第 5 章　底层通信网络与应用

物理层完成基于信号的比特传输，着眼于信号处理与传输；数据链路层实现了数据通信，着眼于带有校验和流量控制的块数据处理，实现了数据通信的基本功能。至此，基于物理层、数据链路层就可实现基本的数据通信，组成数据通信网络，即我们常用的局域网、广域网、城域网等，这些网是当年网络研发的基本网络产品，也是当今承载 Internet 互联的底层主体，或者称作 Internet 的底层网络。历史上，局域网、广域网、城域网等产品有很多，目前社会上广泛应用的是以太网和无线局域网。

本章以局域网诞生时代的各种问题追寻为主线，对以太网的原理和技术进行剖析，揭示发明人精准地解决难题的智慧。技术创新也直接推动以太网演变升级，在产品市场上呈现出强大的生命力。在内容呈现上，力图让读者置身于工匠环境思考问题，亲历问题，体验创新，体会创新的价值。底层网络路线图如图 5-1 所示。

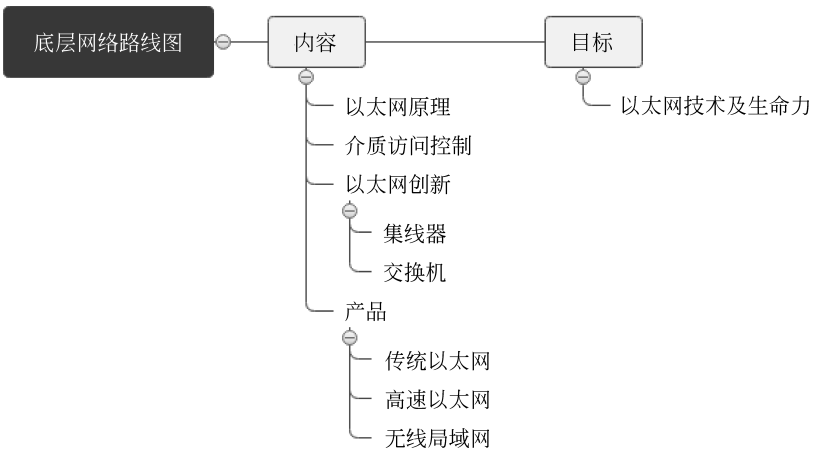


图 5-1　底层网络路线图

5.1　以太网的构成及原理

5.1.1　以太网及其构成

以太网的研发主要解决本地计算机间的数据通信问题。以太网是一种基带总线局域网，用无源电缆作为总线传送数据帧，以最简单的方式实现本地计算机的可靠互联。

当初提出以太网的方案是基于下面的思路：要寻找很简单的方法把一些相距不太远的计算机互相连接起来，使它们可以很方便和很可靠地进行较高速率的数据通信。

最早的以太网采用总线结构，将许多计算机都连接到一根总线上构成网络，如图 5-2 所示。当初认为这种连接方法既简单又可靠，因为在那个时代普遍认为："有源器件不可靠，而无源的电缆线才是最可靠的。"

计算机是通过网卡接到总线上的。网卡又称作网络适配器(适配器的概念更一般化，我们称之为网卡，以适应大众习惯)，现在大多数计算机主板上都集成有以太网网卡。网卡是计算机与网络的接口设备，负责网络主机数据的收发，实现以太网协议。网卡功能结构图如图 5-3 所示。由图可见，网卡实现了网络体系中的数据链路层和物理层。

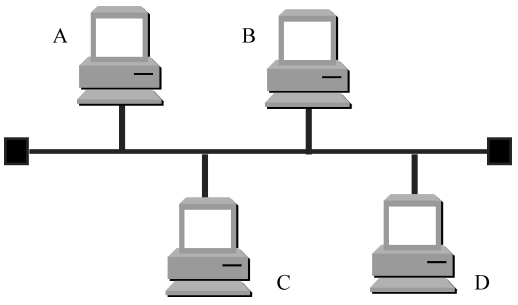


图 5-2　总线型以太网

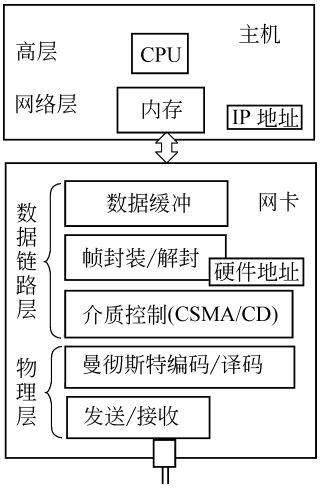


图 5-3　网卡功能结构图

网卡接收和发送各种帧时不使用计算机的 CPU，这时 CPU 可以处理其他任务。当网卡收到有差错的帧时，就把这个帧丢弃，而不必通知计算机。网络层及高层协议是在主机中运行的。当网卡收到正确的帧时，它就使用中断通知该计算机并交付给协议栈中的网络层。当计算机要发送数据时，就由网络层把 IP 数据报向下交给网卡，组装成帧后发送到局域网。计算机的硬件地址就存储在网卡的 ROM 中。

网卡到局域网的通信是通过电缆或双绞线以串行传输方式进行的，而网卡和计算机之间的通信则是通过计算机主板上的 I/O 总线以并行传输方式进行的。因此，网卡的一个重要功能就是进行数据串行传输和并行传输的转换。由于网络上的数据率和计算机总线上的数据率并不相同，因此在网卡中装有对数据进行缓存的存储芯片。

5.1.2　以太网的帧传输方式

人们常把局域网上的计算机称为"主机""工作站""站点"或"站"。

总线型结构的以太网上，当一台计算机发送数据时，总线上的所有计算机都能检测到这个数据。这种方式就是广播通信方式。但我们并不总是要在局域网上进行一对多的广播通信。发给某主机的数据帧，其他主机没有必要接收。为了在总线上实现一对一的通信，每台计算机网卡上都拥有一个代表本主机的地址，该地址具有全球唯一性。发送数据

帧时，须在帧的头部写明接收站的地址。现在的电子技术可以很容易地做到：仅当数据帧中的目的地址与本网卡 ROM 中的硬件地址一致时，网卡才能接收这个帧。不是发送给自己的帧，网卡不予接收。这样，具有广播特性的总线上就实现了一对一的通信。

考虑到是近距离通信，以太网没有采用可靠传输，帧传输控制简单，主要特点为：

第一，采用较为灵活的无连接的工作方式，即不必先建立连接就可以直接发送数据。

第二，网卡发送数据时不给帧编号，也不要求对方确认，但有帧的校验机制。这样做的理由是局域网信道的质量很好，产生差错的概率很小。

因此，以太网提供的服务是不可靠的交付，即尽最大努力地交付。目的站收数据帧时，进行差错检测，一旦检测到有差错的帧就丢弃，其他什么也不做。在因特网上，出错的帧是否需要重传由高层决定。如果高层使用 TCP，那么 TCP 就会发现丢失了一些数据。经过一定的时间后，TCP 就把这些数据传递给以太网进行重传。但以太网并不知道这是重传帧，而是将其当作新的数据帧发送。

初期，以太网发送数据都使用曼彻斯特(Manchester)编码。但是，曼彻斯特编码有缺点，不能有效利用带宽提高数据速率，其频带宽度比原始的基带信号增加了一倍(因为每秒传送的码元数加倍了)，所以在以后的高速以太网中改用其他形式的编码。

综上所述，以太网是无连接、无确认的不可靠交付。以太网上，各站点共享总线，以广播方式发送数据帧，以地址区分用户，对帧进行选择性接收。因此，没有差错控制、流量控制等可靠机制，传输控制非常简单。在以太网上需要重点解决的是信道争用问题，如图 5-4 所示。

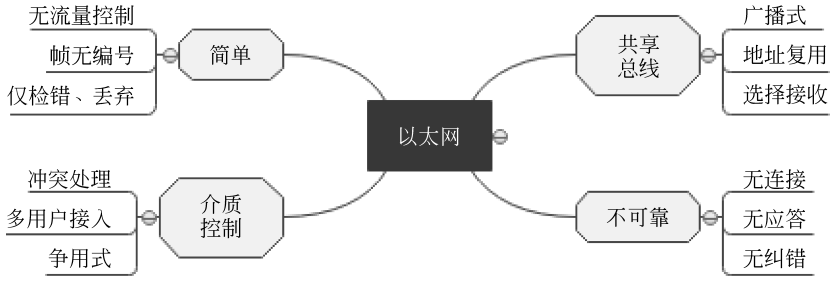


图 5-4　以太网的特点示意图

5.2　以太网的数据链路技术

5.2.1　介质访问控制协议 CSMA/CD

为何介质访问控制构成了以太网的核心问题？

以太网介质访问控制协议是以太网工作的基础，决定了以太网的特性，继而决定了以太网的应用。介质访问控制技术集中体现了以太网研发的创新智慧，挖掘分析、深刻理解其原理技术对创新应用至关重要。下面探讨和体验以太网中要面对的一系列问题及解决

思路。

以太网的总线型结构如图 5-2 所示，各站点共享传输介质，同时发送数据就会在总线上发生信号冲突，导致主机不能正确接收数据帧。这就要求各站点不能同时发送数据，必须互斥地使用信道。因此，必须使用一个策略，为站点合理分配信道的使用权，对介质访问进行控制。

1. 介质访问控制协议及原理

以太网采用的介质访问控制策略是载波监听多路访问/冲突检测，即 CSMA/CD (Carrier Sense Multiple Access with Collision Detection)协议。也就是说，每个站点在发送数据时先运行 CSMA/CD 算法，按其规则申请总线，得到总线的控制权才能发送数据。

(1) 策略分布在各主机上，由各主机自觉实施，发送数据帧前执行，属于“自主遵守规则，保证系统顺畅”。

(2) CSMA/CD 协议的规则，可简单地概括为“先听后发，边发边听，冲突立止，延时重发”。

“先听后发”的意思是“发送前先监听”，即每一个站在发送数据之前先要检测一下总线上是否有其他站在发送数据，如果有，则暂时不发送数据，要等待信道变为空闲时再发送。如果信道空闲，则立即发送。其实，总线上并没有什么“载波”，“载波监听”就是用电子技术检测总线上有没有其他计算机发送的数据信号。

“边发边听”即“边发送边监听”，监听的目的是“冲突检测”，是一个冲突发现机制。冲突后，总线上传输的信号产生了严重的失真，接收站点无法从中恢复出有用的信息。因此，一旦发现冲突，网卡就要立即停止发送，让出总线，然后等待一段随机时间后重发，这就是“冲突立止，延时重发”的含义。

“先听后发”机制保证了拿到总线使用权的站点可以完整地发送完数据帧，站点一旦确定无疑地抓住了信道，冲突就不会再发生。发送过程中不会再有其他站点的冲撞，这降低了系统的冲突概率。

(3) CSMA/CD 策略意图。CSMA/CD 是分布式地控制，就像开车上路，忙就等，闲就行，既不碰撞，又不空路。

2. CSMA/CD 机制分析

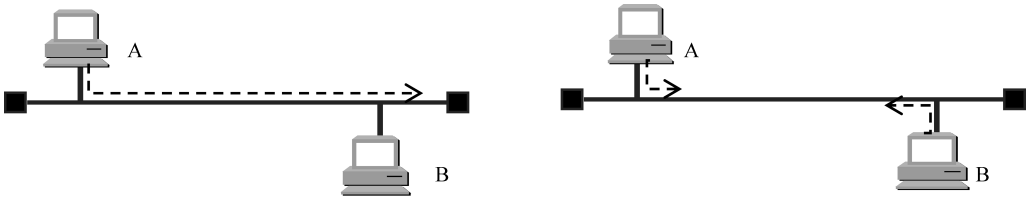
CSMA/CD 的前身是 CSMA，是由其改进而来，冲突检测的意义在哪里？

冲突检测的目的是发现冲突，以便站点发现冲突之后立即停止帧的后续数据的发送，让出信道，以提高信道利用率；而 CSMA 机制下，即使冲突了，也不知道，会把帧发完。

“先听后发”还会冲突吗？哪种情况下会冲突？

“先听后发”，站点听到信道忙就不会再去打扰，如图 5-5(a)所示；如果两个站点都先后“同时”听到信道为“空闲”而“同时”发数据帧，导致冲突出现，如图 5-5(b)所示。这种同时发送导致冲突，看起来是小概率事件，但当网络上主机很多，很繁忙时，可能就小事变大事，成为大概率事件了。

上述情形出现是因为电信号在总线上传输是有时延的。传输时延的存在意味着站点



(a) A 的信号能被 B 听到，B 则等待　　(b) 同时听到空闲，则 A、B 都发送数据帧

图 5-5　导致冲突的情形

发出的信号不能立即传遍总线，此时，总线另一端的站点没检测到信号，误认为信道空闲而发送数据，从而造成冲突。当然，由于电磁波在电缆上的传输速率接近光速（约230 000km/s），因此在局域网上的这个传输时延很小。但这段时间内的冲突概率不一定小，因为网络繁忙时等待发数据的站点可能很多，所以还要“边发边听”地检测冲突。

冲突有什么现象？站点怎么检测冲突？

冲突检测的原理是：当几个站同时向总线上发送数据时，总线上的信号就会叠加，信号电压变化幅度将会增大。当网卡检测到的信号电压变化幅度超过一定的门限值时，就说明有信号叠加，表明产生了冲突。

注意，检测是在网卡处进行的，即网卡边发送数据、边检测总线上的信号电压的变化情况，以便判断自己在发送数据时其他站是否也在发送数据。

如前所述，冲突检测是有意义的，但冲突检测也是有开销的。

主机发送后多久才会知道有冲突？

有最晚的冲突吗？检测到最晚的冲突需要多久？

在图 5-6 中，信号叠加（冲突）最早出现在总线中间的某个位置，但此时主机并不知道发生冲突。因为发送者进行冲突检测时，监听的是本网卡处的信号，所以只有当造成冲突的另一路信号到达检测者的网卡时，才会检测出冲突，这是需要时间的。从主机发送数据开始到主机检测到冲突的这段时间不妨记为 T_{cd}，$T_{cd}=2L_{ac}/C$，其中 C 为信号在总线上的传播速率，如图 5-6 所示。显然，对不同的冲突，T_{cd} 大小是不同的，对主机 A 来说，冲突发生得越晚，检测到冲突也会越晚，有没有一个最晚的冲突呢？检测到最晚的冲突需要多久？

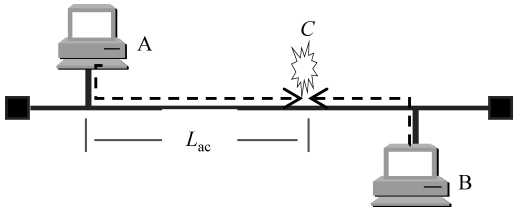


图 5-6　信号冲突示意图

因为是局域网，总线长度不限，所以答案是肯定的。应该是在总线上最远处、最晚发生的冲突将最晚检测出来。因此，如图 5-7 所示的情形是合理的。

设图 5-7 中的站 A 和站 B 是局域网两端相距最远的两个站点，A、B 间总线长度为 L，单程端到端传播时延记为 τ，则 $\tau=L/C$（C 为电磁波在电缆中的传输速率）。

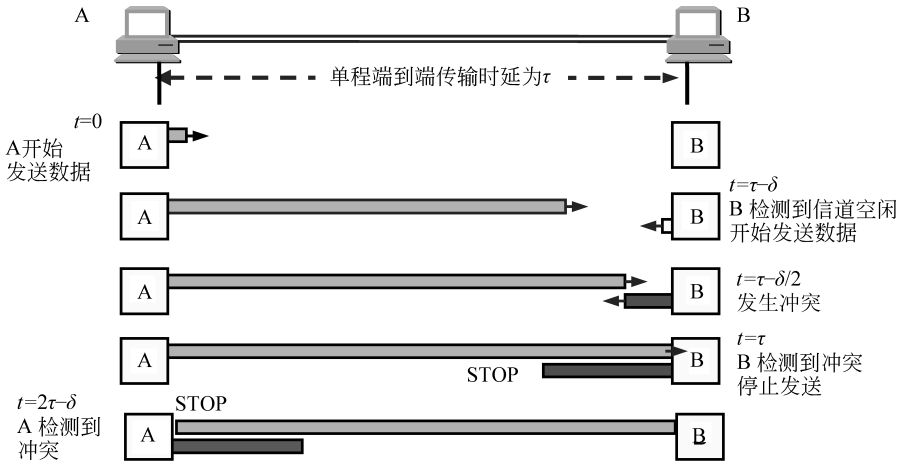


图 5-7　冲突检测示例

假设 A、B 要相继发送数据。

在 $t=0$ 时，A 检测到信道空闲，开始发送数据。

在 $t=\tau-\sigma$ 时（这里，σ 远小于 τ，趋于 0），A 发送的信号即将到达 B 但还没有到达 B 时，B 检测到信道为空闲，因此 B 开始发送数据。

经过时间 $\sigma/2$ 后，即在 $t=\tau-\sigma/2$ 时，A 的信号和 B 的信号在总线上相遇，发生冲突。但这时 A 和 B 都不能检测到此冲突，尚不知道发生了冲突。

在 $t=\tau$ 时，A 的信号到达主机 B，B 检测到了冲突，于是停止发送数据。此时 A 仍然检测不到这个冲突。

在 $t=2\tau-\sigma$ 时，B 发出的信号到达了 A 主机，A 检测到了冲突，因而也停止发送数据。

当冲突发生处离 B 很近，σ 趋于 0 时，A 主机发现冲突的时间即为 2τ。由假设知，这是最晚到达 A 的冲突。所以，A 主机从发送数据开始，经历 2τ 的检测时间就会检测到所有的冲突。也就是说，A 主机如果在 2τ 的时间内没有检测到冲突，以后也就不会有冲突发生了。这意味着，A 主机的这次发送成功了，也意味着站点发数据时要边发送边监听，这种小心翼翼的过程是有限的，不需要全程监听，从而减少开销。从另一方面理解，因为一旦信号传遍总线，其他站点就会监听到总线忙而放弃发送，导致冲突的时间阈是有限的，可以找到一个最晚的冲突。

从图 5-7 可看出，最先发送数据帧的 A 站在发送数据帧后经过时间 2τ 检测不到冲突，就说明 A 站真正抢到了信道，此帧能成功发送。否则，在时间 2τ 内 A 随时都可能因检测到冲突而放弃信道。也就是说，在 2τ 时间内还不能说 A 拿到了总线使用权，只有通过争用期的“考验”才能肯定争得了信道。因此，以太网的端到端往返时间 2τ 称为争用期（contention period），它是一个很重要的参数，又称为冲突窗口（collision window），因为

冲突都发生在这个时段。信道空闲后，可能要经过多个这样的冲突窗口，才能被某站点争到，完成一个帧的传输。

研究最晚冲突的意义在于发现了数据发送过程中的有限检测，而不必全程检测。冲突的有限检测减少了系统开销，但也导致后来以太网的诸多局限性。

3. 以太网的局限性

以太网为什么是局域的？

以太网帧长有什么限制？

以太网把争用期定为 51.2μs。这个参数的确定，实际上也同时确定了以太网的最大跨度（即总线的最大长度）。换句话说，以太网的最大跨度是有限的。

为了简化问题，我们不考虑中继器的转发时延，取 $\tau = L/C$，τ 为 25.6μs，所以总线长度 $L = C\tau$，其值是确定的。也就是说，如果两台计算机间的距离超过这个值，将导致这个最晚的冲突不能在 2τ 时间内被检测到，从而使冲突检测失效，局域网不能正常工作。另外，这个单程时延 τ 还包括中继器的转发时延，因此，实际应用以太网的总线长度比这个计算值要小很多。

思考：故意加长总线，接入 PC，通信会怎么样？

看似一马平川，实则陷阱重重。

其实，导致冲突检测失效的情况还不止这些，如图 5-8 所示。当 B 的信号到达 A 时，A 早已完成了帧的发送，从而在 A 处的总线上不会出现两路信号叠加的情况，这也导致 A 检测不到冲突，冲突检测失效，局域网不能正常工作。

图 5-8　短帧导致冲突检测失效

这个问题也必须解决，不能让这种情况出现。导致这种情况出现的根本原因是 A 主机要发送的数据太少，没能坚持到 B 的信号到来，即没能坚持 2τ 的时间。坚持 2τ 的时间需要有多少数据呢？不难计算，对于 10Mb/s 的以太网，在 2τ（即 51.2μs）的时间，发送了 512b 的数据，即 64B。综上我们规定，以太网的帧最小不能低于 64B，也就是说，以太网有一个最小帧的要求。或者说，在以太网上，小于 64B 的帧一定是坏帧。

以太网上有小于 64B 的帧吗？

事实上，在竞争窗口中因冲突停止导致以太网上有大量小于 64B 的帧片段，这些都是坏帧，习惯上称为碎片。

可见，冲突的处理是以太网中核心的问题，饱含学问。冲突的问题处理好了，剩下算法中的最后一个问题——延时重发。

为何延时重发，立即重发不行吗？

首先，因冲突停止了发送，但主机交给网卡的数据发送任务并没有完成，所以，网卡需要试图重发数据帧，以完成主机的托付。

为什么不立即重发呢？发生冲突了，说明总线上要发数据的站点不止一个，如果都立即重发，肯定会再次冲突。所以，延时重发是必要的。

怎么延时，如果各站点都延时相同的间隔，会有效吗？读者自明。延时的目的是各站点在冲突后相互避让，避开冲突，至于如何避让，还有很多学问需要讨论。

4. 冲突避让算法

发生冲突后，站点停止发送，并试图重发。假如规定各主机都立即重发，肯定会发生第二次冲突、第三次冲突……，永远不止，使系统死锁。所以，在重发时为了相互避让，要延迟一定的时间间隔再重发，能不能延时相同的时间呢？显然不行，只有延迟不同的时间，才能实现相互避让。

以太网使用截断二进制指数退避(truncated binary exponential backoff)算法避让冲突。在这种算法中，以争用期 2τ 为退避时间间隔，各站点的退避时间 T 是 2τ 的整数倍，即 $T=n(2\tau)$。如果 n 不同，各站点的退避时间 T 就不同。n 的取值与冲突的次数有关，设冲突的次数为 k，n 就在整数集合[0,1,2,…,(2^k-1)]中随机取值。例如，n 取到 3，则站点就要延时 6τ 后重发。例如，对于一个主机来说，如果是

第一次冲突。$k=1$，$2^k=2$，n 从[0,1]中随机取一个值。重发可能会出现第二次冲突。

第二次冲突后。$k=2$，$2^k=4$，n 从[0,1,2,3]中随机取一个值。

第三次冲突后。$k=3$，$2^k=8$，n 从[0,1,2,3,4,5,6,7]中随机取一个值。

……

显然，冲突的次数越多，各重发站延时取值越分散，越容易相互避让开，因而减小发生冲突的概率，有利于整个系统的稳定。考虑到冲突的次数超过 10 次以后，取值范围已经足够大，故让 k 值都取 10 并不再增大(截断指数)。工作站如果重发 16 次还不能成功，就要放弃发送。

明白了算法原理，请思考算法中 n 取值的学问。

n 的取值范围是大点好，还是小点好？

这需要主机考虑群体特性进行抉择。退避算法是在各主机上执行的，遇到冲突，以太网上有多少个站点等待发送，参与信道竞争，主机是不知道、不明确的。n 的取值范围越大，各站点重发的时机越分散，避让效果越好，尤其是在站点多、网络忙的时候；但在网络负载小、较清闲时可能出现总线空闲，站点还迟迟傻等的状况，系统效率低。

主机如何知道此时网络忙、闲，合理选取 n 的取值范围呢？

退避算法中采用了一种模糊处理方式，根据冲突次数逐步增加 n 的取值范围，兼顾了退让和效率，实现双赢。这让我们再次体会到研发人员在技术创新中的心思和处理问题的智慧。

至此，如果认为算法已经非常圆满，那么请看算法中的最后一条：工作站如果重发 16 次还不能成功，就要放弃发送。

请问，为什么要放弃？

CSMA/CD算法看似简单，但一路走来，倍感陷阱重重。研发者缜密的思维、智慧的处理让我们的思想生花，对当初的研发创新由衷敬佩。

根据以上所讨论的，可以把CSMA/CD协议的要点归纳如下。

(1) 网卡从网络层获得一个分组，加上以太网的头部和尾部，组成以太网帧，放入网卡的缓存中，准备发送。

(2) 若网卡检测到信道空闲(即在96比特时间内没有检测到信道上有信号)，就立即发送这个帧。若检测到信道忙，则继续检测并等待信道转为空闲(加上96比特时间)，然后发送这个帧。

(3) 在发送过程中继续检测信道，若一直未检测到冲突，就顺利把这个帧成功发送完毕。若检测到冲突，则中止数据的发送，并发送人为干扰信号。

(4) 在中止发送后，网卡就执行指数退避算法，等待 n 倍51.2μs后，返回到步骤(2)。

以太网还规定，帧间最小间隔为9.6μs。这样做是为了使刚刚收到数据帧的站来得及清理接收缓存，做好接收下一帧的准备。

5. CSMA/CD以太网的特性

由CSMA/CD算法可见，以太网站点在总线使用权分配上是一种竞争机制，哪个站点能够获得总线使用权限是随机的。以太网具有随机性。

每一个站在自己发送数据之后的一小段时间内，存在遭遇冲突的可能性。并且在以后的冲突重发中，也不能保证什么时候会得到信道完成发送，甚至最后因重发不成功而不得不放弃发送。也就是说，在以太网中，各主机共同竞争信道，想发送数据的主机什么时候获得总线的使用权，完成数据发送具有不可预期性、不确定性。这一特点称为以太网发送的不确定性。

以太网的随机性和不确定性决定了以太网应用的局限性，尤其不适合实时性要求高的网络控制传输。但现实社会中，以太网的确太普遍、太廉价，以至于后期有大量的人员研究如何对以太网进行改造，以便能用于网络控制领域。

同时可见，以太网是以半双工、无连接的方式提供不可靠的传输服务。在CSMA/CD控制下，每个时刻总线上只能有一路传输，所以，发送站点不能同时接收数据。后来，在IEEE 802.3x标准中又定义了全双工以太网。全双工以太网能够同时发送和接收数据，可以提供两倍带宽；它不再使用CSMA/CD，网络长度不受冲突域的限制；使用和半双工以太网一样的帧格式。全双工以太网使用交换机通过点对点链路连接计算机，传输链路、网络接口和交换机也必须支持全双工模式。

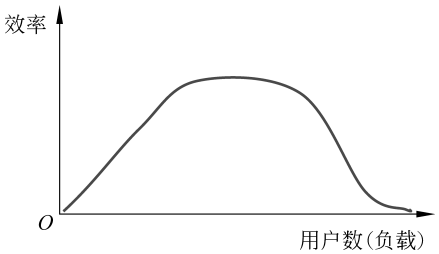


图5-9 以太网的效率

冲突的存在，决定了以太网上站点数量的上限。随着以太网上发送数据的用户增多，以太网的工作效率是变化的，如图5-9所示。为了克服这些缺点，以太网后期的研发、改进从未间断，一系列的创新持续赋予以太网强大的生命力。本

章后续将探讨网桥等系列设备的研发是如何持续改进以太网性能的。

扩展阅读：CSMA/CD 算法探源

CSMA/CD 技术的产生不是一蹴而就的，它最早来源于 Aloha，为了提高效率，后来对 Aloha 进行改进形成 CSMA 算法，以太网研发者加上冲突检测之后用于以太网总线控制。Aloha 是美国夏威夷大学 20 世纪 70 年代初研制成功的一种使用无线广播技术的分组交换网络，目的是解决夏威夷群岛之间的通信问题；Aloha 是最早、最基本的无线数据通信协议，其控制算法比较简单、原始，站点“想说就说”，繁忙时冲突多；CSMA 的主要思想是“先听后说”，以减少相互干扰；CSMA/CD 的高明之处是加入了冲突检测机制，目的是“冲突立止”。请查找和搜索 Aloha 无线分组网，了解更多关于 Aloha、CSMA 算法的细节，并结合 CSMA/CD 原理分析解决以下问题。

（1）Aloha 无线分组网的背景应用，Aloha 控制算法。

（2）了解 CSMA 种类及算法，体会它的改进是如何利于效率提升的。

（3）冲突检测的意义是什么？从 Aloha 到 CSMA/CD 是如何一步步进行技术改进，提高网络效率的？

（4）体会技术进步的过程、规律，体验技术创新中的智慧与思维。

5.2.2 以太网物理地址

CSMA/CD 解决了介质访问控制问题，以太网的传输还要解决用户标识和帧构建等问题。

以太网中有众多站点用户，要区分和寻址用户，首先要标识用户。以太网中用 MAC 层的硬件地址标识工作站，同时在以太网帧中指明接收站。

以太网中使用的硬件地址是 6B 的网卡编号，具有全球唯一性。它在网卡生产时固化在网卡的 ROM 里，又称为物理地址，因为在 MAC 帧中也用这种地址，故也称为 MAC 地址。确切地说，每个网卡都有一个局域网地址。如果连接在局域网上的主机或路由器安装有多个网卡适配器，那么这样的主机或路由器就有多个“地址”，这种“地址”是网络接口的标识符。

现在 IEEE 的注册管理机构（Registration Authority，RA）是局域网全球地址的法定管理机构，它负责分配地址字段的 6B 中的前三个字节（即高 24 位）。世界上凡要生产局域网网卡的厂家，都必须向 IEEE 购买由这三个字节构成的这个号（即地址块），这个号的正式名称是组织唯一标识符（Organizationally Unique Identifier，OUI），通常叫作公司标识符（companyid）。例如，3Com 公司生产的网卡的 MAC 地址的前三个字节是 02-60-8C。地址字段中的后三个字节（即低 24 位）则由厂家自行指派，称为扩展标识符（extended identifier），只要保证生产出的网卡没有重复地址即可，如图 5-10 所示。可见，一个地址块有 2^{24} 个不同的地址。用这种方式得到的 48 位地址称为 MAC-48，它的通用名称是 EUI-48。