网技术。在过去十年里令以太网发展速度放缓的唯一原因，是因为无线局域网（Wi-Fi 格式）实在是太方便了。（而且 Wi-Fi 和有线以太网之间的兼容

性很好。)但有线和无线网络功能很大程度上是互补的,所以即使越来越多计算机不会动用到以太网端口,甚至根本就没有配备这么一个端口,以太网仍时刻准备着为人们提供高速而可靠的传输服务。这些特性一直是共享无线介质梦寐以求的。

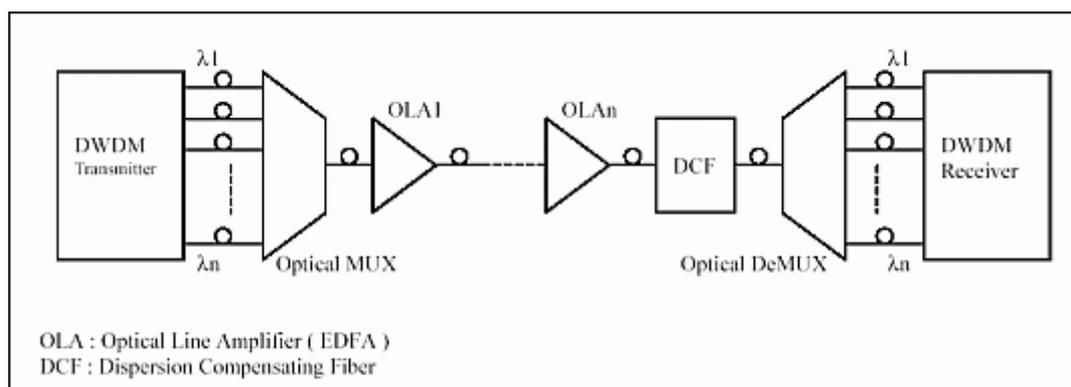


万千兆位以太网?

未来还会出现速度高达 1000Gbps 的一万千兆位以太网吗?从一方面看,在光纤里以 100Gbps 的速度传输已经难度很大了,所以这实在很悬;不过话又说回来,若时光退回 1975 年,也没几个人能想象得到如今的学生能带着配备 10Gbps 端口的笔记本去上课吧?



CPU 设计师们已经通过多核并行技术解决了类似的问题。10 千兆位以太网在 UTP 电缆中使用了四绞线，这已经体现出“并行”的概念了。很多以光纤为介质、速度达 40Gbps 和 100Gbps 的衍生类型以太网，也已经通过采用不同波长激光来实现了利用并行数据流传输。海底电缆更是已经通过使用密集波分复用 (dense wavelength division multiplexing, DWDM) 实现了在一根光纤电缆里有一万千兆位总带宽。所以纵观时局，现在又是以太网将现有技术据为己用、取长补短，积极降低成本的好机会。



或者连这样的步骤都不必了。当我通过电邮向拉迪亚·珀尔曼申请在本文中引用她的《算法小诗》时，她提及一种被称为多链接半透明互联

(Transparent Interconnection of Lots of Links, TRILL) 的新技术。这种技术实现了用“多链接”而不仅仅是一个快速链接来搭建灵活、高速的以太网网络。看来无论如何，高速以太网的未来都必然要糅合进“并行”的概念了。



我实在是迫不及待地想知道，未来 30 年的科技发展，还能为以太网带来怎样的变化？