

参考资料

- 西田佳史. TCP 詳説 [EB/OL]. パシフィコ横浜: Internet Week 99, 1999.
- 〈RTP: 实时应用程序传输协议〉(RFC 3550).
- T.Bova, T.Krivoruchka. Reliable UDP Protocol [EB/OL]. IETF Internet Draft, 1999.
- 石森礼二. モバイル対戦アクションゲームの通信最適化テクニック [EB/OL]. SQUARE ENIX オンラインゲーム・テクニカルオープンカンファレンス, 2018.
- 〈第 2.5 代和第 3 代无线网络上的 TCP〉(RFC 3481).
- 〈流控制传输协议〉(RFC 4960).
- 〈数据报拥塞控制协议〉(RFC 4340).

第 **2** 章

TCP/IP 的变迁

随着互联网的普及而不断进化的协议

TCP/IP 从诞生以来就随着互联网的普及而逐渐推广开来，并一直发展至今。其中，伴随着新技术与新应用服务的普及，初版 TCP 无法解决的问题一个个暴露出来。

因此，为了解决这些问题，TCP 不断改良，逐渐加入了许多现在常用的功能。其中最典型的便是拥塞控制算法，它可以在网络上的数据流量增加时防止网络拥塞，避免拥塞崩溃。

本章将结合时代技术背景和当时的流行应用服务等，介绍从 TCP 问世之初到现在为止的发展历程。TCP 发展过程中出现的各种难题究竟是如何被解决的？TCP 又是如何发展到现在这个样子的？理解了这些问题，想必会对现代 TCP 的各种机制和算法有更深入的理解。

2.1

TCP 黎明期

1968 年—1980 年

下面将详细介绍不同时期 TCP 相关技术的发展情况。这里以年表的形式整理了从 1968 年互联网的前身——阿帕网项目启动开始到现在为止 TCP 和互联网领域的大事件，请根据需要参阅（表 2.1）。

表 2.1 TCP 和互联网领域的主要事件

时间	年份	概要
20 世纪 60 年代	1968	阿帕网项目启动
	1969	UNIX 问世
20 世纪 70 年代	1970	ALOHAnet 问世
	1972	阿帕网公开实验
	1974	TCP 问世
	1976	公开密钥理论诞生
	1980	以太网标准公开
20 世纪 80 年代	1980	拥塞崩溃问题浮出水面
	1981	当前 TCP 的标准文档 RFC 791、792、793 公开
	1982	SMTP 正式 RFC 化
	1983	阿帕网全面使用 TCP/IP
	1983	域名规范化
	1983	4.2 BSD（默认支持 TCP 的 UNIX OS）公开
	1984	开始使用域名
	1984	引入 Nagle 算法
	1987	NSFnet 启动
	1988	Tahoe 出现 拥塞控制算法的起点

(续)

时间	年份	概要
20 世纪 90 年代	1990	阿帕网关闭，商用互联网开始出现
	1990	万维网问世
	1990	Reno 算法出现
	1991	万维网公开
	1995	Windows 95 发售
	1995	NSFnet 停止运行
	1995	PHS 服务启动
	1999	IPv6 投入使用
	1999	无线 LAN (IEEE 802.11a) 出现
21 世纪 00 年代	2001	维基百科 (Wikipedia) 诞生
	2002	RFC 3261 (IP 电话) 发布
	2003	Skype 诞生
	2004	Facebook、mixi、Firefox 诞生
	2005	YouTube 诞生
	2005	CTPC 出现
	2006	Twitter、AWS (Amazon Web Services)、Niconico 动画诞生
	2006	云计算出现
21 世纪 10 年代	2008	iPhone 发售、App Store 出现
	2010	3.9G (LTE) 出现
	2015	4.0G (LTE-Advanced) 出现

本节将从 1968 年阿帕网项目启动开始到 TCP 问世、TCP/IP 基本成形的时期称为 TCP 的黎明期。

阿帕网项目启动 (1968 年) 分组交换的出现

现在的互联网主要使用的通信方式是分组交换，但在分组交换的概念流行起来之前，“线路交换”才是最常见的方式 (图 2.1)。

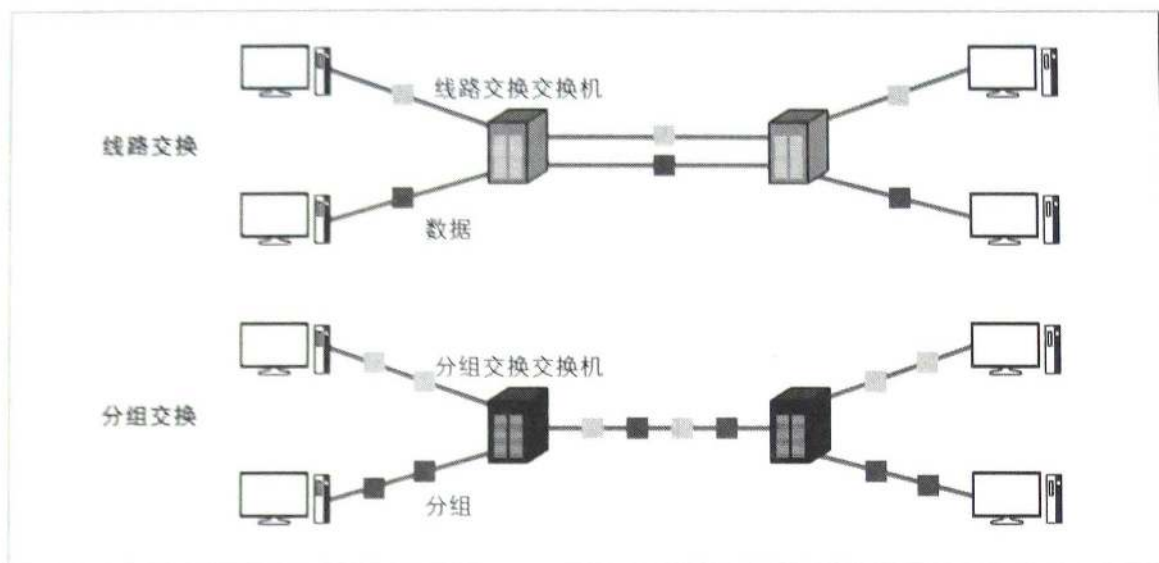


图 2.1 线路交换与分组交换

——线路交换

线路交换类似于传统的电话网络，需要独占从拨打方（发送方）终端设备到接听方（接收方）终端设备的电气线路。如果发送方设备与接收方设备的地理位置不同，通信则会经过多个通信局，而且要维持专用的线路。

20 世纪 60 年代前半期，线路交换方式的远程计算机连接通过电话线路成功得以实现。1966 年，横跨几乎整个美国大陆的计算机通信实验也成功进行。然而，此时的通信连接是通过线路交换实现的，只能是一对一（Point-to-Point, P2P）通信，而且必须保证计算机间的独占线路。

这种方式被指出存在若干问题。首先，计算机间数据通信的时间较短，如果采用线路交换方式建立连接，就需要一直保持连接，效率很低。也就是说，由于要保证临时的独占线路，所以在连接的过程中，其他所有的通信都无法使用链路上的任何线路，即必须为每个通信分别准备线路。

其次，因为只能进行一对一的连接，所以如果有三台以上的计算机要进行通信，就必须一条条地更换线路，非常麻烦。不仅如此，也有人指出，如果进行线路交换的交换局受到攻击，所有的通信都会中断，这也非常危险。

此外，还存在一个问题：当时各个设备制造厂商都使用安装了自己独家系统的设备，想要互相通信，就必须准备基于各式各样系统的终端设备。也就是从这个时期开始，人们逐渐意识到计算机通信中约定俗成的规范，也就是“协议”的必要性。

—— 分组交换

针对以上问题开发出来的便是被称为分组交换的通信方式。

分组交换首先将计算机间收发的信息拆分为多个被称为分组的单位，然后确定发送链路，最终将创建的各个分组经过多个网络中的设备发送到目的地。

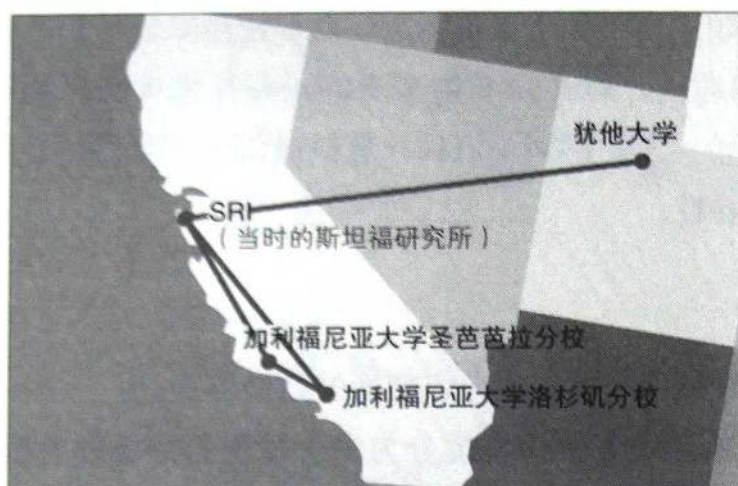
使用分组交换之后，通信就无须确保一对一的独占线路了，多台计算机也能共享线路。因此，这种做法不仅提高了线路的利用效率，而且因为无须一条条地切换线路，所以实现高效通信也不在话下。此外，通信也不再需要“交换局”这种必经节点，所以能更好地抵御攻击。

—— 阿帕网 世界上首个使用分组交换的网络搭建工程

接下来，便是 1968 年由美国 ARPA（Advanced Research Project Agency，美国高级研究计划局^①）发起的阿帕网，它是以构建分组交换式计算机网络为目的的代表性工程。

1969 年，研究人员搭建了由斯坦福研究所、加利福尼亚大学洛杉矶分校、加利福尼亚大学圣芭芭拉分校、犹他大学等 4 个节点相连组成的网络，这便是初期的阿帕网（图 2.2）。阿帕网使用了 NCP（Network Control Protocol，网络控制协议）——注意，并非 TCP/IP 协议——是世界上第一个使用分组交换的计算机网络。

^① 现在是美国国防高级研究计划局，即上一章提到的 DARPA。——译者注



※ 参考: SRI International 官方网站

图 2.2 初期的阿帕网

此外，阿帕网项目首次采用了 RFC (Request For Comments) 这一用于互联网技术标准化发布的流程。

UNIX 问世 (1969 年) OS 和 TCP/IP 的普及

通信协议指的是多台计算机接入网络时，为了让它们能互相进行通信而由人们事先定好的一些规则。也就是说，要正确地完成通信，所有接入网络的计算机都必须遵循同一套通信协议。类比来说，在使用同一种语言的小团体中，如果有一位不懂这门语言，那么他就无法与小团体中的其他人沟通交流。

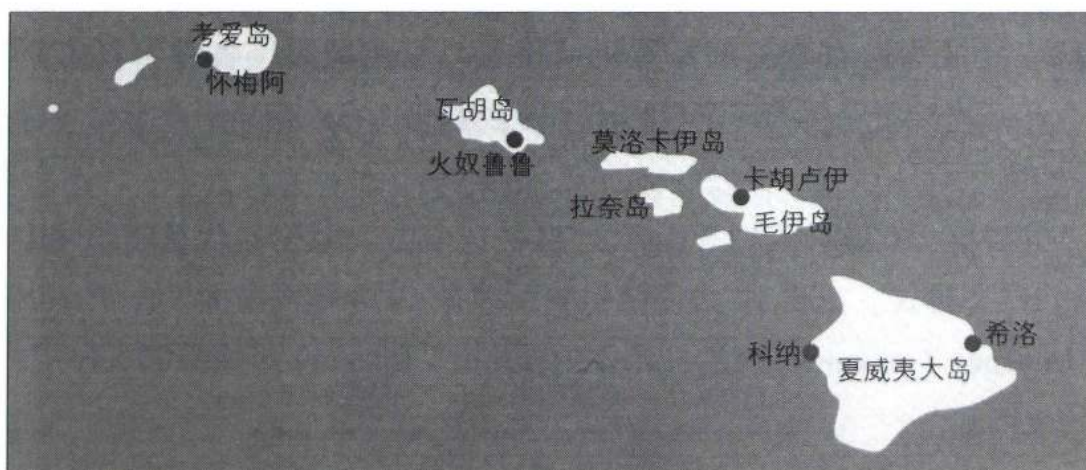
因此，比较重要的问题是，即使确定了新的通信协议，如果搭载协议的设备没有普及，那么设备间也无法通信。如今的计算机设备一般在操作系统中就安装了通信协议功能。在搭建阿帕网的同一年，即 1969 年，UNIX 操作系统在 AT&T 的贝尔实验室中被开发出来。

UNIX 在开发初期是由汇编语言实现的，1973 年用 C 语言进行了重写，这大大地提高了 UNIX 在各种计算机上的可移植性。后来，UNIX 的源代码被免费提供给大学和研究机构，它成了一个可以自由修改的操作系统，各种版本被开发出来，这使 UNIX 在研究机构和教育机构中广泛地普及开来。其中最具有代表性的，便是由加利福尼亚大学伯克利分校开发的 BSD (Berkeley Software Distribution)。这个 UNIX 系统不仅在后来成为阿

帕网所使用的操作系统，还作为首个默认搭载 TCP/IP 的操作系统，为 TCP/IP 的普及做出了巨大的贡献。

世界上首个无线分组交换网络， 搭建 ALOHAnet (1970 年) 冲突避免技术的前身

1970 年，连接夏威夷大学各个校区的计算机网络 ALOHAnet 完成了搭建（图 2.3）。这是世界上首个无线分组交换网络，它连接了分布在夏威夷群岛的各校区。



※ 出处：Wireless Communication, Jean Paul Linnartz' Reference Website

图 2.3 ALOHAnet 的节点分布

接入 ALOHAnet 中的计算机，可以在任意时刻向任意一台计算机发送数据。但是，此时会出现很严重的问题。具体来说，就是如果多个节点同时发送数据，那么这些数据的信号之间会发生冲突，导致数据全部损坏，进而导致无法通信。针对这一难题，ALOHAnet 所用的对策可以说是网络发展中十分重要的一项技术，它与后来被标准化的 CSMA（Carrier Sense Multiple Access，载波侦听多路访问）一脉相承。

下面大致介绍一下 ALOHAnet 所用的避免冲突的方法。各个节点在想发送数据时直接发送数据，如果发生冲突，就过一段时间重新尝试。为了让发送数据的节点探测是否出现了冲突，需要由名为 hub 的节点在收到分

组之后，立即将分组数据返回给发送节点。发送节点在确认收到了返回的数据后，认为当前数据已经发送成功，并开始发送下一个分组。如果 hub 没有返回数据，发送节点会认为发生了冲突，并在等待一小段时间之后，重新发送分组数据。

ALOHAnet 的带宽利用率并不高，但是它所采用的避免冲突的策略对后世的技术有很大的影响。同时，这种被称为介质访问控制的技术的重要性也开始被广泛认知。

TCP 问世（1974 年）焕然一新网络基本策略

后来也不断有新计算机接入前文所介绍的阿帕网。到了 1974 年，阿帕网已经成为遍布全美的计算机网络了。与此同时，各种协议纷繁芜杂，相互连接中发生的问题也层出不穷。因此，有人指出需要整合各种网络协议。

针对这一问题，人们开始开发 TCP/IP 协议，目的是代替阿帕网中使用的 NCP 协议。TCP/IP 协议最初的标准是 1974 年发布的 Specification of Internet Transmission Control Program（Internet 传输控制程序规范，RFC 675）。

TCP/IP 协议最基本的策略就是最大限度地缩减网络的功能。一旦网络的功能多样化，必然会导致消耗增加、相互连接困难，以及搭建与维护困难等一系列问题。为了规避这些问题，人们倾向于使用更加简单的网络，这正是 TCP/IP 之后流行的重要原因之一。

此时，还有一个重要的情况需要关注。当时的主流观点是，以阿帕网为代表的各种网络，应该由其自身来“确保数据在网络中的传输可靠性”。但是 TCP/IP 颠覆了这种观点，引入了新的思路，也就是由发送数据一方的计算机来确保通信的可靠性。换句话说，由错误等情况引起的数据无法送达的问题的检测和重传控制，都交由发送方终端设备负责。

像这样将网络的功能控制到必要的最小范围内，便可以降低各种网络互相连接的门槛。1980 年左右，TCP/IP 的基本架构完成；1983 年初，阿帕网的通信协议从过去的 NCP 完全切换到 TCP/IP。此外，如前文所述，

1983 年 UNIX 系操作系统 BSD 也开始默认搭载 TCP/IP 协议，这进一步促进了 TCP/IP 的普及。

以太网标准公开（1980 年）IEEE 802.3 与 CSMA/CD

以太网标准的原型是 20 世纪 70 年代以 ALOHAnet 为基础开发出来的。后来，它于 1980 年由 IEEE 提出和公开，并在经过升级后，于 1983 年以 IEEE 802.3 的名称正式确定下来。

以太网有许多特点，其中第一点便是继承了分组交换这一特性。以太网将待发送的数据分割为帧，网络中的设备以帧为单位传输数据。

此外，初期的以太网在逻辑上是“总线式”结构（图 2.4）。换句话说，多台计算机是连接在同一根同轴线缆上的。在这种情况下，由任意一台终端设备发送出的数据可以被所有连接在这个网络上的计算机接收。也就是说，这个网络从原理上就无法进行一对一通信，所有的通信都是广播式的。

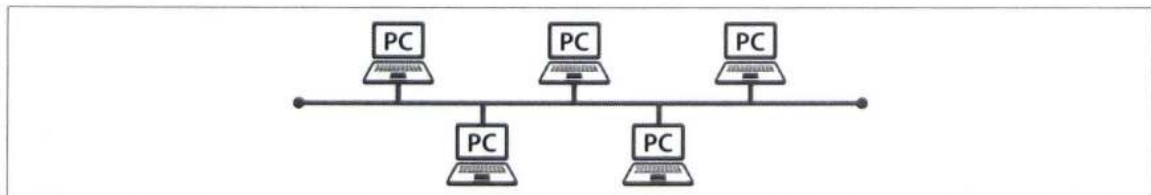


图 2.4 总线式网络

然后，所有已连接的设备共享同一个通信介质，一旦多台终端设备同时发送数据，就会发生信号冲突，导致数据丢失。发生冲突的范围称为冲突域（collision domain）。因此，当有多台终端想要发送数据时，需要按顺序来发送。实现这一功能的就是 CSMA/CD（Carrier Sense Multiple Access with Collision Detection，带有冲突检测的载波侦听多路访问），这项技术也被认为是以太网的代表名片。CSMA/CD 的控制流程如图 2.5 所示。

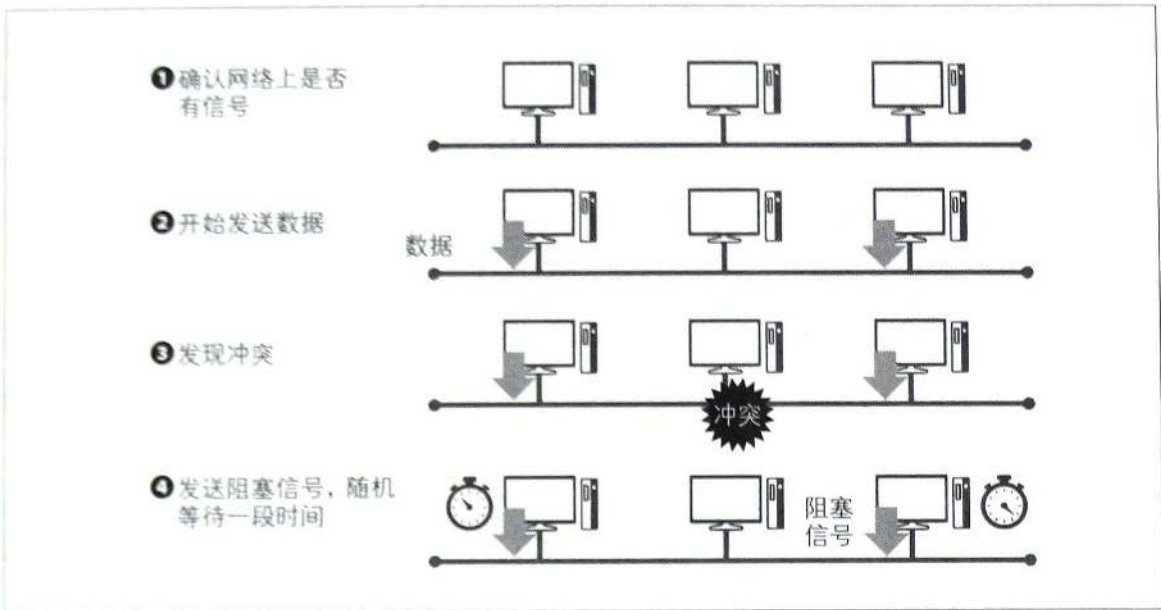


图 2.5 CSMA/CD 的控制流程

各台终端设备首先持续接收网络上的信号并进行检测 (❶), 当确认没有其他终端设备发送数据时, 开始发送数据 (❷)。如果发现因与其他终端设备发送的数据冲突而导致数据出错 (❸), 就发送一个称为阻塞信号 (jam signal) 的特殊信号 (❹), 通知其他终端设备检测到了冲突。之后, 检测出冲突的设备停止发送数据, 随机等待一段时间后再重传数据。

随后, 逐渐高速化的以太网作为 OSI 参考模型第 1 层、第 2 层的网络协议, 与作为第 3 层、第 4 层协议的 TCP/IP 一起, 广泛应用于全世界的有线 LAN 中。

2.2

TCP 发展期

1980 年—1995 年

TCP/IP 协议是为了实现通信设备间数据传输的可靠性而开发的协议。进入 20 世纪 80 年代之后, TCP/IP 不断发展, 添加了拥塞控制等诸多新功能。本节就介绍一下 20 世纪 80 年代到 20 世纪 90 年代前半期这段“TCP 发展期”。

拥塞崩溃问题浮出水面（1980 年）网络流量增加

拥塞指网络出现拥堵的情况。在 TCP/IP 刚开发出来时，由于网络流量不大，拥塞这一现象并不为大众所知，也没有成为问题。因此，TCP/IP 等当时的网络协议都没有拥塞控制功能。所谓拥塞控制功能，指的就是控制或避免拥塞的功能。

进入 20 世纪 80 年代，互联网上的流量逐步增长，拥塞问题日渐突出。特别是 TCP 网络一旦进入拥塞状态，如果不进行任何拥塞控制，就很难摆脱拥塞的状况。

在 TCP 中，发送的数据包如果丢失，发送方设备就会进行重传。在拥塞之后更容易发生数据包丢失，而这会导致重传更加频繁。因此，拥塞越严重，丢失的数据包便越多，进而重传也就越多，导致拥塞进一步恶化，这是一个恶性循环。这种状态如果一直持续下去，在最坏的情况下，网络可能会崩溃（图 2.6）。这种现象称为**拥塞崩溃**。进入 20 世纪 80 年代后，人们开始担心拥塞崩溃的发生。

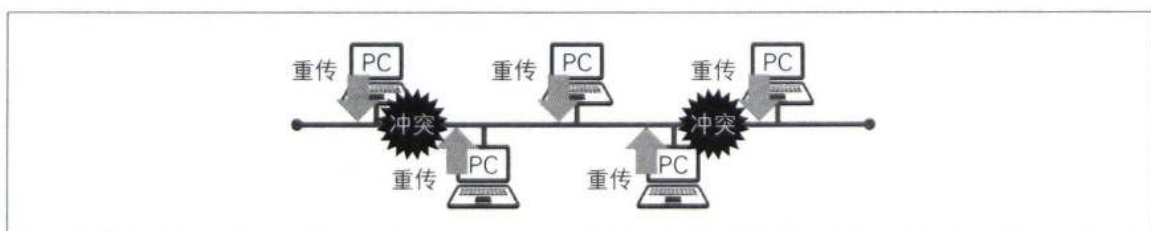


图 2.6 拥塞崩溃

引入 Nagle 算法（1984 年） 用于防止拥塞崩溃的拥塞控制相关技术的先驱

1984 年的 Congestion Control in IP/TCP Internetworks（IP/TCP 互联网上的拥塞控制，RFC 896）中提出了 Nagle 算法，其目的是减少 TCP/IP 网络上待发送数据包的数量。可以肯定地说，Nagle 算法是一切为了防止 TCP/IP 网络出现拥塞崩溃而进行的拥塞控制相关技术的“鼻祖”。

Nagle 算法的提出源于 Telnet 等应用程序每次发送的都是 1 字节左右的极小的数据单位。哪怕数据的有效载荷^① (payload) 只有 1 个字节, 当其在网络中被发送出去时, 也必须依次加上 TCP 的 20 个字节和 IP 的 20 个字节首部, 此外还要加入以太网 14 个字节的首部和 FCS 的 4 个字节, 最终组成长达 59 个字节的数据。当时的通信速度远比现在慢得多, 如此巨大的系统开销影响极大。如果 Telnet 会话的关键操作以 1 字节为单位发送数据, 那么就必须连续大量地发送这种小包, 这很可能引起拥塞。

因此, Nagle 算法采取以下策略, 即“将多个待发送的数据缓存起来, 合并发送”(图 2.7)。

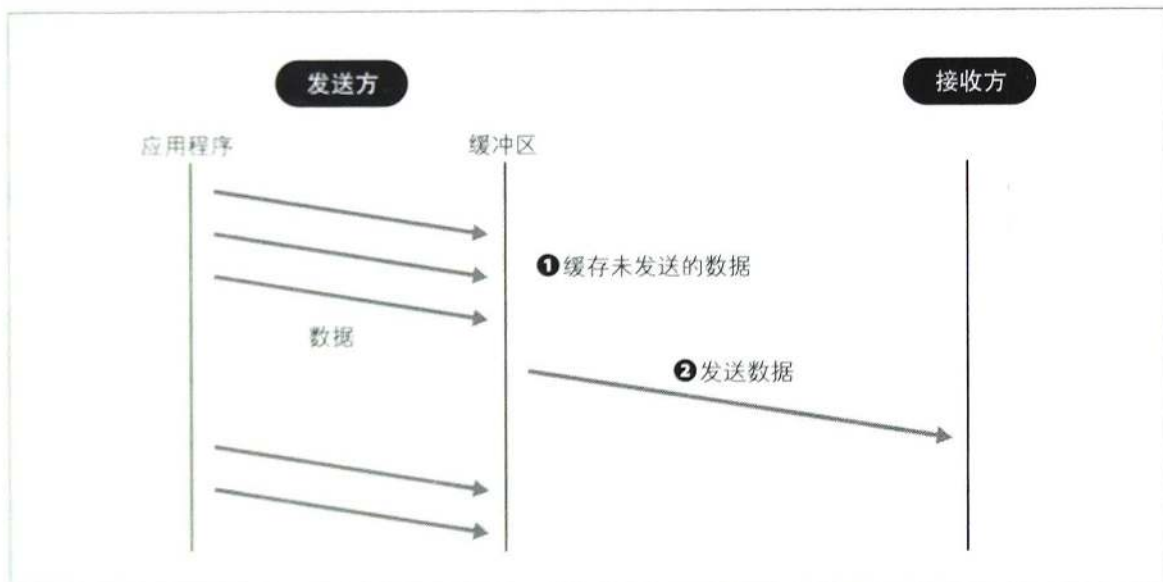


图 2.7 Nagle 算法的流程示意图

Nagle 算法的具体步骤如下所示。

- ① 发送方将未发送的数据存储在缓冲区中;
- ② 如果未发送数据的累计大小超过 MSS , 或者所有已发送的包都收到了对应的 ACK, 又或者出现了超时, 就把数据发送出去。

然而, Nagle 算法并非适用于所有环境的万能算法, 而是针对当时的

^① 真正待发送的实际数据本体。

网络环境和特定的应用程序的有效对策。话虽如此，Nagle 算法对于后世技术的影响也不可谓不大，它不仅让人们认识到了 TCP/IP 网络中拥塞崩溃问题的危险性，更促进了针对性解决方案的引入。

引入拥塞控制算法（1988 年） 根据情况调节数据发送量

1986 年 10 月，接入阿帕网的 NSFnet 发生了拥塞崩溃。NSFnet 由 NSF（National Science Foundation，美国国家科学基金会）于 1986 年建立，是一个为访问超级计算机提供支持的网路。拥塞崩溃导致其吞吐量由 32 Kbit/s 降到 40 bit/s，实际上降低到了原先千分之一左右的水平。虽然拥塞崩溃早在许多年前便一直令人担忧，但直到这次真正发生了，崩溃所造成的巨大影响才开始被人们广泛了解。

为了防止再次出现此类拥塞崩溃事件，拥塞控制算法 Tahoe 于 1988 年被提了出来。下一章会详细介绍拥塞控制算法，所以本节就只介绍一下此算法的大致流程与意义。

Tahoe 的工作流程如图 2.8 所示。发送方先缓慢增加数据发送量（❶），当检测到拥塞发生时（❷），就减少数据发送量（❸）。Tahoe 的目标是通过这一系列操作，使网络恢复到正常状态。

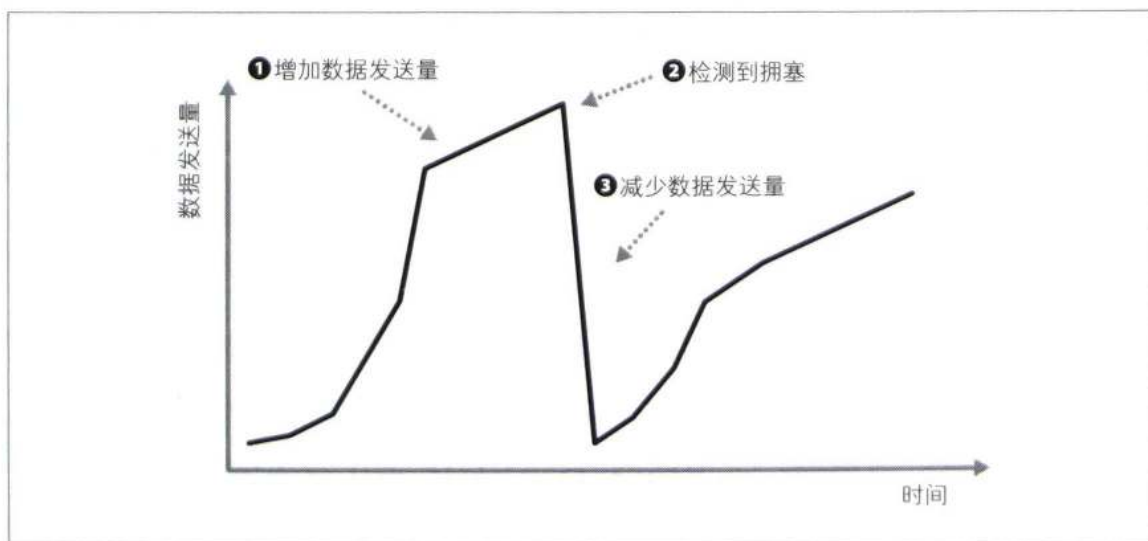


图 2.8 Tahoe 的拥塞控制流程

总的来说，各台终端设备不再以自己获得发送机会为中心行动，而是开始考虑“根据实际情况调整发送量”。Tahoe 在检测到拥塞之后，会将数据发送量调节到一个非常小的值，因此相比后面开发的算法，其带宽利用率相对较低，然而从“为 TCP 增加了拥塞控制”这一点来看，Tahoe 的作用仍然举足轻重。

由应用程序推动的迁移 往互联网的迁移与万维网的诞生（1990 年）

20 世纪 80 年代后半期，“互联网”一词开始成为指代由阿帕网与 NSFnet 相连而构成的网络的专有名词。

前文已经介绍过，TCP/IP 是将必要功能缩减到最小集的简单协议。不仅如此，TCP/IP 无须在意底层的物理网络构成，因此具有“与现有的其他网络通信十分容易”的优点。各个相互连接的网络，就这样形成了世界规模的 TCP/IP 网络。虽然后来阿帕网项目停止了，但那时“互联网”这一专有名词已经固定了下来。

随后，万维网（World Wide Web，WWW）于 1990 年被提出，接着 Web 网页、Web 浏览器也被开发出来。世界上第一个 Web 网站已被 CERN（European Organization for Nuclear Research，欧洲核子研究组织）复原出来，所以现在我们可以一览其风采（图 2.9）。

在万维网中，组成 Web 页面的文档都是由 HTML 之类的超文本（hypertext）语言编写的。超文本指的是这样一种机制：在文档中插入其他文档的链接（hyperlink，超链接），使网络上的文档可以互相引用。后来，HTML 由于语法也比较简单易懂，所以广泛地普及开来。

此外，这个时期互联网服务提供商（Internet Service Provider，ISP）开始出现，互联网接入逐步走向商业化。随后，万维网作为互联网上主要的应用程序被广泛地推广，与此同时，TCP/IP 也作为互联网所使用的协议而飞速普及。