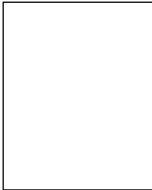


# The Linux Operating System

**S Balakrishnan**



**S Balakrishnan is a research scholar at the Supercomputer Education and Research Centre, Indian Institute of Science. His interests includedesign of special purpose computer architectures for digital signal processing and operating systems.**

**This article discusses the Linux Operating System – one of the most popular operating systems used on PCs available free-of-cost on the Internet.**

## Introduction

Linux is a free Unix-like operating system that has become popular with PC users around the world. An independent POSIX (see *Box 1*) implementation, Linux does true multitasking and includes virtual memory, shared libraries, demand loading, memory management, TCP/IP networking and other features that are available with current full featured commercial operating systems (see *Box 2*). A feature of the Linux operating system that distinguishes it from other operating systems is that it is one of the few operating systems whose source code is also available as free software under the GNU (see *Box 3*) General Public License (GPL). The GNU GPL is intended to safeguard and guarantee the freedom of any user of free software to share, modify and also share the modified software. This is in contrast to the licenses given for proprietary software that prohibits its users to share or modify software.

Unix? What's that?  
Is that like Linux?

### Box 1. The POSIX Standard

POSIX stands for Portable Operating System Interface for Computer Environments. POSIX was first adopted by the United States Federal Information Processing Standard (FIPS) in 1988 and revised in 1990. FIPS 151-1, POSIX, adopts the Institute of Electrical and Electronics Engineers (IEEE) Standard 1003.1-1988. The IEEE 1003.1-1988 Standard defines an interface to an operating system environment. The main aim of the FIPS 151-1, POSIX standard was to permit Federal Agencies in the United States of America to exercise control over the production, management and use of information resources by promoting the portability of applications at the source code level. This in turn is expected to reduce the costs of maintaining and transferring information technology applications.



### Box 2. Defining Terms

**Multi-tasking** is a method by which the operating system allows multiple programs to share a computer system to give an user the illusion that the programs are running simultaneously. Multi-tasking can be done either in a preemptive or in a cooperative manner. Preemptive multi-tasking is also known as true multitasking. In preemptive multi-tasking the operating system allocates an unit of time to each process to run. This unit of time is called a time-slice or quanta (typically 1/10th of a second). If a process uses up its time slice the operating system arranges to run a process with the highest priority to execute its time-slice next. In a priority based preemptive multi-tasking model of execution, interacting processes are given a higher priority over processes that have recently completed their time-slice. In cooperative multi-tasking an application has to voluntarily give up use of the processor to enable another process to run. The application periodically has to inform the operating system about its willingness to give up the processor by making system calls. The problem with cooperative multi-tasking is that an application can either by mistake or by malicious intent be made to monopolize use of the processor thereby stopping all other processes from running.

**Virtual Memory** is a scheme employed by the operating system to provide means of executing programs whose code occupy more space than the size of the on-board semiconductor memory. To achieve this, least recently used parts of a program are copied temporarily from memory into hard disk and copied back on demand.

**Shared Libraries** are used with dynamic linking to share commonly used routines. Each reference to a library routine is replaced with a stub that indicates how the appropriate routine can be located in memory. A stub initially executes and replaces itself with the address of the appropriate library routine. The next time around the same code segment is reached, the library routine is executed directly with no extra cost of locating the memory-resident portion of the library.

**Demand Loading** is a method of loading only parts of the program that is currently being executed into primary memory (RAM) from secondary memory (disk). The program is logically divided into segments or pages and is loaded into primary memory on demand, hence the term demand loading.

**Memory Management** is a scheme by which an operating system shares the memory in a computer system among several programs or several modules of the same program.

**TCP/IP:** TCP stands for Transport Cotnrol Protocol while IP stands for Internet Protocol. These are communication protocols used for transmitting data over a computer network.

## Linux History

In 1991 Linus Torvalds a student at the University of Helsinki, Finland got interested in a small operating system for PCs called



**Box 3. GNU is Not Unix!**

Conceived in 1983 to make cooperative development and use of software possible without the constraints imposed by the usage of proprietary software, the GNU project has developed a 'free software' system name 'GNU'. GNU is upwardly compatible with Unix. The term free software here stands for fact that users of the software have the *freedom* to 1) copy and give it away 2) change the program in any manner the user wishes, by having full access to the source code and 3) distribute an improved version of the sources or programs.

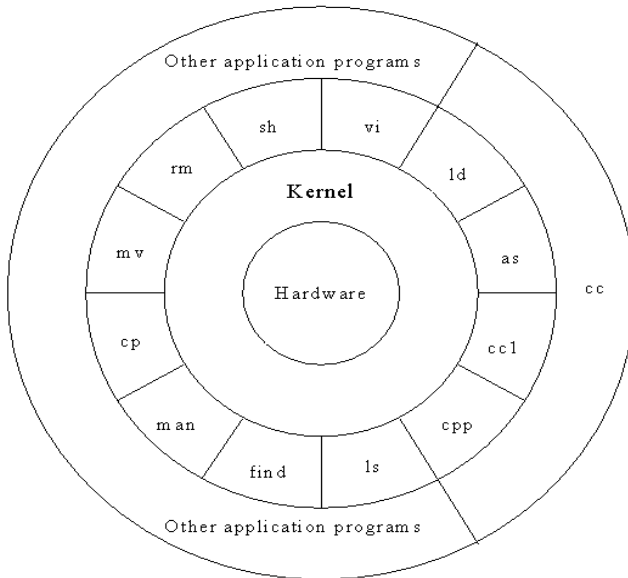
To use a computer system an user needs an operating system (see *Box 4*) that can run on it. Having free software for the operating system helps further the cause of cooperative use of software. GNU achieved the initial goal of a free Unix like operating system by the early 1990s except for the kernel. When Linux was developed and combined with other parts of the operating system already developed, a complete GNU operating system resulted.

Developing just operating systems is not the 'be all and end all' of the GNU project. The GNU project envisages to provide users any software that they wish to have without being fettered by the shackles of using proprietary software.

MINIX. MINIX was developed and written by Andrew S Tannenbaum of the Vrije Universiteit, Amsterdam, The Netherlands. MINIX as described in his book *Operating Systems: Design and Implementation*, was used for educational purposes and was written to elucidate the internals of UNIX-like operating systems. Linus Torvalds however decided to put in more functionality into the system than that promised by MINIX. He released version 0.02 of the operating system in 1991 and worked steadily on it till 1994 when he released version 1.0 of the Linux kernel (see *Box 4*). Programmers around the world then got together and decided to give a POSIX compliant UNIX-like system based on Linus's operating system to users around the world under the GNU GPL. The copyright for Linux is owned by Linus Torvalds and a few other contributors. The full-featured version of the Linux kernel available now is version 2.2. Further work is in progress to update the Linux source code and release it for public use.



### Box 4. UNIX Like Operating Systems



The figure depicts the architecture of a UNIX like operating system. An UNIX system can be considered to be composed of layers of software built around the central hardware that provides basic services to a user. The layer of software nearest to the hardware providing an abstraction of the hardware to the user, thereby insulating the user from hardware idiosyncrasies is commonly called the operating system or simply the *system kernel*. User programs are built around the system kernel and are independent of the underlying hardware. Application programs can also be built using a

combination of user programs and thus form a layer above user programs. A full-fledged UNIX system kernel includes basic support for multitasking, virtual memory, shared libraries, demand loading, memory management and TCP/IP networking.

UNIX kernels have been predominantly written using a high level programming language like C. The set of C programs that comprise the kernel is called the source code of the kernel.

### Linux Distributions

Linux *per se* refers only to the kernel of the operating system. A full fledged UNIX system however requires a set of applications and support software that make the system user friendly. Hundreds of programmers around the world have contributed to bring the Linux system to its current state of robustness. The Linux kernel along with a combination of these contributed utilities and applications forms a complete operating system. The Linux (see Box 5) operating system in any form or distribution is neither public domain nor is it shareware. Linux is free software. In this mode of distribution, any person or





### Box 5. The Linux Mascot

Linux has an official mascot, the Linux Penguin which was chosen by Linus Torvalds himself based on his impression on how it represented the operating system. According to Linus Torvalds the Linux Penguin is a contented, cute and cuddly creature. The Penguin is sitting down with a beatific smile because it is contented after having gorged itself with herrings.

organization can distribute or sell copies of the operating system but must also include the source code of the same in the same way the binaries were distributed or sold. This applies to every copy of even a modified distribution of Linux. Linux is called free software because of the fact that the source code is available to any user freely. This only means that persons or organizations can charge money for a distribution as long as they supply the source code too.

The Linux operating system is distributed in various user-friendly installation packages. Most of these distributions are available on CDROM and are easy to install. Most of the distributions are commercially available while others are available only through the Internet. Some of the distributions available on CDROM are Slackware, Redhat Linux, Caldera OpenLinux, Debian Linux, LinuxPro, Trans-Ameritech Linuxware and the S.u.S.E.

### The Linux Development Model

Linux is developed following what is referred to as the 'Open Development Model'. In this model all update versions are released to the public irrespective of whether the version is considered 'production quality' or not. However, to enable an user of the operating system to easily figure out whether he/she is working with a stable version or not, the following scheme is used: Linux versions are indicated using a number of the type *n.b.c*. *n* refers to the major version number. A Linux version in

"Only wimps use tape backup. *Real* men just upload their important stuff on ftp, and let the rest of the world mirror it."

— *Linus Torvalds*



## Linux: Choice of a GNU Generation

### *Photograph of Linus Benedict Torvalds*



which  $b$  is an even number is considered stable. Else it is considered a beta version which might be unstable and might crash and is intended only for developers. Linux versions that have only bug fixes applied with no major change in the kernel will only have the  $c$  portion of the version number incremented. The advantage of using an open development model for Linux is that the most current development version is always public. Whenever new functionality is added to the system and released, bugs are also detected and corrected quickly, leading to faster development of stable 'production quality' versions. This development model owes its success to the high speed at which information can be disseminated via the Internet. Thousands of programmers examine any new added functionality almost simultaneously and fix bugs, if any, rapidly.

### **Linux – Myths and Reality**

Linux has been an object of many myths that have arisen because of what is referred to as FUD by the Internet savvy. FUD stands for Fear, Uncertainty and Doubt. In this section we will try to dispel a few myths about Linux.

#### ***Myth: Linux is difficult to install***

**Truth:** This myth has taken root from the fact that early Unix systems were in general installed by experts and had very terse installation guide books. Linux, on the other hand, has been packaged in such a manner that even novices can install it from a CDROM. Almost all distributions of Linux have a character mode graphical user interface. There is complete interactive control over the packages installed on the system. Linux can also be installed over a network. Linux also comes with a host of application software like Netscape that are easy to configure and install.

#### ***Myth: Linux is hard to network***

**Truth:** Linux supports a wide range of networking protocols and standards. Linux supports AppleTalk of Mac, IPX of Novell,



Microsoft's Networking System and also TCP/IP. Armed with complete information about the networking hardware in use on a system one can network Linux in a time span ranging anywhere from two hours to a day.

Linux – "It's computing, Jim, but not as we know it".

***Myth: The Linux operating system is insecure***

**Truth:** This myth stems from the fact that Unix-like systems are reputedly insecure. In the case of Linux there is additional fear and doubt due to the fact that Linux is open source. The contention is that 'since the source code is freely available, people will easily find weaknesses in the system'. However, by the same argument security bugs are located by experts as soon as they appear and are fixed early. Other operating systems like Windows NT and Next Step rely on what is known as 'security through obscurity', to hide potential security problems from malicious hackers. Thus it is the hackers or crackers who discover security bugs and exploit them first before problems become known to all.

***Myth: Linux does not have office software.***

**Truth:** While Linux does not come with office software, there exist commercial software as well as some free office software for download through the Internet. Some commercial software that run on a Linux environment and have features very similar to MS Office are Star Office, Applix ware and WordPerfect. Due to the popularity of Linux, most major application vendors have come up with ports of their applications to Linux.

***Myth: Linux runs only on Intel processor based PCs***

**Truth:** Though Linux was originally created on a PC running MINIX and is popular on PCs, it can also be used on workstations using other processors. Linux has been ported to work on work-stations with the PowerPC, Sun's Sparc, Digital's Alpha and the Macintosh with the m68k processor to name a few. There is on-going effort to make Linux available for a wide range of platforms.



## Suggested Reading and Websites

- [1] Andrew S Tannenbaum and Albert S Woodhull, *Operating Systems: Design and Implementation*, Second Edition, Prentice Hall of India, 1997.
- [2] Mark G Sobel, *Hands-on Linux*, Addison Wesley Longman Publishers, 1997.
- [3] Mark G Sobel, *A Practical Guide to Linux*, Addison Wesley Longman Publishers, 1997.
- [4] *The Linux official web site*, <http://www.linux.org/>.
- [5] *The Redhat web site*, <http://www.redhat.com/>.
- [6] A feature from the Wired Magazine, <http://www.wired.com/wired/5.08/linux.html>
- [7] *The Cathedral and the Bazaar*, Eric. S. Raymond, <http://www.tuxedo.org/~esr/writings/cathedral-bazaar/>

## Linux – Miscellaneous Features

**Graphical User Interface:** Most distributions of the Linux operating system come with a windows environment called X. The X-windows environment is fully customizable and is fast. X comes with different window managers and customizable desk top features that make it very flexible and easy to use. Some of the window managers can give a look and feel of a MS windows environment.

**Java:** The Java platform is based on the concept that any software should run on any kind of computer, consumer gadget or any other device. This means that Java software is operating system and machine independent and works on any compatible device that supports the Java platform. To run Java applications efficiently, the operating system should be able to support the Java platform. The Linux operating system has support to run Java applications with kernel integration to the interpreter, compilers, and applets.

**File Systems:** Linux can recognize a wide range of file system formats. This means that, given for example, a system on which you have the option of running either Windows NT or Linux, we can look up files in the Windows NT partition without restarting the system, while running Linux. Some of the standard file systems that are supported are 1) A native Linux advanced extended file system 2) The MINIX file system 3) The FAT file system of MSDOS 4) The ISO9660 file system used on CDROMs and 5) th VFAT file system used in Windows NT.

**Linux Documents:** The Linux documentation Project's (LDP) goal is to develop good reliable documentation for the Linux operating system. LDP takes care of documentation ranging from online documentation to printed manuals for installing, using and running Linux. All documentation pertaining to Linux and produced by LDP are available freely and distributed over the Internet.



### Box 6. The Y2K Problem

The year 2000 problem commonly referred to as the Y2K problem is due to the fact that early computer hardware and software tended to use two digit year fields. Therefore, computers that use hardware and software using two digit date fields will register the year 2000 as 00. This means that calculations that try to find the difference between two dates spanning the 20th and the 21st century will land up with negative numbers.

The Y2K problem can strike at six levels of PC hardware and software, the RTC (real-time clock), the BIOS, operating system, Network operating system application software and interfaces with other systems like databases that require date calculations.

**Y2K Compliance:** The Linux operating system is Y2K (see *Box 6*) compliant. Date on an Unix system is calculated in terms of seconds elapsed since 0000 hours UTC, January 1, 1970. Linux uses a 32 bit signed representation for this value. This means that this counter will overflow on January 19, 2038, when the number of seconds reaches 2,147,483,647, the maximum number that can be represented in a 32-bit signed integer. This problem may automatically be solved when Linux moves on to using a 64 bit representation for this number.

### Concluding Remarks

Recent surveys on the use of computer systems and customer's preferences for one operating system over the other indicate that quite a few people are moving from the Windows NT Operating system to Linux. Trends as published in the Internet, indicate that engineers working long hours at a computer keyboard have switched over to Linux from NT because of the latter's need for constant rebooting. Some users switch to Linux in their quest for cost-effective computation. For example a team of graphics artists behind the simulation of the movie *Titanic* used a bunch of machines running the Linux operating system as a compute farm. A Linux desktop can be configured to look like Windows and also has application suites that are functionality equivalent to Microsoft Office. Another point in favor of Linux is that, since the source code for Linux is widely available, work-around for hardware defects are reported and patched into the kernel almost overnight.

*Address for Correspondence*  
S Balakrishnan  
Research Scholar  
Supercomputer Education  
and Research Centre,  
Indian Institute of Science,  
Bangalore 560 012.  
Email: sbalki@serc.iisc.ernet.in

