UNIX操作系统设计与实现

陈华瑛　李建国　主编

電子工業出版社

UNIX操作系统
设计与实现

中国科学院计算技术研究所
陈华瑛　李建国　　主编

電子工業出版社

(京)新登字 055 号

内容提要

本书是一本完整地描述加里福尼亚大学伯克利分校研制的 UNIX 系统版本的设计和实现的权威性著作。全书共分十三章，分别描述了 UNIX 的历史和目的，4.3BSD 设计概貌、核心服务、进程管理、I/O 系统概述、文件系统、设备驱动程序、终端处理、进程间通信、网络通信、网络协议、系统启动。

每章末都有习题。练习题共分三级，可供不同层次的读者思考和选作。

本书可供计算机科技人员和大学生阅读参考。

UNIX 操作系统设计与实现

陈华瑛　李建国　主编

责任编辑　王昌铭

*

电子工业出版社出版(北京市万寿路)

电子工业出版社发行　各地新华书店经售

北京顺义李史山胶印厂制版印刷

*

开本：787×1092 毫米 1/16 印张：19 字数：462 千字

1992 年 3 月第 1 版　1992 年 3 月第 1 次印刷

印数 6000 册　定价：11.60 元

ISBN7-5053-1685-0/TP・370

目　录

前　言

UNIX 操作系统，由于它的开放性、可移植性和多用户、多任务等特点，不仅深受用户欢迎，而且为广大计算机厂商所青睐。当今，工作站上几乎清一色地运行 UNIX 操作系统，并且 UNIX 已遍及微型机、小型机、工作站、大型机、小巨型机，甚至巨型机等各种类型计算机。大多数多处理机、图形处理系统和向量处理系统也均选用 UNIX 作为操作系统。虽然 UNIX 设计的初衷是设计成一个分时系统，但现在已有不少厂家的 UNIX 实现中加入了实时功能。UNIX 已成为计算机界公认的标准操作系统，应用范围也越来越广。

自从 1969 年 UNIX 发表以来，UNIX 系统已发展出许多不同的、后来又复聚的流派，其中主要的、占领导地位的是两大流派。一是以 AT&T 为首的，它们开发了 AT&T 专利的 UNIX 系统 V，目前普遍使用的是系统 V 3.2 版，最新版本系统 V 4.0 版亦已开始使用。另一是加里福尼亚大学伯克利分校的计算机系统研究组(CSRG)研制的伯克利软件发行版本(BerkeleySoftware Distributions)，目前普遍使用的是 4.2BSD 和 4.3BSD，最新版本是 4.3BSD。伯克利版本，以它引进了新技术而著称。它已在 UNIX 中引进了许多有用的机制和程序，如 2BSD 中的正文编辑程序 VI，3BSD(第一个 VAX 上的 UNIX 系统) 中的请页虚存支持，4.0BSD 中的性能改进，4.1BSD 中的作业控制、自动配置和长标识符，以及 4.2BSD 和 4.3BSD 中引进的可靠信号、改进的网络、巧妙的进程间通信(IPC) 原语、快速文件系统和进一步的性能改进等等。4.2BSD 和 4.3BSD 中 TCP/IP 网络协议的实现已被广泛采用，许多厂家，不管他们的基础系统是 BSD 还是系统 V 或者甚至是 VMS，均采用了伯克利的网络实现。伯克利版本对 UNIX 的这些发展有许多已经合入系统 V 中。4BSD 对 POSIX(IEEE std 1003.1)操作系统接口标准及其他有关标准也有着很大的影响。POSIX 的一些特性：如可靠的信号，作业控制，进程的多个访问组，以及目录操作子程序等都来自 4.3BSD。

鉴于伯克利版本对 UNIX 的重大贡献，我们在 1990 年初组织人力精读了由 S. T. Leffler, M. K. Mekusick, M. J. Karels, J. S. Quarterman 著的《The Design and Implementation ofthe 4.3BSD UNIX Operating System》一书，该书初版于 1989 年，是第一本完整描述伯克利最新版本 4.3BSD 设计和实现的权威性著作。书中介绍了 4.3BSD 的内部结构和实现 4.3BSD 的系统功能中所用的概念、数据结构和算法。书中着重对 BSD4.3 和 AT&T 系统 V UNIX 版本的不同处作了较详细的描述，并对其设计思想及背景作了清晰的阐述。该书对研究、开发和使用 UNIX 系统，特别是 4.3BSDUNIX 实现中的一些新技术、新特点有很大的参考价值。为此，我们在精读该书的同时还对伯克利版本部分源代码进行了分析，在深入理解 4.3BSD 的设计和实现基础上编写了本书。操作系统实现者，系统程序员，UNIX 应用开发人员，管理人员以及对 UNIX 感兴趣的用户均可在阅读本书中受益。它亦可作为操作系统的高级教程。我们保留了原书中所有的习题和参考书目。习题分成三级，分别用无星号、一个星号、二个星号来标志。不带星号的习题，只要理解了书中内容，答案就可从书中的叙述内容找到。带一个星号的习题则除了运用书中所阐述的概念外还要作些推理分析。带两个星号的习题则提出了一些较大的设计项目或尚未解决的研究课题。书中用黑体字来标识一些术语、系统调用名、子程序名、变量名和结构名等。子

程序名后面随一对括号来标识，以和变量名区分。

全书分成五个部分：

第一部分，概述。前三章为导论性的，介绍了操作系统的概貌，并为书中后面的章节作了铺垫。第一章，历史和目的，概述了系统的发展历史，着重于它的研究方向。第二章，4.3BSD 的设计概貌，叙述系统所提供服务的性质，并概述了核心的内部结构。还讨论了研制系统时所作的设计决策。第三章，核心服务，阐述系统调用是如何完成的，并详细描述了核心的一些基本服务。

第二部分，进程。这部分的第一章至第四章，进程管理——是以后几章的基础，描述了进程的结构，调度进程执行所用的算法，以及为确保常驻核心数据结构的存取一致性所用的同步机制。第五章，存储管理，详细讨论了虚拟存储管理系统。

第三部分，I/O 系统。首先是第六章，I/O 系统概述，阐述 I/O 系统接口，并描述了支持该接口的机制的结构。随后三个章节讲述 I/O 系统三个主要部分的详细情况。第七章，文件系统，叙述实现文件系统的数据结构和算法。第八章，设备驱动程序，阐述块设备驱动程序和字符设备驱动程序，还叙述了决定哪些物理设备和系统相联的自动配置方法。提供了一个详细的磁盘设备驱动程序例子。第九章，终端处理，描述对字符终端的支持，并描述了面向字符的设备驱动程序。

第四部分，进程间通信。第十章，进程间通信，叙述提供相关进程间或无关进程间通信的机制。第十一章和第十二章，网络通信和网络协议是密切相关的，第十一章描述的机制大部分是依据特定的协议，如第十二章所述的 TCP/IP 协议实现的。

第五部分，系统运行。第十三章，系统启动，讨论系统启动、关闭、和配置，并阐述进程级的系统初始化，从核心初始化到用户注册进入系统。

参加本书编写工作的有陈华瑛、李建国、王建农、王钢、秦方中、穆小茜、徐力、戴征及刘翔等同志。由于时间仓促、水平有限，文中错误在所难免，敬请读者指正。

参考文献

Bach, 1986.

M.J. Bach, The Design of the UNIX Operating System, Prentice-Hall, Englewood Cliffs, NJ (1986)

Bentley & Kernighan, 1986,

J.Bentley & B. Kernighan, "Tools for Printing Indexes," Computing Science Technical Report 128, AT&T Bell Laboratories, Murray Hill, NJ (1986)

CSRG, 1986.

CSRG, "UNIX Programmer's Manual, 4.3 Berkeley Software Distribution, Virtual VAX-11 Version," Six Volumes and an Index Volume, University of California Computer Systems Research Group, Berkeley, CA (April 1986).

Kernighan & Ritchie, 1978.

B.W.Kernighan & D.M. Ritchie, "The C Programming Language," Prentice-Hall, Englewood Cliffs, NJ (1978).

Kernighan & Pike, 1984.

B. W. Kernighan & R. Pike. "The UNIX Programming Environment," Prentice-Hall, Englewood Cliffs, NJ (1984) .

Kernighan & Ritchie, 1988.

B. W. Kernighan & D. M. Ritchie, "The C Programming Language," 2nd ed. Prentice-Hall, Englewood Cliffs, NJ (1988) .

Libes & Ressler, 1988.

D. Libes & S. Ressler, "Life with UNIX, Prentice-Hall," Englewood Cliffs, NJ (1988) .

O'Dell, 1987.

M. O'Dell, "UNIX: The World View," Proceedings of the 1987 Winter USENIX Conference, pp. 35-45 (January 1987) .

Organick, 1975.

E.I. Organick, "The Multics System: An Examination of Its Structure," MIT Press, Cambridge, MA (1975) .

Peterson & Silberschatz, 1985.

J. Peterson & A. Silberschatz, "Operating System Concepts," Addison-Wesley, Reading, MA (1985) .

Quarterman et al., 1985.

J.S. Quarterman, A. Silberschatz, & J.L. Peterson, "4.2BSD and 4.3BSD as Examples of the UNIX System," ACM Computing Surveys 17 (4), pp.379-418 (December 1985) .

Ritchie & Thompson, 1978.

D. M. Ritchie & K. Thompson, "The UNIX Time-Sharing System," Bell System Technical Journal 57 (6 Part 2), pp. 1905-1929 (July-August 1978) . The original version [Comm. ACM 7 (7), pp. 365-375 (July 1974)] described the 6th edition; this citation describes the 7th edition.

Schwartz, 1987.

M. Schwartz, "Telecommunication Networks," Addison-Wesley, Reading, MA (1987) .

Stallings, 1985.

R. Stallings, "Data and Computer Communications," Macmillan, New York, NY (1985) .

Tanenbaum, 1988.

A.S. Tanenbaum, "Computer Networks," 2nd ed, Prentice-Hall, Englewood Cliffs, NJ (1988) .

第一章　历史和目的

1.1　UNIX 系统的历史

UNIX 系统是较早的广泛使用的操作系统之一，已有近 20 年的历史，但它的许多有特色的、有用的功能都是最近几年发展的。

起源

UNIX 系统的第一个版本是 1969 年由 Ken Thompson 在贝尔实验室作为一个个人研究项目利用 PDP-7 机器研制的。Thompson 还设计了语言 B，许多早期的系统都用 B 重写了。后来，他和 Dennis Ritchie 进行了短期使用，D. Ritchie 不仅参加了系统的设计和实现，而且发明了 C 程序设计语言，以后的版本均用 C 语言写。最初系统的精致设计[Ritchie，1978]和后来 15 年来的发展[Ritchie，1984a]，[Compton，1985]使 UNIX 系统成为一个重要的功能很强的操作系统[Ritchie，1987]。

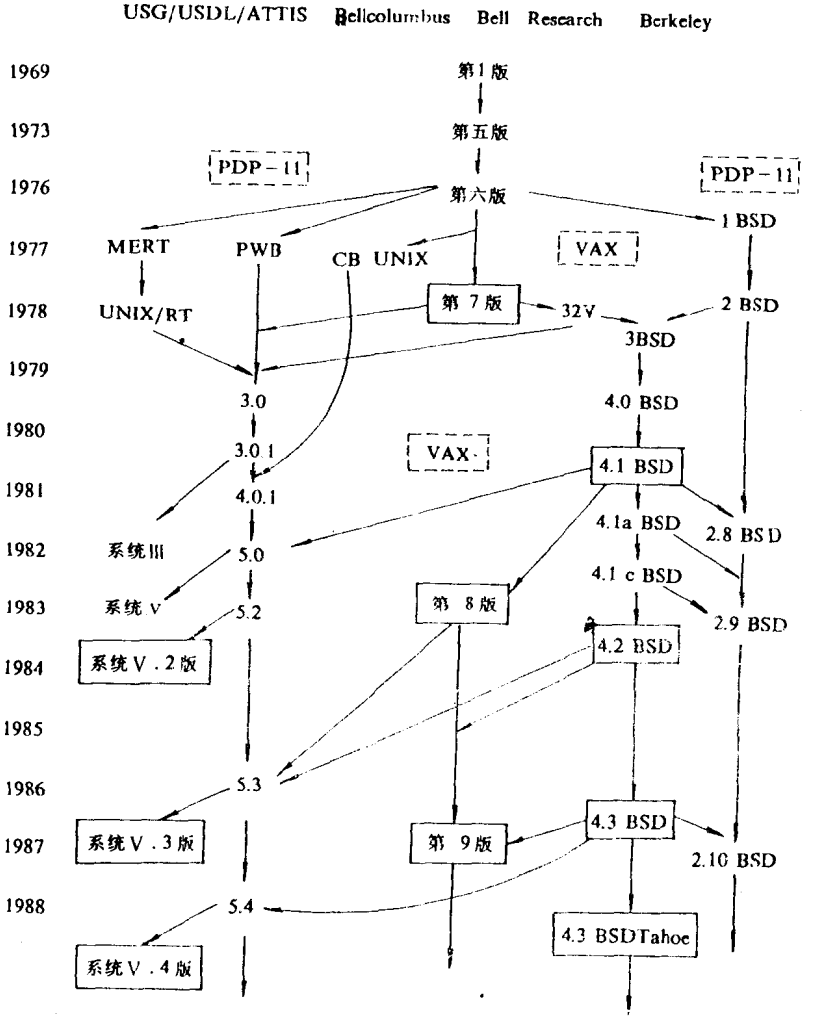

图 1.1UNIX 系统族树

Ritchie, Thompson 和其他早期研究 UNIX 的研究人员以前是在 Multics 设计项目工作的[Peirce, 1985; Organick, 1975], Multics 对这新的操作系统有很大的影响. 甚至"UNIX"这个名字就是关于 Multics 的双关语: Multics 要作许多事的地方, UNIX 试图去做好一件事. UNIX 文件系统的基本结构, 命令介释程序 (shell) 作为一个用户进程的思想, 通用的文件系统接口组织和许多其他系统特征均来自 Multics.

也应用了一些其他操作系统的思想, 如麻省理工学院 (MIT) 的 CTSS. 建立新进程的 fork 操作来自伯克利的 GENIE (SDS-940, 后来是 XDS-940) 操作系统. 允许用户建立进程, 使得可以容易地实现一条命令用一个进程而不是用过程调用来运行, 这是取自 Multics.

UNIX 系统的发展至少有三个主要的流派. 图 1.1 给出了变革的概貌, 特别是导向 4.3BSD 和导向系统 V 的那些分支[chambers & Quarterman, 1983; Uniejewski, 1985]. 图中的日期是大概的, 我们并没有想在图中列出所有的影响. 图中指名的某些系统在本书中没有提到, 列出它们是为了更清楚我们所考察的系统之间的关系.

研究 UNIX

前几个主要版本是贝尔实验室的研究用系统. 除了系统的最早版本外还包括 **UNIX 分时系统, 第六版**, 通常称之为 V6, 这是 1976 年第一个在贝尔实验室外广泛使用的版本. 系统用发布时的 UNIX 程序员手册的版本号来标识.

UNIX 系统和其他操作系统有三个重要区别: 1.系统是用高级语言写的. 2.系统用源码形式发布. 3.系统提供了功能很强的原语, 而这些原语往往只有在昂贵得多的硬件上运行的操作系统才提供的. 大多数系统源码是用 C 而不是用汇编语言写的. 在当时, 一般均认为操作系统必须用汇编语言写, 以提供合理的效率以及能访问硬件. C 语言本身是一种恰到好处的高级语言, 它既是高级的、能容易地在广泛类型的计算机硬件上编译, 又不是那么复杂或限制很大而使系统程序员必须转去用汇编语言写以取得合理的效率或功能. UNIX 操作系统的百分之三需访问硬件的部分一如环境切换, 需要用汇编语言写. 虽然 UNIX 的成功并不单是由于用高级语言写而带来的, 但是用 C 语言是关键的第一步[Ritchie 等, 1978; Kernighan 和 Ritchie, 1978; Kernighan 和 Ritchie, 1988]. Ritchie 的 C 语言是继承了 Thompson 的 B 语言导出的[Rosler, 1984], 而 B 语言是从 BCPL 发展来的[Richards 和 Stevens, 1982]. C 继续扩展[Tuthill, 1985; X3J11, 1988], 并且有了一个变种 C^{++}, 它能更容易地作数据抽象[Stroustrup, 1984; USENIX, 1987].

UNIX 的第二个重要特点是, 它的早期版本贝尔实验室是以源码形式向外界其他研究单位发布的. 由于提供了源码, 系统的创立者确保了其他机构可以不仅使用 UNIX 系统, 而且可深入系统内部设法修改和补充系统. 容易在系统中采纳新的思想是修改系统的关键因素. 每当出现一个新系统企图超过 UNIX 时, 就会有人对这新系统进行分析而把它的中心思想加入 UNIX. 这种在充满新思想的环境中使用一个小型的、用高级语言写的、容易理解的系统的独特能力, 使 UNIX 系统的发展远远超过了它一开始的粗糙系统的面貌.

UNIX 的第三个重要特点是它提供各个用户可以并发地运行多个进程, 以及把多个

进程连接为若干命令的管道线．当时，只有在大型的昂贵的机器上运行的操作系统有运行多个进程的功能，并且并发进程的数目通常是由系统管理员严格控制的．

早期大多数 UNIX 系统是在 PDP-11 上运行的，在当时，PDP-11 价格较低廉功能却较强．但是，至少有一个早期的第六版 UNIX 在一个体系结构十分不同的机器上运行[Miller，1978]．PDP-11 有一个不方便之处就是地址空间小．引入了 32 位地址空间的机器，特别是 VAX-11／780，使 UNIX 有可能扩大它的服务，使它能提供虚拟存储和网络．根据研究小组在不同的硬件上提供类一UNIX 功能的早期经验可得出如下结论：移植整个操作系统和在另一操作系统下复制它的服务一样容易．以可移植性为特定目标的第一个 UNIX 系统是 **UNIX 分时系统，第七版（V7）**，它可在 PDP-11 和 Interdata 8／32 上运行，还有一个 VAX 变种，称为 **UNIX／32V 分时系统版本 1.0（32V）**．贝尔实验室的研究小组还研制了 **UNIX 分时系统，第八版（V8）**．目前他们的系统是 **UNIX 分时系统，第九版（V9）**．

AT&TUNIX 系统Ⅲ和系统Ⅴ

在 1978 年发布第七版以后，研究小组开始向 UNIX 支持小组（USG）以外发行系统了．以前 USG 是在内部作为 **UNIX 程序员工作台（PWB）** 发布系统的，有时也向外界发布[Mohr，1985]．

第七版后第一个向外界发布的系统是在 1982 年发布的 **UNIX 系统Ⅲ（System Ⅲ）**，它加进了第七版、32V 的特点，还包括了研究小组以外的一些其他小组开发的几个 UNIX 系统的特点．包括了 UNIX／RT（一个实时 UNIX 系统）的特点，以及许多来自 PWB 的特点．USG 在 1983 年发表了 **UNIX 系统Ⅴ（System Ⅴ）**，这个系统大部分是从系统Ⅲ导出的．AT&T 董事会决定把贝尔实业公司从 AT&T 分出来后使 AT&T 得以将系统Ⅴ投入市场[Wilson，1985；Bach，1986]．

USG 又演变为 UNIX 系统开发实验室（USDL），它在 1984 年发布了 **UNIX 系统Ⅴ第 2 版**．系统Ⅴ的 2.4 版把请页[Miller，1984；Jung，1985]，包括写时复制(Copy-on-write) 和共享存储引入系统Ⅴ．系统Ⅴ的实现并不基于伯克利的请页系统．USDL 靠着 ATT 信息系统（ATTIS）而获得成功，ATTIS 在 1987 年发布了 UNIX 系统Ⅴ第三版．这个系统包括了流技术（STREAMS），这是采自 V.8 的进程间通信（IPC）机制[Presotto Ritchie，1985]．

其他组织

UNIX 系统易于修改这个特点，使得许多组织均可进行发展工作．如 Rand 公司，它负责发展了第十章提到的兰德信口（Rand Ports）；Bolt Beranek 和 Newman（BBN），第十二章讨论的 4.2BSD 网络实现就是直接继承了它们所提出的网络实现；Illinois 大学，它作了早期的网络工作；Harvard；Purdue；和 DEC 公司．

按照运行 UNIX 系统的机器数目而言，最广泛使用的 UNIX 操作系统或许应是 Microsoft 公司的 XENIX．XENIX 原来基于第七版，但后来又基于系统Ⅴ．最近，Microsoft 和 AT&T 已经同意合并这两个系统．

伯克利软件发行版本

除了贝尔实验室和 AT&T UNIX 研究小组以外，最有影响的就是加里福尼亚大学伯克利分校了[McKusick，1985]。伯克利的 UNIX 软件是以**伯克利软件发行版本（BSD）**形式发布的；例如 4.3BSD。伯克利的第一项 VAX UNIX 工作是 1979 年由 William Joy 和 Ozalp Babaoglu 作的，在 32V 中加入虚拟存储、请页和页替换，这就是 3BSD [Babaoglu Joy，1981]。3BSD 提供大虚拟存储空间是由于开发大型程序，如伯克利研制的 **Franz** LISP 的需要。这项存储管理工作得到国防部先进研究项目局（DARPA）的信任而决定投资支持伯克利研究组进一步研究一个为 DARPA 各项目承担单位使用的标准 UNIX 系统（4BSD）。

这个项目的目的之一是提供 DARPA 互连网的网络协议支持，TCP／IP[Cerf & Cain，1983]。网络实现是通用的，能在各种不同的网络设施间进行通信，从局部网如以太网和令牌网（Token rings），到远程网如 DARPA 的阿帕网（ARPANET）。

我们把 3BSD 以后所有的伯克利 VAX UNIX 系统均称为 4BSD，虽然实际上有好几个版本—4.0BSD，4.1BSD，4.2BSD，4.3BSD 和 4.3BSD Tahoe。从 1977 年 VAX 最初投入运行开始到系统Ⅲ的发布（从 1979 到 1982），4BSD 一直是 VAX 族选用的操作系统，并且为许多研究单位或网络安装单位所采用。大多数组织愿意买 32V 的许可证，然后向伯克利订购 4BSD。贝尔系统内部许多计算机系统运行的是 4.1BSD（现在，许多系统运行 4.3BSD）。

为 DARPA 研制的 4BSD 工作受指导委员会的领导，这个指导委员会包括许多事业单位和科学研究所的著名人士。伯克利 DARPA UNIX 计划的最盛期是 1983 年 4.2BSD 的发布；后来伯克利进一步发展产生了 4.3BSD，且工作仍在继续发展。

UNIX 在全球

许多计算机厂家，包括几乎所有在市场上占有主要份额的厂家，或宣布或引入运行 UNIX 系统或类似 UNIX 系统的计算机。还有许多公司出售有关的外部设备、软件包，提供维护支持、培训、资料等等。硬件系统则从微型机到小型机、多处理机、以及大型机甚至到巨型机。大多数系统使用系统Ⅴ、4.2BSD、4.3BSD 或它们的结合，不过还有一些机器运行基于系统Ⅲ、4.1BSD 和第七版的软件。有一些 PDP-11 运行 2BSD 和其他 UNIX 变种，甚至还有一些第六版的系统仍在正常运行。

UNIX 系统在科学研究领域也成果累累。Thompson 和 Ritchie 由于系统的设计被授予 ACM 的图林奖[Ritchie，1984b]。UNIX 以及有关的专为教学设计的一些系统，（如 Tunis [Ewens 等，1985；Holt，1983]）都广泛用于操作系统的课程中。UNIX 系统遍及全世界的大学和研究机构，在工业和商业方面使用得更广泛。

1.2 BSD 和其他系统

计算机系统研究组（CSRG）不仅吸收了 UNIX 系统的特点，而且吸收了其他操作系统的特点。4BSD 的许多终端驱动程序来自 TENEX／TOPS-20。作业控制（从概念上而言，不是指实现）是来自 TOPS-20 和 MIT 的不兼容分时系统（ITS）。虚存接口先是为 4.2BSD 而提出的，因为是由几个商业厂商实现的，所以它基于首先出现在

TENEX／TOPS-20 的文件—映射（file-mapping）和页式（page-level）接口。在设计新的功能时，常常参考 Multics 系统。

效率问题一直是 CSRG 工作中考虑的主要因素。在和 VAX 的专利操作系统 VMS 作比较后，已经作了一些效率上的改进[Kashtan，1980；Joy，1980]。

另外有一些 UNIX 变种采用了若干 4BSD 的特点。AT&T UNIX 系统 V[AT&T，1987]和 IEEE 1003.1 POSIX 标准[P1003.1，1988]已采纳了下列文件系统的系统调用（参见第六章）：

- 重新命名文件和目录的 **rename**
- 建目录和删目录的 **mkdir** 和 **rmdir**

上述系统调用中还采用了目录—访问子程序。

此外，POSIX 和有关的国家标准局（NBS）联邦信息处理标准（FIPS）已经采纳了

- 可靠的信号（第四章）
- 作业控制（第二章）
- 多个文件—存取许可权组（第六章）

X／OPEN 组织原来仅由欧洲的厂家组成，现在有一些美国公司也参加了，它产生的 X／OPEN 可移植性指南[X／OPEN，1987]，是一份描述核心接口和许多用户用实用程序的资料。X／OPEN 已经采纳了 POSIX 中许多功能。其他类似的标准和指南也期望采纳它们。IEEE1003.1 标准也是一个 ISO 的国际标准草案，称为 SC22WG15。因此，POSIX 或许将被全世界大多数类—UNIX 系统所接受。

4BSD 的**套接口**（Socket）进程间通信机制（参见第 10 章）是为可移植性设计的，并立即就移植至 AT&T 的系统Ⅲ，虽然系统Ⅲ从来没公布过。TCP／IP 网络协议的 4BSD 实现广泛用作为许多系统，从运行系统 V 的 AT&T3B 机器到 VMS、IBM PC，进一步实现的基础。

CSRG 还始终和各个基于 4.2BSD 和 4.3BSD 系统的生产厂家保持密切的联系。这种联合开发便于进一步发展 4.3BSD，使系统不断前进。

用户团体的影响

许多伯克利 UNIX 的开发工作是由用户团体负责进行的。一些思想和要求不仅出自 DARPA 这个主要的指导—投资组织，而且还来自各地的公司和大学的许多系统用户。

伯克利研究工作者不仅接受来自用户团体的思想，而且吸收源自实际软件的思想。澳大利亚、加拿大、欧洲和美国的许多大学和其他单位都对 4BSD 有过贡献，其中包括如自动配置和磁盘限量等一些主要特点。有些思想，如 **fcntl** 系统调用取自系统 V，虽然由于许可证和价格因素妨碍了在 4BSD 中使用系统Ⅲ或系统 V 的任何实际代码。除了在发行版本中包括这些贡献外，CSRG 还发行了一组用户开发的软件。

用户团体开发软件的一个例子是 4.3BSDTahoe 版本中采用的公用域时区处理（public-domain time-zone-handling）软件包。它是由 Arthar Olson、Robert Elz 和 Guy Harris 组成的国际组织设计和实现的，部分还出自于 USENET 通信组 **Comp.std.unix** 的讨论。这个软件包采用的时区转换规则完全独立于 C 库，而是放在一些文件里，使得要改变时区规则时不需要改系统代码；这对只有 UNIX 的二进代码系统时特别有用。这个

方法也使各个进程可选用不同的规则，而不需要全系统只能用同一组规则．该版本还有一个包括世界各地许多地区（从中国到澳大利亚、欧洲）使用规则的大型数据库．因此，UNIX 系统的发行就简化了，不需要为不同地点建立不同的软件包，而只要包括整个数据库就行了．

伯克利通过电子邮件（地址为 **bsd-bugs@ berkeley.edu**）接收有关系统故障和修改的信息，UNIX 软件办公室 MT XINU 接受委托负责编汇发行故障表．许多故障的修复被包括在以后发行的版本中．有一个普遍性的、(包括 4.3BSD) 固定的 UNIX 讨论在 DARPA Internet 电子邮件、名单 UNIX-WIZARDS 中，它在 USENET 网络中通信组名为 **comp.unix.wizards**; Internet 和 USENET 都是国际范围的．还有一个 USENET 通信组专门用于 4BSD 故障的: **comp.bugs.4bsd**．伯克利很少是直接从这些通信组和故障表中吸取思想的，因为很难从那么多委托信息中筛选．但是，现在有一个 CSRG 专门负责确认故障修复的仲裁通信组，称为 **comp.bugs.4bsd.bug.fixes**，在这些通信组中的讨论有时导出一些以后版本中采纳的新功能．

1.3 4BSD 的设计目的

4BSD 是为研究单位，(近来也为商业用户)，研制的研究系统 (research system)．研制者在编写系统时考虑了许多设计课题．设计中采用了一些不同于常规的考虑，但是带来了很好的市场效果．

早期的系统是由技术-驱动的，采用了当时在其他 UNIX 系统中还不能用的硬件新技术．这些新技术有：虚拟存储支持；第三厂家（非 DEC 的）外围设备的设备驱动程序；独立于终端的屏幕使用支持库，还开发了许多用这些库的应用，包括屏幕编辑程序 vi．和 AT&T 版本在 32V 中提供很少几种外围设备支持相比，4BSD 支持大量第三厂家的外围设备．这也是 4BSD 之所以广泛使用的一个重要因素．直到其他厂家开始提供它们自己的基于 4.2BSD 的系统支持以前，对于必需使硬件费用尽量少的大学来说，除了使用 4BSD 没有其他选择方案．

独立于终端的屏幕支持虽然现在看来是很平常的，但在当时对伯克利软件的普及是非常重要的．

4.2 BSD 的设计目的

DARPA 要求伯克利研制 4.2BSD，作为 VAX 上的标准研究操作系统．4.2BSD 设计中包括了许多新的功能：全面修改过的能支持大的进程地址空间的虚存系统，快速文件系统，进程间通信机制和网络支持．快速文件系统和修改后的虚拟存储系统是从事计算机辅助设计和制造 (CAD／CAM)、图象处理和人工智能 (AI) 所需要的．进程间通信功能是在分布式系统方面从事研究工作所需要的．网络支持最初是由于 DARPA 计划需要把研究人员通过每秒 56K 位的 ARPA 互连网连系起来．(虽然伯克利也感兴趣于在高速局部网上能获得良好的性能)．

并没有企图提供一个真正的分布式操作系统[Popek，1981]，而是采用了传统的 ARPANET 的资源共享目标．选择资源共享设计有三个理由：

・系统分布广泛而且要求管理自治．当时，一个真正的分布式操作系统需要有一个集

中管理特权（central administrative authority）.

• 一些已知的算法还不能很好地适应紧耦合系统.

• 伯克利的许可权是加入当前已经证明了的软件技术，而不是研制新的未经证明的技术. 因此，提供了一些容易实现的方法用于远程注册进入（rlogin, telnet），文件传输（rcp, ftp）和远程命令执行（rsh），但是所有的宿主机在用户面前仍保持是独立的系统.

由于时间的限制，4.2BSD 系统发行时并没有包括原来设计时企图包括的所有功能. 特别是修改过的虚存系统并未包括进去. 但是，CSRG 继续不断地努力跟踪快速发展的硬件技术. 网络系统支持广泛类型的硬件设备，包括 10 兆位／秒以太网、环型（ring）网和 NSC 的 Hyperchannel 等的接口. 核心源代码已经模块化并重新组织，使它易于移植到新的体系结构，从微处理机到大型的机器.

4.3 BSD 的设计目的

4,2BSD 存在一些问题，这促使了 4.3BSD 的产生. 因为 4.2BSD 中包括了许多新的功能，其中一些功能在实现中有些故障和错误，特别是在 TCP 协议实现中. 一些功能的性能不够好，其中部分是由于引入了符号链接. 还有一些功能由于时间不够而没有包括在内. 此外，如 TCP／IP 子网和路由支持没有及时作详细说明就加在 4.2BSD 中了.

商用系统通常要求版本保持向上兼容性，以使已有的应用软件不废弃. 但是，要维持兼容性越来越困难，所以大多数研究系统只维持少得可怜的向上兼容性. 作为一种折衷办法，BSD 通常对一次版本发行（one release）向上兼容，并对不兼容功能明确地标出. 这种方法使得可以逐步转移到新的接口，而又不限制系统逐步发展. 特别是 4.3BSD 很好地考虑了和 4.2BSD 的向上兼容性，这是应用程序可移植性的需要.

C 语言与 4.3BSD 的接口和 4.2BSD 只有少数终端接口命令以及一条 IPC 系统调用用一个参数（**select**: 参见 2.6 节）等地方不一样. 4.3BSD 中，建立信号处理程序的系统调用加了一个标志，使进程可以请求使用 4.1BSD 的信号含义而不是 4.2BSD 的含义（见 4.7 节）. 该标志的唯一目的是使已有的依赖于老的信号含义的应用程序不用重新编写可以继续使用.

4.2BSD 和 4.3BSD 在实现方法上改变很多，这些改变用户是见不到的. 例如，支持多个网络协议（如除了支持 TCP／IP 外还支持 XEROX NS）的实现作了改进. 因为 4.3BSD 保持了以前各伯克利系统的先进性，本书对 4.3BSD 的考察就是研究了伯克利系统的主要特点.

4.3BSD 的第二个版本，后来称之为 4.3BSD Tahoe，加进了对计算机控制台公司（CCI）加强系列 6（Tahoe）小型计算机的支持. 虽然总体上类似于原来的 VAX 用 4.3BSD 版本，但还包括了许多修改和新的特点.

未来的伯克利版本

4.3BSD 不是完善的. 特别是虚存系统需要全部替换掉. 新的虚存系统要提供能不怎么依赖于 VAX 体系结构的、又能更好地适合于大容量内存和当前可用的慢速磁盘的算法. 终端驱动程序已经仔细地设计以保持和第七版甚至第六版的兼容性. 这个特点曾经是

有用的，但现在不那么有用了，特别是考虑到它的命令和选项不够标准化．CSRG 计划发布一个和 POSIX 兼容的终端驱动程序，因为系统 V 将遵循 POSIX，所以这个终端驱动程序亦将和系统 V 兼容．一般来说，POSIX 兼容性是一个设计目标．

BSD 当前计划要做的另一些工作包括：研究国际标准化组织（ISO）网络协议的实现，TCP／IP 性能的进一步改进和加强，以及其他网络协议的实现．

4.3BSD 最关键的缺点是缺少了一个分布式文件系统．正如网络协议那样，没有一个分布式系统能在所有情况下提供足够的速度和功能．正如需要运行若干个不同的网络协议那样，常常需要支持若干不同的分布式文件系统协议．因此，将要开发一个文件系统的标准接口，它十分类似于 Sun Microsystem 的网络文件系统（NFS）方案，但比它更通用，使得既可支持本地文件系统也能支持远程文件系统，就像 4.3BSD 可支持多种网络协议那样[Sandberg，1985]．

IPC 处理模块的灵活配置这项工作最初是贝尔实验室在 UNIX 第八版中作的[Presotto Ritchie，1985]．流 I／O 系统（stream I／O system）基于 UNIX 的字符 I／O 系统．它允许用户进程打开一个不加工终端端口（raw terminal port），然后插入适当的核心处理模块，就像人们作一般的终端行编辑一样．处理网络协议的模块也可以插入．把终端处理模块放在网络处理模块的上层能灵活有效地在核心中实现网络虚拟终端．但是，流模块的问题是它们从本质上来讲是线性的．因此对处理与基于数据报文的网络中多路复用有关的扇入扇出来说是不够的；这种多路复用是在专用模块下层的设备驱动程序完成的．第八版的流 I／O 系统被系统 V.3 中采用作为 STREAMS 系统．

4.2BSD 的网络功能设计采用了不同的方法，它基于套接口（socket）和灵活的多层网络体系．这个设计使一个系统能支持多种网络协议，使用流、数据报文和其他类型的存取方式．协议模块可以处理单一传输介质上不同连接的数据的多路复用，以及从每个网络设备接收到的不同协议和连接的数据的多路分解．

计划要重新设计协议的内部分层，准备应用 V8 流 I／O 系统的思想．将使用套接口而不用字符设备接口，多路分解将在内部由核心中的网络协议处理．但是和流相似地，在多路复用上层的核心协议模块之间的接口将遵循统一的约定．这个约定允许把终端处理模块加入一个网络流中，产生有效的网络虚拟终端连接，并使核心能支持基于标准传输协议的远程过程协议．最后，这个接口将提供一个机制，把核心协议结构扩充至用户进程，使得可试制新的协议及实施网络控制功能．

1.4 发行工程

CSRG 一直是一个小规模的软件开发人员小组．有限的资源条件要求仔细的软件工程管理．精心的协调不仅对 CSRG 成员来说是必须的，而且对与系统开发有关的一般单位的成员也是需要的．主要的版本发行通常在主要新功能的发布（3BSD，4.0BSD，4.2BSD），和故障修复、效率改进（4.1BSD，4.3BSD）两者间交替．这样交替既使版本发行能及时，又能及时提供新功能的改进和校正，以及消除由于新功能而产生的性能上的问题．及时发行加入新功能的版本反映了 CSRG 的重点是提供一个用户可信赖的、可靠的、强壮的系统．

CSRG 研制工作的版本发行分三个步骤：alpha，beta，以及最后版，如表 1.1 所