SNELL & WILMER L.L.P.
Alan L. Sullivan (3152)
Todd M. Shaughnessy (6651)
Amy F. Sorenson (8947)
15 West South Temple, Suite 1200
Salt Lake City, Utah 84101-1004
Telephone: (801) 257-1900
Facsimile: (801) 257-1800

CRAVATH, SWAINE & MOORE LLP
Evan R. Chesler (admitted pro hac vice)
David R. Marriott (7572)
Worldwide Plaza
825 Eighth Avenue
New York, New York 10019
Telephone: (212) 474-1000
Facsimile: (212) 474-3700

Attorneys for Defendant/Counterclaim-Plaintiff International Business Machines Corporation

IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF UTAH

THE SCO GROUP, INC.

Plaintiff/Counterclaim-Defendant,

v.

INTERNATIONAL BUSINESS MACHINES CORPORATION,

Defendant/Counterclaim-Plaintiff.

UNSEALED DECLARATION OF BRIAN W. KERNIGHAN [Docket No. 252]

Civil No. 2:03CV0294 DAK

Honorable Dale A. Kimball

Magistrate Judge Brooke C. Wells

SNELL & WILMER, L.L.P.
Alan L. Sullivan (3152)
Todd M. Shaughnessy (6651)
Amy F. Sorenson (8947)
15 West South Temple, Suite 1200
Salt Lake City, Utah 84101-1004
Telephone: (801) 257-1900
Facsimile: (801) 257-1800

CRAVATH, SWAINE & MOORE LLP
Evan R. Chesler (admitted pro hac vice)
David R. Marriott (7572)
Worldwide Plaza
825 Eighth Avenue
New York, New York 10019
Telephone: (212) 474-1000
Facsimile: (212) 474-3700

Attorneys for Defendant/Counterclaim-Plaintiff International Business Machines Corporation

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF UTAH

THE SCO GROUP, INC. Plaintiff/Counterclaim-Defendant, vs. INTERNATIONAL BUSINESS MACHINES CORPORATION, Defendant/Counterclaim-Plaintiff.	**DECLARATION OF BRIAN W. KERNIGHAN** Civil No. 2:03CV0294 DAK Honorable Dale A. Kimball Magistrate Judge Brooke Wells

FILED UNDER SEAL

I. Introduction

1. I have been asked by counsel for IBM to evaluate the opinions set out in the declaration of Sandeep Gupta submitted in opposition to IBM's motion for partial summary judgment of non-infringement with respect to its Linux activities. Specifically, I have been asked to address Mr. Gupta's conclusion that there is "substantial similarity" between certain "routines" and "groupings of code" in Linux and copyrighted works allegedly owned by SCO.

2. My analysis and conclusions are based upon the principles described in *Gates Rubber v. Bando*, 9 F.3d 823 (10th Cir. 1993), which I understand to describe the appropriate methodology for determining substantial similarity. My analysis and conclusions are also based upon my experience and expertise in the field of computer science.

3. In summary, I find fundamental errors in Mr. Gupta's conclusions. His conclusions of substantial similarity are flawed because he fails to exclude from comparison unprotectable elements of the allegedly copyrighted code, and he uses an indefensible standard for what qualifies as "substantially similar" code.

4. If unprotectable elements are excluded from the comparison and an appropriate standard of similarity is applied, there is no similarity between the parts of Linux identified by Mr. Gupta and the allegedly copyrighted works.

5. Errors in methodology aside, the code Mr. Gupta identifies as infringing does not constitute a substantial part of the allegedly copyrighted works. SCO purports to hold copyrights in at least several versions of Unix System V source code, which over the years have ranged in size from several hundred thousand lines of code (in early versions)

to many millions of lines of code in a current version. The alleged similarities identified by Mr. Gupta amount to less than three hundred lines of code, none of which is qualitatively or quantitatively significant.

6. The next section describes my background and qualifications to address the issues addressed herein. The following section describes the flaws in Mr. Gupta's conclusions and explains why, when the segments of material in Linux to which SCO claims rights are compared to protectable elements of the allegedly copyrighted works, there is no similarity between them.

II. Background and Qualifications

7. I am a professor of Computer Science at Princeton University in Princeton, New Jersey. Exhibit I provides more details of my technical background and experience. I received my undergraduate degree in Engineering Physics from the University of Toronto in 1964, and a PhD in Electrical Engineering from Princeton in 1969.

8. From 1969 to 2000 I was a member of technical staff in the Computing Science Research Center at Bell Labs in Murray Hill, New Jersey, where I worked on programming languages, software tools, programming methodology, and other areas, in the group that developed the Unix operating system, the C and C++ programming languages, and numerous other software systems. My personal research included the development of a number of software systems, some of which are in widespread use today, such as the AWK programming language, the AMPL modeling language, and a variety of document preparation tools. From 1981 to 2000, I was head of the Computing Structures Research department. I was named a Bell Labs Fellow in 1996.

9. I am the co-author of eight books on programming, which have been translated altogether into 24 languages, and over 60 technical articles, and hold four software patents. Together with Dennis Ritchie (one of the two inventors of Unix), I wrote *The C Programming Language* in 1978, which was the first book describing C, and which has since become a standard text on the language. I am also the co-author, with Rob Pike, of *The Unix Programming Environment* and *The Practice of Programming*.

10. I was an advisor to Prentice-Hall for their software series from 1978 to 1990, and have been the advisor for Addison-Wesley's Professional Computing series since 1990. I am an editor for the journal *Software Practice and Engineering*. From 1993 to 1995, I was a member of the computer science grant selection committee for the National Science and Engineering Research Council (Canada). I have served on advisory committees at Stanford University, Carnegie-Mellon University, and Princeton.

11. From 1996 to 1999, I was a member of the National Research Council's Panel for Information Technology, and was co-chair from 1997 to 1999. This panel reviews and assesses the work of the information technology laboratory at the National Institute of Standards and Technology (NIST), in the areas of computing, mathematics, software testing, networking, security, and the like.

12. I am a member of the National Academy of Engineering ("NAE"), to which I was elected in 2002, and I am currently a member of the NAE Peer Committee for the Computer Science and Engineering section.

13. I have not testified as an expert in a trial or deposition in the last four years.

14. I have been retained by counsel for IBM in this lawsuit and am being compensated at a rate of $400 per hour.

III. Analysis of the Gupta Declaration

15. Mr. Gupta concludes that there are six categories of "routines" or "groupings of code" in Linux that are substantially similar to the allegedly copyrighted works: (1) the Read-Copy-Update ("RCU") routine; (2) the user level synchronization ("ULS") routines; (3) IPC code; (4) certain "headers and interfaces"; (5) UNIX System V init code; and (6) Executable and Linking Format ("ELF") code (¶ 3). Mr. Gupta's conclusions regarding these six categories of routines or groupings of code appear to form the basis of SCO's contention that IBM's Linux activities infringe SCO's alleged copyrights.

16. In *Gates Rubber*, the Tenth Circuit sets out a three-part test, known as the abstraction-filtration-comparison test, for determining whether a computer program (such as Linux) is substantially similar to and therefore may infringe the copyright in another computer program (such as Unix). 9 F.3d at 834-39. While flexible, the test ordinarily contemplates that a court (1) "dissect the program according to its varying levels of generality", (2) "filter out those elements of the program which are unprotectable" at each level of abstraction; and (3) "compare the remaining protectable elements with the allegedly infringing program" *Id.* at 834.

17. Mr. Gupta does not describe the methodology by which he arrives at his conclusions. Even so, it is clear that Mr. Gupta's methodology, and therefore his conclusions, are indefensible. Mr. Gupta does not appear to have dissected the allegedly copyrighted

works according to their varying levels of generality, but the real problem with his analysis is that he fails to filter unprotectable elements before comparing the works at issue and he employs a mistaken standard of similarity.

18. As is explained in *Gates Rubber*, the unprotectable elements of an allegedly copyrighted work must be filtered out of the analysis before an allegedly infringing work may be found to be substantially similar. *9 F.3d* at 838-39. "Filtration should eliminate from the comparison the unprotectable elements of ideas, processes, facts, public domain, information, merger material, scènes à faire material, and other unprotectable elements suggested by the particular facts of the program under examination." *Id.* at 834. To the extent there are similarities between the routines and groupings of code identified by Mr. Gupta, they concern unprotectable elements. They lack originality or constitute ideas, procedures, processes, systems, methods of operation, concepts, principles or discoveries, public domain information, merger material, and/or scènes à faire material.

19. Moreover, substantial similarity may be found, according to the Tenth Circuit, only where "those protectable portions of the original work that have been copied constitute a substantial part of the original work -- *i.e.*, matter that is significant in the plaintiffs' program". Gates Rubber, 9 F.3d. at 839. Putting aside Mr. Gupta's failure to filter unprotectable elements, much of what Mr. Gupta is willing to call "substantially similar," or even "identical," plainly is not, even to an untrained eye. In some instances, code that Mr. Gupta discusses as if it were a single block is in fact his own composite of code scattered throughout a large body of code, sometimes even coming from different files.

Finally, the amount of code at issue is an insignificant portion of the whole — not only quantitatively but also qualitatively.

20. Even without employing all potential filters, I have no difficulty concluding that all of the similarities identified by Mr. Gupta either should have been filtered as unprotectable or are simply not similar, much less "identical".

21. Two allegedly copied regions (described by Mr. Gupta as the "RCU routine" and the "ULS routines") are unprotectable ideas or processes expressed at a very high level of abstraction, not protectable expression. Mr. Gupta's own language makes the point, labeling the material at issue a "routine" (¶¶ 3, 5, 10) and "method" (¶¶ 6, 7). He also describes the routines at issue as "perform[ing] the same five acts" (¶ 11) and "includ[ing] the same nine (9) acts" (¶ 36), but each of these acts is generic, far removed from a specific expression. Moreover, the actual code that Mr. Gupta says is "substantially similar" is not similar at all. Mr. Gupta's own Exhibits A, H and I illustrate the point—even to a non-programmer, the Linux code is completely different from the Unix code—and detailed examination of the code shows no relationship.

22. Another two items identified by Mr. Gupta (the IPC header files and the ELF header file) are no more protectable. The IPC material is in the public domain, because, as I understand it, it was included in Unix System V 2.0 and Unix System V 3.2 without copyright notices and distributed before 1989. Moreover, to the extent it is even original, the IPC material constitutes ideas, concepts or principles, merger material and scènes à faire material. Furthermore, in places, Mr. Gupta's conclusions of similarity depend on his selecting isolated lines of code from disparate places and putting them

together as if contiguous blocks of code were involved (which they are not) and important differences did not exist (which they do).

23. The ELF header file is classic scènes à faire material. The contents of the file are determined by software standards, target industry practice and demands and computer industry programming practices. Indeed, the substantive content in the ELF header file comes from a published and widely distributed standard — the "Tool Interface Standard (TIS) Executable and Linking Format (ELF) Specification", version 1.2, attached as Exhibit XVI— material that explicitly grants permission for use in the interests of interoperability, by an industry consortium that included SCO's alleged predecessor-in-interest, The Santa Cruz Operation, Inc.

24. Mr. Gupta's claims for "certain UNIX System V headers and interfaces" (¶¶ 63-72) and "SYS V init code" (¶¶ 73-76) concern code that is not found in the Linux kernel.[1] That code is therefore outside what I understand to be the scope of the IBM Counterclaim at issue. Mr. Gupta's claims as to this code nevertheless suffer from the defects described above: a failure to filter non-protectable material and an inappropriate standard of similarity.

25. Each of these six regions involves only elements that are unprotectable and/or dissimilar, as is demonstrated in detail in Exhibits II through VII.

[1] The "kernel" is the fundamental part of an operating system. It is the piece of software responsible for providing secure access to the machine's hardware. Since there are many programs, and access to the hardware is limited, the kernel is also responsible for deciding when and how long a program should be able to make use of a piece of hardware. [from WordIQ Dictionary, www.wordiq.com]

26. Putting aside the errors in Mr. Gupta's methodology, the code he identifies as infringing does not constitute a substantial part of the allegedly copyrighted works. SCO purports to hold copyrights in several versions of Unix System V source code, which over the years have ranged in size from several hundred thousand lines of code (in early versions) to many millions of lines of code in a current version. The alleged similarities identified by Mr. Gupta amount to a total of less than three hundred lines.

27. The allegedly infringed portions of the allegedly copyrighted works are not significant or important parts of the allegedly copyrighted works, considered as a whole — either quantitatively or qualitatively.

V. Conclusions

28. Mr. Gupta's declaration contains errors of methodology and fact. His conclusions of substantial similarity are flawed because he fails to exclude from comparison unprotectable elements of the allegedly copyrighted code and he uses an indefensible standard for what qualifies as "substantially similar" code.

29. If unprotectable elements are excluded from the comparison and an appropriate standard of similarity is applied there is no similarity between the six parts of Linux identified by Mr. Gupta and the allegedly copyrighted works.

30. Errors in methodology aside, the code Mr. Gupta alleges to be infringing does not constitute a substantial part of the allegedly copyrighted works, amounting to less than three hundred lines, which pales in comparison to the millions of lines in a current version of Unix.

31. I declare under penalty of perjury that the foregoing is true and correct.

Brian W. Kernighan

8/20/04

Date

Princeton, NJ

Place

Materials Considered During Preparation of this Declaration

Computer Associates v. Altai, 982 F.2d 693 (2d Cir. 1992)

Gates Rubber, Inc., v. Bando American, Inc., 9 F.3d 823 (10th Cir. 1993)

Mitel, Inc. v. Iqtel, Inc. 124 F.3d 1366 (10th Cir. 1997)

System Calls (Solaris 2.4 manual section 2). SUN Microsystems, Inc., 1994.

The Single UNIX Specification, Version 2. The Open Group, 1997.
`http://www.opengroup.org/onlinepubs/007908799/toc.htm.`

The Single UNIX Specification, Version 3. The Open Group, 2004.
`http://www.opengroup.org/onlinepubs/009695399/nframe.html.`

Maurice J. Bach, *The Design of the UNIX Operating System*. Prentice-Hall, 1986.

W. Richard Stevens, *Advanced Programming in the UNIX Environment*. Addison-Wesley, 1992.

Éric Lévénez, *Unix Timeline*. `http://www.levenez.com/unix.`

Code cited by Gupta for both Unix and Linux

All exhibits to Gupta declaration

Web site lxr.linux.no (Linux Cross-reference) for Linux source code

Code produced by SCO for Unix source code

EXHIBIT 1

BRIAN W. KERNIGHAN

Brian Kernighan received his undergraduate degree in Engineering Physics from the University of Toronto in 1964, and his PhD in Electrical Engineering from Princeton University in 1969.

In 1969, he joined Bell Laboratories in Murray Hill, New Jersey, in the Computing Science Research Center, the group that produced such fundamental computer systems as the Unix operating system and the C programming language. His research areas included programming languages, software tools, computer-aided design, document preparation systems, and other systems for use within AT&T. From 1981 to 2000, he was head of the Computing Structures Research Department, where in addition to his own research he managed a group of individual PhD researchers. He was named a Bell Labs Fellow in 1996.

Upon retirement from Bell Labs in 2000, he joined the Department of Computer Science at Princeton University as a full professor.

Dr. Kernighan has published over 60 technical papers, is the co-author of eight widely used books on programming, including *The C Programming Language* and *The Unix Programming Environment*, and holds four software patents.

Dr. Kernighan has been a member of the editorial board of the journal *Software Practice & Experience* for many years. From 1978 to 1990, he was the advisor for Prentice-Hall's series on computing, and since 1990 has been the advisor for Addison-Wesley's Professional Computing series. He has served on computer advisory committees at Princeton, Carnegie Mellon, and Stanford. From 1996 to 1999, he was a member of the National Research Council's Panel for Information Technology, which assesses the National Institute of Standards and Technology (NIST) in computing, mathematics, networks, security, etc., and was co-chair in 1997-1999. He is currently a member of the Computer Science and Engineering Peer Committee of the National Academy of Engineering. He has also been a consultant for Bell Labs and Google.

Dr. Kernighan has won a number of awards for research and teaching, including the Usenix Association Lifetime Achievement Award in 1997 and the Princeton University 250th Anniversary Visiting Professorship for Distinguished Teaching for 1999-2000. He was elected a member of the National Academy of Engineering in 2002.

Bibliography for Brian W. Kernighan for Last Ten Years

1. R. Fourer, D. M. Gay and B. W. Kernighan, An Introduction to the AMPL Modeling Language for Mathematical Programming, *Mathematech*, Vol. 1, 1, pp. 49-56, Spring 1994.
2. S. J. Fortune, D. M. Gay, B. W. Kernighan, O. Landron, R. A. Valenzuela and M. H. Wright, WISE Design of Indoor Wireless Systems: Practical Computation and Optimization, *IEEE Computational Science and Engineering*, Vol. 2, 1, pp. 58-68, Spring 1995.
3. B. W. Kernighan, Experience with Tcl/Tk for Scientific and Engineering Visualization, Proc. Tcl/Tk Workshop 95, pp. 269-278, Toronto Ontario, July 1995.
4. S. J. Fortune, D. M. Gay, B. W. Kernighan, O. Landron, R. A. Valenzuela and M. H. Wright, *Prediction Of Indoor Electromagnetic Wave Propagation For Wireless Indoor Systems*, US Patent 5 450 615, September 12 1995.
5. B. W. Kernighan and C. J. Van Wyk, Extracting Geometric Information from Architectural Drawings, Workshop on Applied Computational Geometry, pp. 167-176, Philadelphia PA, May 27-28 1996. (Reprinted in Springer *Lecture Notes in Computer Science*, Vol 1148).
6. B. W. Kernighan and C. J. Van Wyk, Timing Trials, Or, the Trials of Timing: Experiments with Scripting and User-interface Languages, *Software Practice and Experience*, Vol. 28, 8, pp. 819-843, July 1998.
7. B. W. Kernighan and R. Pike, *The Practice of Programming*, Addison-Wesley, 1999.
8. B. W. Kernighan and R. Pike, Regular Expressions: Languages Algorithms and Software, *Dr. Dobbs Journal*, pp. 19-22, April 1999 (Excerpted from *The Practice of Programming*).
9. B. W. Kernighan and R. Pike, Finding Performance Improvements, *IEEE Software*, pp. 61-65, March April 1999 (Excerpted from *The Practice of Programming*).
10. B. S. Baker, K. F. Church, J. F. Helfman and B. W. Kernighan, *Methods and Apparatus for Detecting and Displaying Similarities in Large Data Sets*, US Patent 5 953 006, September 14 1999.
11. R. Fourer, D. M. Gay and B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*, Second edition. Duxbury, 2002.
12. R. Fourer, D. M. Gay and B. W. Kernighan, Design Principles and New Developments in the AMPL Modeling Language, in *Modeling Languages in Mathematical Optimization*, Josef Kallrath, ed., Kluwer, 2004.

EXHIBIT 2

DOCUMENT MAINTAINED UNDER SEAL
AT SCO'S REQUEST

EXHIBIT 3

DOCUMENT MAINTAINED UNDER SEAL
AT SCO'S REQUEST

EXHIBIT 4

DOCUMENT MAINTAINED UNDER SEAL
AT SCO'S REQUEST

EXHIBIT 5

DOCUMENT MAINTAINED UNDER SEAL
AT SCO'S REQUEST

EXHIBIT 6

DOCUMENT MAINTAINED UNDER SEAL
AT SCO'S REQUEST

EXHIBIT 7

DOCUMENT MAINTAINED UNDER SEAL
AT SCO'S REQUEST

EXHIBIT 8

NAME

sys/ipc.h - interprocess communication access structure

SYNOPSIS

```
#include <sys/ipc.h>
```

DESCRIPTION

The *<sys/ipc.h>* header is used by three mechanisms for interprocess communication (IPC): messages, semaphores and shared memory. All use a common structure type, **ipc_perm** to pass information used in determining permission to perform an IPC operation.

The structure **ipc_perm** contains the following members:

```
uid_t    uid    owner's user ID
gid_t    gid    owner's group ID
uid_t    cuid   creator's user ID
gid_t    cgid   creator's group ID
mode_t   mode   read/write permission
```

The **uid_t**, **gid_t**, **mode_t** and **key_t** types are defined as described in *<sys/types.h>*.
Definitions are given for the following constants:
Mode bits:
IPC_CREAT
Create entry if key does not exist.
IPC_EXCL
Fail if key exists.
IPC_NOWAIT
Error if request must wait.
Keys:
IPC_PRIVATE
Private key.
Control commands:
IPC_RMID
Remove identifier.
IPC_SET
Set options.
IPC_STAT
Get options.
The following is declared as a function and may also be defined as a macro. Function prototypes must be provided for use with an ISO C compiler.

```
key_t  ftok(const char *, int);
```

APPLICATION USAGE

None.

FUTURE DIRECTIONS

None.

SEE ALSO

ftok(), *<sys/types.h>*.

NAME

sys/msg.h - message queue structures

SYNOPSIS

```
#include <sys/msg.h>
```

DESCRIPTION

The <**sys/msg.h**> header defines the following constant and members of the structure **msqid_ds**.

The following data types are defined through **typedef**:

msgqnum_t

Used for the number of messages in the message queue.

msglen_t

Used for the number of bytes allowed in a message queue.

These types are unsigned integer types that are able to store values at least as large as a type **unsigned short**.

Message operation flag:

MSG_NOERROR

No error if big message.

The structure **msqid_ds** contains the following members:

```
struct ipc_perm msg_perm   operation permission structure
msgqnum_t       msg_qnum   number of messages currently on queue
msglen_t        msg_qbytes maximum number of bytes allowed on
queue
pid_t           msg_lspid  process ID of last msgsnd()
pid_t           msg_lrpid  process ID of last msgrcv()
time_t          msg_stime  time of last msgsnd()
time_t          msg_rtime  time of last msgrcv()
time_t          msg_ctime  time of last change
```

The **pid_t**, **time_t**, **key_t** and **size_t** types are defined as described in *<sys/types.h>*.

The following are declared as functions and may also be defined as macros. Function prototypes must be provided for use with an ISO C compiler.

```
int       msgctl(int, int, struct msqid_ds *);
int       msgget(key_t, int);
ssize_t   msgrcv(int, void *, size_t, long int, int);
int       msgsnd(int, const void *, size_t, int);
```

In addition, all of the symbols from *<sys/ipc.h>* will be defined when this header

is included.

APPLICATION USAGE

None.

FUTURE DIRECTIONS

None.

SEE ALSO

msgctl(), *msgget()*, *msgrcv()*, *msgsnd()*, *<sys/types.h>*.

NAME

sys/sem.h - semaphore facility

SYNOPSIS

```
#include <sys/sem.h>
```

DESCRIPTION

The <**sys/sem.h**> header defines the following constants and structures.

Semaphore operation flags:
SEM_UNDO
Set up adjust on exit entry.
Command definitions for the function *semctl()*:
GETNCNT
Get **semncnt**.
GETPID
Get **sempid**.
GETVAL
Get **semval**.
GETALL
Get all cases of **semval**.
GETZCNT
Get **semzcnt**.
SETVAL
Set **semval**.
SETALL
Set all cases of **semval**.
The structure **semid_ds** contains the following members:

```
struct ipc_perm    sem_perm  operation permission structure
unsigned short int sem_nsems number of semaphores in set
time_t             sem_otime last semop^) time
time_t             sem_ctime last time changed by semctl()
```

The **pid_t**, **time_t**, **key_t** and **size_t** types are defined as described in *<sys/types.h>*.
A semaphore is represented by an anonymous structure containing the following members:

```
unsigned short int semval    semaphore value
pid_t              sempid    process ID of last operation
unsigned short int semncnt   number of processes waiting for
```

```
semval
                                to become greater than current value
unsigned short int semzcnt      number of processes waiting for
semval
                                to become 0
```

The structure **sembuf** contains the following members:

```
unsigned short int sem_num   semaphore number
short int          sem_op    semaphore operation
short int          sem_flg   operation flags
```

The following are declared as functions and may also be defined as macros. Function prototypes must be provided for use with an ISO C compiler.

```
int    semctl(int, int, int, ...);
int    semget(key_t, int, int);
int    semop(int, struct sembuf *, size_t);
```

In addition, all of the symbols from *<sys/ipc.h>* will be defined when this header is included.

APPLICATION USAGE

None.

FUTURE DIRECTIONS

None.

SEE ALSO

semctl(), *semget()*, *semop()*, *<sys/types.h>*.

NAME

sys/shm.h - shared memory facility

SYNOPSIS

```
#include <sys/shm.h>
```

DESCRIPTION

The <**sys/shm.h**> header defines the following symbolic constants and structure:

Symbolic constants:

SHM_RDONLY

Attach read-only (else read-write).

SHMLBA

Segment low boundary address multiple.

SHM_RND

Round attach address to SHMLBA.

The following data types are defined through **typedef**:

shmatt_t

Unsigned integer used for the number of current attaches that must be able to store values at least as large as a type **unsigned short**.

The structure **shmid_ds** contains the following members:

```
struct ipc_perm shm_perm   operation permission structure
size_t          shm_segsz  size of segment in bytes
pid_t           shm_lpid   process ID of last shared memory operation
pid_t           shm_cpid   process ID of creator
shmatt_t        shm_nattch number of current attaches
time_t          shm_atime  time of last shmat()
time_t          shm_dtime  time of last shmdt()
time_t          shm_ctime  time of last change by shmctl()
```

The **pid_t**, **time_t**, **key_t** and **size_t** types are defined as described in *<sys/types.h>*. The following are declared as functions and may also be defined as macros. Function prototypes must be provided for use with an ISO C compiler.

```
void *shmat(int, const void *, int);
int   shmctl(int, int, struct shmid_ds *);
int   shmdt(const void *);
int   shmget(key_t, size_t, int);
```

In addition, all of the symbols from *<sys/ipc.h>* will be defined when this header is included.

APPLICATION USAGE

None.

FUTURE DIRECTIONS

None.

SEE ALSO

shmat(), *shmctl()*, *shmdt()*, *shmget()*, *<sys/types.h>*.

EXHIBIT 9

When a process is done with a shared resource that is controlled by a semaphore, the semaphore value is incremented by 1. If any other processes are asleep, waiting for the semaphore, they are awakened.

To implement semaphores correctly, the test of a semaphore's value and the decrementing of this value must be an atomic operation. For this reason, semaphores are normally implemented inside the kernel.

A common form of semaphore is called a *binary semaphore*. It controls a single resource and its value is initialized to 1. In general, however, a semaphore can be initialized to any positive value, with the value indicating how many of units of the shared resource are available for sharing.

System V semaphores are, unfortunately, more complicated than this. Three features contribute to this unnecessary complication.

1. A semaphore is not just a single nonnegative value. Instead we have to define a semaphore as a set of one or more semaphore values. When we create a semaphore we specify the number of values in the set.
2. The creation of a semaphore (`semget`) is independent of its initialization (`semctl`). This is a fatal flaw, since we cannot atomically create a new semaphore set and initialize all the values in the set.
3. Since all forms of System V IPC remain in existence even when no process is using them, we have to worry about a program that terminates without releasing the semaphores it has been allocated. The "undo" feature that we describe later is supposed to handle this.

The kernel maintains a `semid_ds` structure for each semaphore.

```
struct semid_ds {
  struct ipc_perm  sem_perm;  /* see Section 14.6.2 */
  struct sem  *sem_base;  /* ptr to first semaphore in set */
  ushort       sem_nsems; /* # of semaphores in set */
  time_t       sem_otime; /* last-semop() time */
  time_t       sem_ctime; /* last-change time */
};
```

The `sem_base` pointer is worthless to a user process, since it points to memory in the kernel. What it points to is an array of `sem` structures, containing `sem_nsems` elements, one element in the array for each semaphore value in the set.

```
struct sem {
  ushort  semval;   /* semaphore value, always >= 0 */
  pid_t   sempid;   /* pid for last operation */
  ushort  semncnt;  /* # processes awaiting semval > currval */
  ushort  semzcnt;  /* # processes awaiting semval = 0 */
};
```

Figure 14.18 lists the system limits (Section 14.6.3) that affect semaphore sets.

EXHIBIT 10

docs.sun.com - Sun Product Documentation

NAME

Intro, intro - introduction to system calls and error numbers

SYNOPSIS

#include <errno.h>

DESCRIPTION

[material below extracted from 19 pages of description]

Semaphore Identifier

A semaphore identifier (**semid**) is a unique positive integer created by a **semget** system call. Each semid has a set of semaphores and a data structure associated with it. The data structure is referred to as **semid_ds** and contains the following members:

```
struct   ipc_perm sem_perm;    /* operation permission struct */
struct   sem * sem_base;       /* ptr to first semaphore in set */
ushort   sem_nsems;            /* number of sems in set */
time_t   sem_otime;            /* last operation time */
time_t   sem_ctime;            /* last change time */
                               /* Times measured in secs since */
                               /* 00:00:00 GMT, Jan. 1, 1970 */
```

Here are descriptions of the fields of the **semid_ds** structure:

sem_perm is an **ipc_perm** structure that specifies the semaphore operation permission (see below). This structure includes the following members:

```
uid_t    uid;    /* user id */
gid_t    gid;    /* group id */
uid_t    cuid;   /* creator user id */
gid_t    cgid;   /* creator group id */
mode_t   mode;   /* r/a permission */
ulong    seq;    /* slot usage sequence number */
key_t    key;    /* key */
```

sem_nsems is equal to the number of semaphores in the set. Each semaphore in the set is referenced by a nonnegative integer referred to as a **sem_num**. **sem_num** values run sequentially from 0 to the value of **sem_nsems** minus 1.

sem_otime is the time of the last **semop** operation.

sem_ctime is the time of the last **semctl** operation that changed a member of the above structure.

A semaphore is a data structure called **sem** that contains the following members:

```
ushort   semval;    /* semaphore value */
pid_t    sempid;    /* pid of last operation */
ushort   semncnt;   /* # awaiting semval > cval */
ushort   semzcnt;   /* # awaiting semval = 0 */
```

semval is a non-negative integer that is the actual value of the semaphore.

sempid is equal to the process ID of the last process that performed a semaphore operation on this semaphore.

semncnt is a count of the number of processes that are currently suspended awaiting this semaphore's semval to become greater than its current value.

semzcnt is a count of the number of processes that are currently suspended awaiting this semaphore's semval to become 0.

EXHIBIT 11

Name	Description	Typical Value
SEMVMX	The maximum value of any semaphore.	32,767
SEMAEM	The maximum value of any semaphore's adjust-on-exit value.	16,384
SEMMNI	The maximum number of semaphore sets, systemwide.	10
SEMMNS	The maximum number of semaphores, systemwide.	60
SEMMSL	The maximum number of semaphores per semaphore set.	25
SEMMNU	The maximum number of undo structures, systemwide.	30
SEMUME	The maximum number of undo entries per undo structures.	10
SEMOPM	The maximum number of operations per semop call.	10

Figure 14.18 System limits that affect semaphores.

The first function to call is `semget` *to obtain a semaphore ID.*

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/sem.h>

int semget(key_t key, int nsems, int flag);
```

Returns: semaphore ID if OK, −1 on error

In Section 14.6.1 we described the rules for converting the *key* into an identifier and discussed whether a new set is created or an existing set is referenced. When a new set is created the following members of the `semid_ds` structure are initialized.

- The `ipc_perm` structure is initialized as described in Section 14.6.2. The `mode` member of this structure is set to the corresponding permission bits of *flag*. These permissions are specified with the constants from Figure 14.14.
- `sem_otime` is set to 0.
- `sem_ctime` is set to the current time.
- `sem_nsems` is set to *nsems*.

nsems is the number of semaphores in the set. If a new set is being created (typically in the server) we must specify *nsems*. If we are referencing an existing set (a client) we can specify *nsems* as 0.

The `semctl` function is the catchall for various semaphore operations.

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/sem.h>

int semctl(int semid, int semnum, int cmd, union semun arg);
```

Returns: (see following)

EXHIBIT 12

function first. As described previously, we don't consider message queues connectionless, since some technique is required to obtain the identifier for a queue. Since all these forms of IPC are restricted to a single host, all are reliable. When the messages are sent across a network, the possibility of messages being lost becomes a concern. Flow control means that the sender is put to sleep if there is a shortage of system resources (buffers) or if the receiver can't accept any more messages. When the flow control condition subsides, the sender should automatically be awakened.

One feature that we don't show in Figure 14.15 is whether the IPC facility can automatically create a unique connection to a server for each client. We'll see in Chapter 15 that streams and Unix stream sockets provide this capability.

The next three sections describe each of the three forms of System V IPC in detail.

14.7 Message Queues

Message queues are a linked list of messages stored within the kernel and identified by a message queue identifier. We'll call the message queue just a "queue" and its identifier just a "queue ID." A new queue is created, or an existing queue is opened by `msgget`. New messages are added to the end of a queue by `msgsnd`. Every message has a positive long integer type field, a nonnegative length, and the actual data bytes (corresponding to the length), all of which are specified to `msgsnd` when the message is added to a queue. Messages are fetched from a queue by `msgrcv`. We don't have to fetch the messages in a first-in, first-out order. Instead, we can fetch messages based on their type field.

Each queue has the following `msqid_ds` structure associated with it. This structure defines the current status of the queue.

```
struct msqid_ds {
  struct ipc_perm  msg_perm; /* see Section 14.6.2 */
  struct msg  *msg_first;  /* ptr to first message on queue */
  struct msg  *msg_last;   /* ptr to last message on queue */
  ulong        msg_cbytes; /* current # bytes on queue */
  ulong        msg_qnum;   /* # of messages on queue */
  ulong        msg_qbytes; /* max # of bytes on queue */
  pid_t        msg_lspid;  /* pid of last msgsnd() */
  pid_t        msg_lrpid;  /* pid of last msgrcv() */
  time_t       msg_stime;  /* last-msgsnd() time */
  time_t       msg_rtime;  /* last-msgrcv() time */
  time_t       msg_ctime;  /* last-change time */
};
```

The two pointers, `msg_first` and `msg_last` are worthless to a user process, as these point to where the corresponding messages are stored within the kernel. The remaining members of the structure are self-defining.

Figure 14.16 lists the system limits (Section 14.6.3) that affect message queues.

EXHIBIT 13

```
#include      <sys/types.h>
#include      <sys/ipc.h>
#include      <sys/msg.h>

#define MSGKEY      75

struct msgform {
      long      mtype;
      char      mtext[256];
};

main()
{
      struct msgform msg;
      int msgid, pid, *pint;

      msgid = msgget(MSGKEY, 0777);

      pid = getpid();
      pint = (int *) msg.mtext;
      *pint = pid;      /* copy pid into message text */
      msg.mtype = 1;

      msgsnd(msgid, &msg, sizeof(int), 0);
      msgrcv(msgid, &msg, 256, pid, 0);      /* pid is used as the msg type */
      printf("client: receive from pid %d\n", *pint);
}
```

Figure 11.6. A Client Process

A process can receive messages of a particular type by setting the *type* parameter appropriately. If it is a positive integer, the kernel returns the first message of the given type. If it is negative, the kernel finds the lowest type of all messages on the queue, provided it is less than or equal to the absolute value of *type*, and returns the first message of that type. For example, if a queue contains three messages whose types are 3, 1, and 2, respectively, and a user requests a message with type −2, the kernel returns the message of type 1. In all cases, if no messages on the queue satisfy the receive request, the kernel puts the process to sleep, unless the process had specified to return immediately by setting the *IPC_NOWAIT* bit in *flag*.

Consider the programs in Figures 11.6 and 11.8. The program in Figure 11.8 shows the structure of a *server* that provides generic service to *client* processes. For instance, it may receive requests from client processes to provide information from a database; the server process is a single point of access to the database, making consistency and security easier. The server creates a message structure by setting

```
#include        <sys/types.h>
#include        <sys/ipc.h>
#include        <sys/msg.h>

#define MSGKEY        75
struct msgform
{
        long mtype;
        char mtext[256];
} msg;
int msgid;

main()
{
        int i, pid, *pint;
        extern cleanup();

        for (i = 0;  i < 20;  i++)
                signal(i, cleanup);
        msgid = msgget(MSGKEY, 0777 | IPC_CREAT);

        for (;;)
        {
                msgrcv(msgid, &msg, 256, 1, 0);
                pint = (int *) msg.mtext;
                pid = *pint;
                printf("server: receive from pid %d\n", pid);
                msg.mtype = pid;
                *pint = getpid();
                msgsnd(msgid, &msg, sizeof(int), 0);
        }
}

cleanup()
{
        msgctl(msgid, IPC_RMID, 0);
        exit();
}
```

Figure 11.8. A Server Process

Messages are formatted as type-data pairs, whereas file data is a byte stream. The *type* prefix allows processes to select messages of a particular type, if desired, a feature not readily available in the file system. Processes can thus extract messages of particular types from the message queue in the order that they arrive, and the kernel maintains the proper order. Although it is possible to implement a message

EXHIBIT 14

The *cmd* argument specifies the command to be performed, on the queue specified by *msqid*.

`IPC_STAT` Fetch the `msqid_ds` structure for this queue, storing it in the structure pointed to by *buf*.

`IPC_SET` Set the following four fields from the structure pointed to by *buf* in the structure associated with this queue: `msg_perm.uid`, `msg_perm.gid`, `msg_perm.mode`, and `msg_qbytes`. This command can be executed only by a process whose effective user ID equals `msg_perm.cuid` or `msg_perm.uid`, or by a process with superuser privileges. Only the superuser can increase the value of `msg_qbytes`.

`IPC_RMID` Remove the message queue from the system and any data still on the queue. This removal is immediate. Any other process still using the message queue will get an error of `EIDRM` on its next attempted operation on the queue. This command can be executed only by a process whose effective user ID equals `msg_perm.cuid` or `msg_perm.uid`, or by a process with superuser privileges.

We'll see that these three commands (`IPC_STAT`, `IPC_SET`, and `IPC_RMID`) are also provided for semaphores and shared memory.

Data is placed onto a message queue by calling `msgsnd`.

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/msg.h>

int msgsnd(int msqid, const void *ptr, size_t nbytes, int flag);
```

Returns: 0 if OK, −1 on error

As we mentioned earlier, each message is composed of a positive long integer type field, a nonnegative length (*nbytes*), and the actual data bytes (corresponding to the length). Messages are always placed at the end of the queue.

ptr points to a long integer that contains the positive integer message type, and it is immediately followed by the message data. (There is no message data if *nbytes* is 0.) If the largest message we send is 512 bytes, we can define the following structure

```
struct mymesg {
  long  mtype;      /* positive message type */
  char  mtext[512]; /* message data, of length nbytes */
};
```

The *ptr* argument is then a pointer to a `mymesg` structure. The message type can be used by the receiver to fetch messages in an order other than first-in, first-out.

A *flag* value of `IPC_NOWAIT` can be specified. This is similar to the nonblocking I/O flag for file I/O (Section 12.2). If the message queue is full (either the total number

EXHIBIT 15

```
    mode_t mode;  /* access modes */
    ulong  seq;   /* slot usage sequence number */
    key_t  key;   /* key */
};
```

All the fields other than `seq` are initialized when the IPC structure is created. At a later time we can modify the `uid`, `gid`, and `mode` fields, by calling `msgctl`, `semctl`, or `shmctl`. To change these values the calling process must either be the creator of the IPC structure, or it must be the superuser. Changing these fields is similar to calling `chown` or `chmod` for a file.

The values in the `mode` field are similar to the values we saw in Figure 4.4, but there is nothing corresponding to execute permission for any of the IPC structures. Also, whereas message queues and shared memory use the terms read and write, semaphores use the terms read and alter. Figure 14.14 specifies the six permissions for each form of IPC.

Permission	Message queue	Semaphore	Shared memory
user-read	`MSG_R`	`SEM_R`	`SHM_R`
user-write (alter)	`MSG_W`	`SEM_A`	`SHM_W`
group-read	`MSG_R >> 3`	`SEM_R >> 3`	`SHM_R >> 3`
group-write (alter)	`MSG_W >> 3`	`SEM_A >> 3`	`SHM_W >> 3`
other-read	`MSG_R >> 6`	`SEM_R >> 6`	`SHM_R >> 6`
other-write (alter)	`MSG_W >> 6`	`SEM_A >> 6`	`SHM_W >> 6`

Figure 14.14 System V IPC permissions.

14.6.3 Configuration Limits

All three forms of System V IPC have built-in limits that we may encounter. Most of these can be changed by reconfiguring the kernel. We describe the limits when we describe each of the three forms of IPC.

> Under SVR4 these values, and their minimum and maximum values, are in the file `/etc/conf/cf.d/mtune`.

14.6.4 Advantages and Disadvantages

A fundamental problem with System V IPC is that the IPC structures are systemwide and do not have a reference count. For example, if we create a message queue, place some messages on the queue, and then terminate, the message queue and its contents are not deleted. They remain in the system until specifically read or deleted: by some process calling `msgrcv` or `msgctl`, by someone executing the `ipcrm`(1) command, or by the system being rebooted. Compare this with a pipe, which is completely removed when the last process to reference it terminates. With a FIFO, although the name stays in the filesystem until explicitly removed, any data left in a FIFO is removed when the last process to reference the FIFO terminates.

EXHIBIT 16

Tool Interface Standard (TIS) Executable and Linking Format (ELF) Specification

Version 1.2

TIS Committee

May 1995

Preface

This Executable and Linking Format Specification, Version 1.2, is the result of the work of the Tool Interface Standards (TIS) Committee--an association of members of the microcomputer industry formed to work toward standardization of the software interfaces visible to development tools for 32-bit Intel Architecture operating environments. Such interfaces include object module formats, executable file formats, and debug record information and formats.

The goal of the committee is to help streamline the software development process throughout the microcomputer industry, currently concentrating on 32-bit operating environments. To that end, the committee has developed specifications--some for file formats that are portable across leading industry operating systems, and others describing formats for 32-bit Windows* operating systems. Originally distributed collectively as the TIS Portable Formats Specifications Version 1.1, these specifications are now separated and distributed individually.

TIS Committee members include representatives from Absoft, Autodesk, Borland International Corporation, IBM Corporation, Intel Corporation, Lahey, Lotus Corporation, MetaWare Corporation, Microtec Research, Microsoft Corporation, Novell Corporation, The Santa Cruz Operation, and WATCOM International Corporation. PharLap Software Incorporated and Symantec Corporation also participated in the specification definition efforts.

This specification like the others in the TIS collection of specifications is based on existing, proven formats in keeping with the TIS Committee's goal to adopt, and when necessary, extend existing standards rather than invent new ones.

About ELF: Executable and Linking Format

The Executable and Linking Format was originally developed and published by UNIX System Laboratories (USL) as part of the Application Binary Interface (ABI). The Tool Interface Standards committee (TIS) has selected the evolving ELF standard as a portable object file format that works on 32-bit Intel Architecture environments for a variety of operating systems.

The ELF standard is intended to streamline software development by providing developers with a set of binary interface definitions that extend across multiple operating environments. This should reduce the number of different interface implementations, thereby reducing the need for recoding and recompiling code.

About This Document

This document is intended for developers who are creating object or executable files on various 32-bit environment operating systems. In order to extend ELf into different operating systems, the current ELF version 1.2 document has been reorganized based on operating system-specific information. It is divided into the following three books:

- Book I: *Executable and Linking Format*, describes the object file format called ELF. This book also contains an appendix that describes historical references and lists processor and operating system reserved names and words.
- Book II: *Processor Specific (Intel Achitecture)*, conveys hardware-specific ELF information, such as Intel Architecture information.
- Book III: *Operating System Specific*, describes ELF information that is operating system dependent, such as System V Release 4 information. This book also contains an appendix that describes ELF information that is both operating system and processor dependent.

Contents

Book III: Operating System Specific (UNIX System V Release 4)

1. Object Files

2. Program Loading and Dynamic Linking

3. Intel Architecture and System V Release 4 Dependencies

List of Figures

Book I: Executable and Linking Format (ELF)

Book II: Processor Specific (Intel Architecture)

Book III: Operating System Specific (UNIX System V Release 4)

Book I:
Executable and Linking Format (ELF)

Contents

Book I: Executable and Linking Format (ELF)

1 Object Files

2 Program Loading and Dynamic Linking

A Reserved Names

Contents

Figures

Introduction

This chapter describes the object file format, called ELF (Executable and Linking Format). There are three main types of object files.

- A *relocatable file* holds code and data suitable for linking with other object files to create an executable or a shared object file.
- An *executable file* holds a program suitable for execution.
- A *shared object file* holds code and data suitable for linking in two contexts. First, the link editor may process it with other relocatable and shared object files to create another object file. Second, the dynamic linker combines it with an executable file and other shared objects to create a process image.

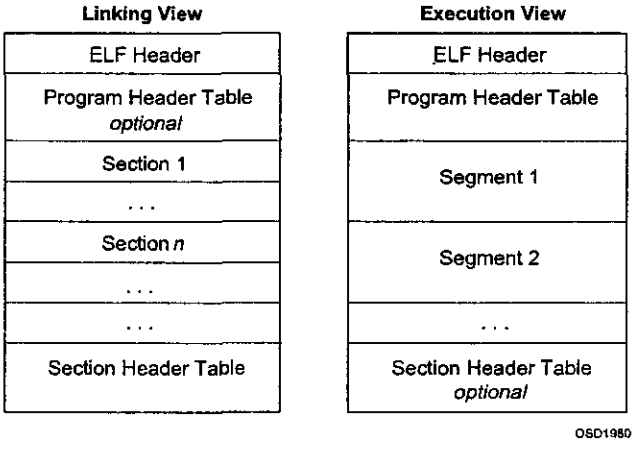
Created by the assembler and link editor, object files are binary representations of programs intended to execute directly on a processor. Programs that require other abstract machines are excluded.

After the introductory material, this chapter focuses on the file format and how it pertains to building programs. Chapter 2 also describes parts of the object file, concentrating on the information necessary to execute a program.

File Format

Object files participate in program linking (building a program) and program execution (running a program). For convenience and efficiency, the object file format provides parallel views of a file's contents, reflecting the differing needs of these activities. Figure 1-1 shows an object file's organization.

Figure 1-1. Object File Format



An *ELF header* resides at the beginning and holds a "road map" describing the file's organization. *Sections* hold the bulk of object file information for the linking view: instructions, data, symbol table, relocation information, and so on. Descriptions of special sections appear later in this section. Chapter 2 also describes *segments* and the program execution view of the file.

A *program header table*, if present, tells the system how to create a process image. Files used to build a process image (execute a program) must have a program header table; relocatable files do not need one. A *section header table* contains information describing the file's sections. Every section has an entry in the table; each entry gives information such as the section name, the section size, and so on. Files used during linking must have a section header table; other object files may or may not have one.

NOTE. Although the figure shows the program header table immediately after the ELF header, and the section header table following the sections, actual files may differ. Moreover, sections and segments have no specified order. Only the ELF header has a fixed position in the file.

Data Representation

As described here, the object file *format* supports various processors with 8-bit bytes and 32-bit architectures. Nevertheless, it is intended to be extensible to larger (or smaller) architectures. Object files therefore represent some control data with a machine-independent format, making it possible to identify object files and interpret their contents in a common way. Remaining data in an object file use the encoding of the target processor, regardless of the machine on which the file was created.

Figure 1-2. 32-Bit Data Types

Name	Size	Alignment	Purpose
`Elf32_Addr`	4	4	Unsigned program address
`Elf32_Half`	2	2	Unsigned medium integer
`Elf32_Off`	4	4	Unsigned file offset
`Elf32_Sword`	4	4	Signed large integer
`Elf32_Word`	4	4	Unsigned large integer
`unsigned char`	1	1	Unsigned small integer

All data structures that the object file format defines follow the "natural" size and alignment guidelines for the relevant class. If necessary, data structures contain explicit padding to ensure 4-byte alignment for 4-byte objects, to force structure sizes to a multiple of 4, and so on. Data also have suitable alignment from the beginning of the file. Thus, for example, a structure containing an `Elf32_Addr` member will be aligned on a 4-byte boundary within the file.

For portability reasons, ELF uses no bit fields.

Character Representations

This section describes the default ELF character representation and defines the standard character set used for external files that should be portable among systems. Several external file formats represent control information with characters. These single-byte characters use the 7-bit ASCII character set. In other words, when the ELF interface document mentions character constants, such as, '/' or '\n' their numerical values should follow the 7-bit ASCII guidelines. For the previous character constants, the single-byte values would be 47 and 10, respectively.

Character values outside the range of 0 to 127 may occupy one or more bytes, according to the character encoding. Applications can control their own character sets, using different character set extensions for different languages as appropriate. Although TIS-conformance does not restrict the character sets, they generally should follow some simple guidelines.

- Character values between 0 and 127 should correspond to the 7-bit ASCII code. That is, character sets with encodings above 127 should include the 7-bit ASCII code as a subset.
- Multibyte character encodings with values above 127 should contain only bytes with values outside the range of 0 to 127. That is, a character set that uses more than one byte per character should not "embed" a byte resembling a 7-bit ASCII character within a multibyte, non-ASCII character.
- Multibyte characters should be self-identifying. That allows, for example, any multibyte character to be inserted between any pair of multibyte characters, without changing the characters' interpretations.

These cautions are particularly relevant for multilingual applications.

NOTE. There are naming conventions for ELF constants that have processor ranges specified. Names such as DT_, PT_, for processor specific extensions, incorporate the name of the processor: DT_M32_SPECIAL, for example. However, pre-existing processor extensions not using this convention will be supported.

Pre-existing Extensions

`DT_JMP_REL`

ELF Header

Some object file control structures can grow, because the ELF header contains their actual sizes. If the object file format changes, a program may encounter control structures that are larger or smaller than expected. Programs might therefore ignore "extra" information. The treatment of "missing" information depends on context and will be specified when and if extensions are defined.

Figure 1-3. ELF Header

```
#define EI_NIDENT       16

typedef struct {
          unsigned char  e_ident[EI_NIDENT];
          Elf32_Half     e_type;
          Elf32_Half     e_machine;
          Elf32_Word     e_version;
          Elf32_Addr     e_entry;
          Elf32_Off      e_phoff;
          Elf32_Off      e_shoff;
          Elf32_Word     e_flags;
          Elf32_Half     e_ehsize;
          Elf32_Half     e_phentsize;
          Elf32_Half     e_phnum;
          Elf32_Half     e_shentsize;
          Elf32_Half     e_shnum;
          Elf32_Half     e_shstrndx;
} Elf32_Ehdr;
```

`e_ident` The initial bytes mark the file as an object file and provide machine-independent data with which to decode and interpret the file's contents. Complete descriptions appear below, in "ELF Identification."

`e_type` This member identifies the object file type.

Name	Value	Meaning
ET_NONE	0	No file type
ET_REL	1	Relocatable file
ET_EXEC	2	Executable file
ET_DYN	3	Shared object file
ET_CORE	4	Core file
ET_LOPROC	0xff00	Processor-specific
ET_HIPROC	0xffff	Processor-specific

Although the core file contents are unspecified, type ET_CORE is reserved to mark the file type. Values from ET_LOPROC through ET_HIPROC (inclusive) are reserved for processor-specific semantics. Other values are reserved and will be assigned to new object file types as necessary.

e_machine This member's value specifies the required architecture for an individual file.

Name	Value	Meaning
ET_NONE	0	No machine
EM_M32	1	AT&T WE 32100
EM_SPARC	2	SPARC
EM_386	3	Intel Architecture
EM_68K	4	Motorola 68000
EM_88K	5	Motorola 88000
EM_860	7	Intel 80860
EM_MIPS	8	MIPS RS3000 Big-Endian
EM_MIPS_RS4_BE	10	MIPS RS4000 Big-Endian
RESERVED	11-16	Reserved for future use

Other values are reserved and will be assigned to new machines as necessary. Processor-specific ELF names use the machine name to distinguish them. For example, the flags mentioned below use the prefix EF_; a flag named WIDGET for the EM_XYZ machine would be called EF_XYZ_WIDGET.

e_version This member identifies the object file version.

Name	Value	Meaning
EV_NONE	0	Invalid versionn
EV_CURRENT	1	Current version

The value 1 signifies the original file format; extensions will create new versions with higher numbers. The value of EV_CURRENT, though given as 1 above, will change as necessary to reflect the current version number.

e_entry This member gives the virtual address to which the system first transfers control, thus starting the process. If the file has no associated entry point, this member holds zero.

e_phoff This member holds the program header table's file offset in bytes. If the file has no program header table, this member holds zero.

e_shoff This member holds the section header table's file offset in bytes. If the file has no section header table, this member holds zero.

e_flags This member holds processor-specific flags associated with the file. Flag names take the form EF_*machine_flag*.

e_ehsize This member holds the ELF header's size in bytes.

`e_phentsize` This member holds the size in bytes of one entry in the file's program header table; all entries are the same size.

`e_phnum` This member holds the number of entries in the program header table. Thus the product of `e_phentsize` and `e_phnum` gives the table's size in bytes. If a file has no program header table, `e_phnum` holds the value zero.

`e_shentsize` This member holds a section header's size in bytes. A section header is one entry in the section header table; all entries are the same size.

`e_shnum` This member holds the number of entries in the section header table. Thus the product of `e_shentsize` and `e_shnum` gives the section header table's size in bytes. If a file has no section header table, `e_shnum` holds the value zero.

`e_shstrndx` This member holds the section header table index of the entry associated with the section name string table. If the file has no section name string table, this member holds the value `SHN_UNDEF`. See "Sections" and "String Table" below for more information.

ELF Identification

As mentioned above, ELF provides an object file framework to support multiple processors, multiple data encodings, and multiple classes of machines. To support this object file family, the initial bytes of the file specify how to interpret the file, independent of the processor on which the inquiry is made and independent of the file's remaining contents.

The initial bytes of an ELF header (and an object file) correspond to the `e_ident` member.

Figure 1-4. `e_ident[]` **Identification Indexes**

Name	Value	Purpose
EI_MAG0	0	File identification
EI_MAG1	1	File identification
EI_MAG2	2	File identification
EI_MAG3	3	File identification
EI_CLASS	4	File class
EI_DATA	5	Data encoding
EI_VERSION	6	File version
EI_PAD	7	Start of padding bytes
EI_NIDENT	16	Size of e_ident[]

These indexes access bytes that hold the following values.

EI_MAG0 to EI_MAG3 A file's first 4 bytes hold a "magic number," identifying the file as an ELF object file.

Name	Value	Meaning
ELFMAG0	0x7f	e_ident[EI_MAG0]
ELFMAG1	'E'	e_ident[EI_MAG1]
ELFMAG2	'L'	e_ident[EI_MAG2]
ELFMAG3	'F'	e_ident[EI_MAG3]

EI_CLASS The next byte, e_ident[EI_CLASS], identifies the file's class, or capacity.

Name	Value	Meaning
ELFCLASSNONE	0	Invalid class
ELFCLASS32	1	32-bit objects
ELFCLASS64	2	64-bit objects

The file format is designed to be portable among machines of various sizes, without imposing the sizes of the largest machine on the smallest. Class ELFCLASS32 supports machines with files and virtual address spaces up to 4 gigabytes; it uses the basic types defined above.

Class ELFCLASS64 is incomplete and refers to the 64-bit architectures. Its appearance here shows how the object file may change. Other classes will be defined as necessary, with different basic types and sizes for object file data.

EI_DATA Byte e_ident[EI_DATA] specifies the data encoding of the processor-specific data in the object file. The following encodings are currently defined.

Name	Value	Meaning
ELFDATANONE	0	Invalid data encoding
ELFDATA2LSB	1	See below
ELFDATA2MSB	2	See below

More information on these encodings appears below. Other values are reserved and will be assigned to new encodings as necessary.

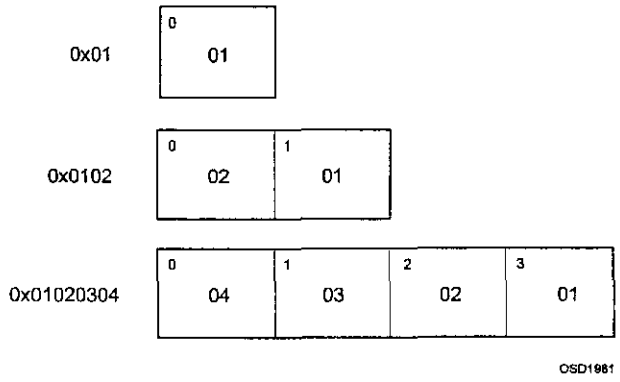
EI_VERSION Byte e_ident[EI_VERSION] specifies the ELF header version number. Currently, this value must be EV_CURRENT, as explained above for e_version.

EI_PAD This value marks the beginning of the unused bytes in e_ident. These bytes are reserved and set to zero; programs that read object files should ignore them. The value of EI_PAD will change in the future if currently unused bytes are given meanings.

A file's data encoding specifies how to interpret the basic objects in a file. As described above, class ELFCLASS32 files use objects that occupy 1, 2, and 4 bytes. Under the defined encodings, objects are represented as shown below. Byte numbers appear in the upper left corners.

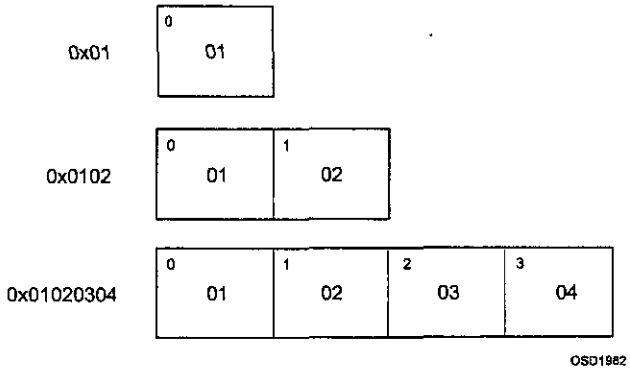
Encoding ELFDATA2LSB specifies 2's complement values, with the least significant byte occupying the lowest address.

Figure 1-5. Data Encoding ELFDATA2LSB



Encoding ELFDATA2MSB specifies 2's complement values, with the most significant byte occupying the lowest address.

Figure 1-6. Data Encoding ELFDATA2MSB



Sections

An object file's section header table lets one locate all the file's sections. The section header table is an array of `Elf32_Shdr` structures as described below. A section header table index is a subscript into this array. The ELF header's `e_shoff` member gives the byte offset from the beginning of the file to the section header table; `e_shnum` tells how many entries the section header table contains; `e_shentsize` gives the size in bytes of each entry.

Some section header table indexes are reserved; an object file will not have sections for these special indexes.

Figure 1-7. Special Section Indexes

Name	Value
SHN_UNDEF	0
SHN_LORESERVE	0xff00
SHN_LOPROC	0xff00
SHN_HIPROC	0xff1f
SHN_ABS	0xfff1
SHN_COMMON	0xfff2
SHN_HIRESERVE	0xffff

SHN_UNDEF — This value marks an undefined, missing, irrelevant, or otherwise meaningless section reference. For example, a symbol "defined" relative to section number SHN_UNDEF is an undefined symbol.

NOTE. Although index 0 is reserved as the undefined value, the section header table contains an entry for index 0. That is, if the `e_shnum` *member of the ELF header says a file has 6 entries in the section header table, they have the indexes 0 through 5. The contents of the initial entry are specified later in this section.*

SHN_LORESERVE — This value specifies the lower bound of the range of reserved indexes.

SHN_LOPROC through SHN_HIPROC — Values in this inclusive range are reserved for processor-specific semantics.

SHN_ABS — This value specifies absolute values for the corresponding reference. For example, symbols defined relative to section number SHN_ABS have absolute values and are not affected by relocation.

SHN_COMMON — Symbols defined relative to this section are common symbols, such as FORTRAN COMMON or unallocated C external variables.

SHN_HIRESERVE	This value specifies the upper bound of the range of reserved indexes. The system reserves indexes between SHN_LORESERVE and SHN_HIRESERVE, inclusive; the values do not reference the section header table. That is, the section header table does *not* contain entries for the reserved indexes.

Sections contain all information in an object file, except the ELF header, the program header table, and the section header table. Moreover, object files' sections satisfy several conditions.

- Every section in an object file has exactly one section header describing it. Section headers may exist that do not have a section.
- Each section occupies one contiguous (possibly empty) sequence of bytes within a file.
- Sections in a file may not overlap. No byte in a file resides in more than one section.
- An object file may have inactive space. The various headers and the sections might not "cover" every byte in an object file. The contents of the inactive data are unspecified.

A section header has the following structure.

Figure 1-8. Section Header

```
typedef struct {
          Elf32_Word    sh_name;
          Elf32_Word    sh_type;
          Elf32_Word    sh_flags;
          Elf32_Addr    sh_addr;
          Elf32_Off     sh_offset;
          Elf32_Word    sh_size;
          Elf32_Word    sh_link;
          Elf32_Word    sh_info;
          Elf32_Word    sh_addralign;
          Elf32_Word    sh_entsize;
} Elf32_Shdr;
```

sh_name	This member specifies the name of the section. Its value is an index into the section header string table section [see "String Table" below], giving the location of a null-terminated string.
sh_type	This member categorizes the section's contents and semantics. Section types and their descriptions appear below.
sh_flags	Sections support 1-bit flags that describe miscellaneous attributes. Flag definitions appear below.
sh_addr	If the section will appear in the memory image of a process, this member gives the address at which the section's first byte should reside. Otherwise, the member contains 0.

sh_offset	This member's value gives the byte offset from the beginning of the file to the first byte in the section. One section type, SHT_NOBITS described below, occupies no space in the file, and its sh_offset member locates the conceptual placement in the file.
sh_size	This member gives the section's size in bytes. Unless the section type is SHT_NOBITS, the section occupies sh_size bytes in the file. A section of type SHT_NOBITS may have a non-zero size, but it occupies no space in the file.
sh_link	This member holds a section header table index link, whose interpretation depends on the section type. A table below describes the values.
sh_info	This member holds extra information, whose interpretation depends on the section type. A table below describes the values.
sh_addralign	Some sections have address alignment constraints. For example, if a section holds a doubleword, the system must ensure doubleword alignment for the entire section. That is, the value of sh_addr must be congruent to 0, modulo the value of sh_addralign. Currently, only 0 and positive integral powers of two are allowed. Values 0 and 1 mean the section has no alignment constraints.
sh_entsize	Some sections hold a table of fixed-size entries, such as a symbol table. For such a section, this member gives the size in bytes of each entry. The member contains 0 if the section does not hold a table of fixed-size entries.

A section header's sh_type member specifies the section's semantics.

Figure 1-9. Section Types, sh_type

Name	Value
SHT_NULL	0
SHT_PROGBITS	1
SHT_SYMTAB	2
SHT_STRTAB	3
SHT_RELA	4
SHT_HASH	5
SHT_DYNAMIC	6
SHT_NOTE	7
SHT_NOBITS	8
SHT_REL	9
SHT_SHLIB	10
SHT_DYNSYM	11
SHT_LOPROC	0x70000000
SHT_HIPROC	0x7fffffff
SHT_LOUSER	0x80000000
SHT_HIUSER	0xffffffff

SHT_NULL	This value marks the section header as inactive; it does not have an associated section. Other members of the section header have undefined values.
SHT_PROGBITS	The section holds information defined by the program, whose format and meaning are determined solely by the program.
SHT_SYMTAB and SHT_DYNSYM	These sections hold a symbol table.
SHT_STRTAB	The section holds a string table.
SHT_RELA	The section holds relocation entries with explicit addends, such as type Elf32_Rela for the 32-bit class of object files. An object file may have multiple relocation sections. See "Relocation" below for details.
SHT_HASH	The section holds a symbol hash table.
SHT_DYNAMIC	The section holds information for dynamic linking.
SHT_NOTE	This section holds information that marks the file in some way.
SHT_NOBITS	A section of this type occupies no space in the file but otherwise resembles SHT_PROGBITS. Although this section contains no bytes, the sh_offset member contains the conceptual file offset.
SHT_REL	The section holds relocation entries without explicit addends, such as type Elf32_Rel for the 32-bit class of object files. An object file may have multiple relocation sections. See "Relocation" below for details.
SHT_SHLIB	This section type is reserved but has unspecified semantics.

SHT_LOPROC through SHT_HIPROC	Values in this inclusive range are reserved for processor-specific semantics.
SHT_LOUSER	This value specifies the lower bound of the range of indexes reserved for application programs.
SHT_HIUSER	This value specifies the upper bound of the range of indexes reserved for application programs. Section types between SHT_LOUSER and SHT_HIUSER may be used by the application, without conflicting with current or future system-defined section types.

Other section type values are reserved. As mentioned before, the section header for index 0 (SHN_UNDEF) exists, even though the index marks undefined section references. This entry holds the following.

Figure 1-10. Section Header Table Entry: Index 0

Name	Value	Note
sh_name	0	No name
sh_type	SHT_NULL	Inactive
sh_flags	0	No flags
sh_addr	0	No address
sh_offset	0	No file offset
sh_size	0	No size
sh_link	SHN_UNDEF	No link information
sh_info	0	No auxiliary information
sh_addralign	0	No alignment
sh_entsize	0	No entries

A section header's sh_flags member holds 1-bit flags that describe the section's attributes. Defined values appear below; other values are reserved.

Figure 1-11. Section Attribute Flags, sh_flags

Name	Value
SHF_WRITE	0x1
SHF_ALLOC	0x2
SHF_EXECINSTR	0x4
SHF_MASKPROC	0xf0000000

If a flag bit is set in sh_flags, the attribute is "on" for the section. Otherwise, the attribute is "off" or does not apply. Undefined attributes are set to zero.

SHF_WRITE	The section contains data that should be writable during process execution.

SHF_ALLOC	The section occupies memory during process execution. Some control sections do not reside in the memory image of an object file; this attribute is off for those sections.
SHF_EXECINSTR	The section contains executable machine instructions.
SHF_MASKPROC	All bits included in this mask are reserved for processor-specific semantics.

Two members in the section header, sh_link and sh_info, hold special information, depending on section type.

Figure 1-12. sh_link and sh_info Interpretation

sh_type	sh_link	sh_info
SHT_DYNAMIC	The section header index of the string table used by entries in the section.	0
SHT_HASH	The section header index of the symbol table to which the hash table applies.	0
SHT_REL SHT_RELA	The section header index of the associated symbol table.	The section header index of the section to which the relocation applies.
SHT_SYMTAB SHT_DYNSYM	This information is operating system specific.	This information is operating system specific.
other	SHN_UNDEF	0

Special Sections

Various sections in ELF are pre-defined and hold program and control information. These Sections are used by the operating system and have different types and attributes for different operating systems.

Executable files are created from individual object files and libraries through the linking process. The linker resolves the references (including subroutines and data references) among the different object files, adjusts the absolute references in the object files, and relocates instructions. The linking and loading processes, which are described in Chapter 2, require information defined in the object files and store this information in specific sections such as .dynamic.

Each operating system supports a set of linking models which fall into two categories:

Static	A set of object files, system libraries and library archives are statically bound, references are resolved, and an executable file is created that is completely self contained.
Dynamic	A set of object files, libraries, system shared resources and other shared libraries are linked together to create the executable. When this executable is loaded, other shared resources and dynamic libraries must be made available in the system for the program to run successfully.

The general method used to resolve references at execution time for a dynamically linked executable file is described in the linkage model used by the operating system, and the actual implementation of this linkage model will contain processor-specific components.

There are also sections that support debugging, such as `.debug` and `.line`, and program control, including `.bss`, `.data`, `.data1`, `.rodata`, and `.rodata1`.

Figure 1-13. Special Sections

Name	Type	Attributes
.bss	SHT_NOBITS	SHF_ALLOC+SHF_WRITE
.comment	SHT_PROGBITS	none
.data	SHT_PROGBITS	SHF_ALLOC + SHF_WRITE
.data1	SHT_PROGBITS	SHF_ALLOC + SHF_WRITE
.debug	SHT_PROGBITS	none
.dynamic	SHT_DYNAMIC	see below
.hash	SHT_HASH	SHF_ALLOC
.line	SHT_PROGBITS	none
.note	SHT_NOTE	none
.rodata	SHT_PROGBITS	SHF_ALLOC
.rodata1	SHT_PROGBITS	SHF_ALLOC
.shstrtab	SHT_STRTAB	none
.strtab	SHT_STRTAB	see below
.symtab	SHT_SYMTAB	see below
.text	SHT_PROGBITS	SHF_ALLOC + SHF_EXECINSTR

`.bss` This section holds uninitialized data that contribute to the program's memory image. By definition, the system initializes the data with zeros when the program begins to run. The section occupies no file space, as indicated by the section type, `SHT_NOBITS`.

`.comment` This section holds version control information.

`.data` and `.data1` These sections hold initialized data that contribute to the program's memory image.

`.debug` This section holds information for symbolic debugging. The contents are unspecified. All section names with the prefix `.debug` are reserved for future use.

`.dynamic` This section holds dynamic linking information and has attributes such as `SHF_ALLOC` and `SHF_WRITE`. Whether the `SHF_WRITE` bit is set is determined by the operating system and processor.

`.hash` This section holds a symbol hash table.

`.line`	This section holds line number information for symbolic debugging, which describes the correspondence between the source program and the machine code. The contents are unspecified.
`.note`	This section holds information in the format that is described in the "Note Section" in Chapter 2.
`.rodata` and `.rodata1`	These sections hold read-only data that typically contribute to a non-writable segment in the process image. See "Program Header" in Chapter 2 for more information.
`.shstrtab`	This section holds section names.
`.strtab`	This section holds strings, most commonly the strings that represent the names associated with symbol table entries. If a file has a loadable segment that includes the symbol string table, the section's attributes will include the `SHF_ALLOC` bit; otherwise, that bit will be off.
`.symtab`	This section holds a symbol table, as "Symbol Table" in this chapter describes. If a file has a loadable segment that includes the symbol table, the section's attributes will include the `SHF_ALLOC` bit; otherwise, that bit will be off.
`.text`	This section holds the "text," or executable instructions, of a program.

Section names with a dot (.) prefix are reserved for the system, although applications may use these sections if their existing meanings are satisfactory. Applications may use names without the prefix to avoid conflicts with system sections. The object file format lets one define sections not in the list above. An object file may have more than one section with the same name.

String Table

This section describes the default string table. String table sections hold null-terminated character sequences, commonly called strings. The object file uses these strings to represent symbol and section names. One references a string as an index into the string table section. The first byte, which is index zero, is defined to hold a null character. Likewise, a string table's last byte is defined to hold a null character, ensuring null termination for all strings. A string whose index is zero specifies either no name or a null name, depending on the context. An empty string table section is permitted; its section header's `sh_size` member would contain zero. Non-zero indexes are invalid for an empty string table.

A section header's `sh_name` member holds an index into the section header string table section, as designated by the `e_shstrndx` member of the ELF header. The following figures show a string table with 25 bytes and the strings associated with various indexes.

Index	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9
0	\0	n	a	m	e	.	\0	V	a	r
10	i	a	b	l	e	\0	a	b	l	e
20	\0	\0	x	x	\0					

Figure 1-14. String Table Indexes

Index	String
0	*none*
1	`name.`
7	`Variable`
11	`able`
16	`able`
24	*null string*

As the example shows, a string table index may refer to any byte in the section. A string may appear more than once; references to substrings may exist; and a single string may be referenced multiple times. Unreferenced strings also are allowed.

Symbol Table

An object file's symbol table holds information needed to locate and relocate a program's symbolic definitions and references. A symbol table index is a subscript into this array. Index 0 both designates the first entry in the table and serves as the undefined symbol index. The contents of the initial entry are specified later in this section.

Name	Value
STN_UNDEF	0

A symbol table entry has the following format.

Figure 1-15. Symbol Table Entry

```
typedef struct {
          Elf32_Word    st_name;
          Elf32_Addr    st_value;
          Elf32_Word    st_size;
          unsigned char st_info;
          unsigned char st_other;
          Elf32_Half    st_shndx;
} Elf32_Sym;
```

`st_name` This member holds an index into the object file's symbol string table, which holds the character representations of the symbol names.

`st_value` This member gives the value of the associated symbol. Depending on the context, this may be an absolute value, an address, and so on; details appear below.

`st_size` Many symbols have associated sizes. For example, a data object's size is the number of bytes contained in the object. This member holds 0 if the symbol has no size or an unknown size.

`st_info` This member specifies the symbol's type and binding attributes. A list of the values and meanings appears below. The following code shows how to manipulate the values.

```
#define ELF32_ST_BIND(i)    ((i)>>4)
#define ELF32_ST_TYPE(i)    ((i)&0xf)
#define ELF32_ST_INFO(b,t)  (((b)<<4)+((t)&0xf))
```

`st_other` This member currently holds 0 and has no defined meaning.

`st_shndx` Every symbol table entry is "defined" in relation to some section; this member holds the relevant section header table index. As Figure 1-7 and the related text describe, some section indexes indicate special meanings.

A symbol's binding determines the linkage visibility and behavior.

Figure 1-16. Symbol Binding, `ELF32_ST_BIND`

Name	Value
STB_LOCAL	0
STB_GLOBAL	1
STB_WEAK	2
STB_LOPROC	13
STB_HIPROC	15

`STB_LOCAL` Local symbols are not visible outside the object file containing their definition. Local symbols of the same name may exist in multiple files without interfering with each other.

`STB_GLOBAL` Global symbols are visible to all object files being combined. One file's definition of a global symbol will satisfy another file's undefined reference to the same global symbol.

`STB_WEAK` Weak symbols resemble global symbols, but their definitions have lower precedence.

`STB_LOPROC` through `STB_HIPROC` Values in this inclusive range are reserved for processor-specific semantics.

In each symbol table, all symbols with `STB_LOCAL` binding precede the weak and global symbols. A symbol's type provides a general classification for the associated entity.

Figure 1-17. Symbol Types, `ELF32_ST_TYPE`

Name	Value
STT_NOTYPE	0
STT_OBJECT	1
STT_FUNC	2
STT_SECTION	3
STT_FILE	4
STT_LOPROC	13
STT_HIPROC	15

`STT_NOTYPE` The symbol's type is not specified.

`STT_OBJECT` The symbol is associated with a data object, such as a variable, an array, and so on.

`STT_FUNC` The symbol is associated with a function or other executable code.

`STT_SECTION` The symbol is associated with a section. Symbol table entries of this type exist primarily for relocation and normally have `STB_LOCAL` binding.

`STT_LOPROC` through `STT_HIPROC` Values in this inclusive range are reserved for processor-specific semantics. If a symbol's value refers to a specific location within a section, its section index member, `st_shndx`, holds an index into the section header table. As the section moves during relocation, the symbol's value changes as well, and references to the symbol continue to "point" to the same location in the program. Some special section index values give other semantics.

`STT_FILE` A file symbol has `STB_LOCAL` binding, its section index is `SHN_ABS`, and it precedes the other `STB_LOCAL` symbols for the file, if it is present.

The symbols in ELF object files convey specific information to the linker and loader. See the operating system sections for a description of the actual linking model used in the system.

`SHN_ABS` The symbol has an absolute value that will not change because of relocation.

`SHN_COMMON` The symbol labels a common block that has not yet been allocated. The symbol's value gives alignment constraints, similar to a section's `sh_addralign` member. That is, the link editor will allocate the storage for the symbol at an address that is a multiple of `st_value`. The symbol's size tells how many bytes are required.

`SHN_UNDEF` This section table index means the symbol is undefined. When the link editor combines this object file with another that defines the indicated symbol, this file's references to the symbol will be linked to the actual definition.

As mentioned above, the symbol table entry for index 0 (STN_UNDEF) is reserved; it holds the following.

Figure 1-18. Symbol Table Entry: Index 0

Name	Value	Note
st_name	0	No name
st_value	0	Zero value
st_size	0	No size
st_info	0	No type, local binding
st_other	0	
st_shndx	SHN_UNDEF	No section

Symbol Values

Symbol table entries for different object file types have slightly different interpretations for the st_value member.

- In relocatable files, st_value holds alignment constraints for a symbol whose section index is SHN_COMMON.
- In relocatable files, st_value holds a section offset for a defined symbol. That is, st_value is an offset from the beginning of the section that st_shndx identifies.
- In executable and shared object files, st_value holds a virtual address. To make these files' symbols more useful for the dynamic linker, the section offset (file interpretation) gives way to a virtual address (memory interpretation) for which the section number is irrelevant.

Although the symbol table values have similar meanings for different object files, the data allow efficient access by the appropriate programs.

Relocation

Relocation is the process of connecting symbolic references with symbolic definitions. For example, when a program calls a function, the associated call instruction must transfer control to the proper destination address at execution. In other words, relocatable files must have information that describes how to modify their section contents, thus allowing executable and shared object files to hold the right information for a process's program image. *Relocation entries* are these data.

Figure 1-19. Relocation Entries

```
typedef struct {
        Elf32_Addr      r_offset;
        Elf32_Word      r_info;
} Elf32_Rel;

typedef struct {
        Elf32_Addr      r_offset;
        Elf32_Word      r_info;
        Elf32_Sword     r_addend;
} Elf32_Rela;
```

`r_offset` This member gives the location at which to apply the relocation action. For a relocatable file, the value is the byte offset from the beginning of the section to the storage unit affected by the relocation. For an executable file or a shared object, the value is the virtual address of the storage unit affected by the relocation.

`r_info` This member gives both the symbol table index with respect to which the relocation must be made, and the type of relocation to apply. For example, a call instruction's relocation entry would hold the symbol table index of the function being called. If the index is `STN_UNDEF`, the undefined symbol index, the relocation uses 0 as the "symbol value." Relocation types are processor-specific; descriptions of their behavior appear in the processor supplement. When the text in the processor supplement refers to a relocation entry's relocation type or symbol table index, it means the result of applying `ELF32_R_TYPE` or `ELF32_R_SYM`, respectively, to the entry's `r_info` member.

```
#define ELF32_R_SYM(i)     ((i)>>8)
#define ELF32_R_TYPE(i)    ((unsigned char)(i))
#define ELF32_R_INFO(s,t)  (((s)<<8)+(unsigned char)(t))
```

`r_addend` This member specifies a constant addend used to compute the value to be stored into the relocatable field.

As shown above, only `Elf32_Rela` entries contain an explicit addend. Entries of type `Elf32_Rel` store an implicit addend in the location to be modified. Depending on the processor architecture, one form or the other might be necessary or more convenient. Consequently, an implementation for a particular machine may use one form exclusively or either form depending on context.

A relocation section references two other sections: a symbol table and a section to modify. The section header's `sh_info` and `sh_link` members, described in "Sections" above, specify these relationships. Relocation entries for different object files have slightly different interpretations for the `r_offset` member.

- In relocatable files, `r_offset` holds a section offset. That is, the relocation section itself describes how to modify another section in the file; relocation offsets designate a storage unit within the second section.
- In executable and shared object files, `r_offset` holds a virtual address. To make these files' relocation entries more useful for the dynamic linker, the section offset (file interpretation) gives way to a virtual address (memory interpretation).

Although the interpretation of `r_offset` changes for different object files to allow efficient access by the relevant programs, the relocation types' meanings stay the same.

Introduction

This chapter describes the object file information and system actions that create running programs. Executable and shared object files statically represent programs. To execute such programs, the system uses the files to create dynamic program representations, or process images. A process image has segments that hold its text, data, stack, and so on. This section describes the program header and complements Chapter 1, by describing object file structures that relate directly to program execution. The primary data structure, a program header table, locates segment images within the file and contains other information necessary to create the memory image for the program.

Given an object file, the system must load it into memory for the program to run. After the system loads the program, it must complete the process image by resolving symbolic references among the object files that compose the process.

Program Header

An executable or shared object file's program header table is an array of structures, each describing a segment or other information the system needs to prepare the program for execution. An object file *segment* contains one or more *sections*. Program headers are meaningful only for executable and shared object files. A file specifies its own program header size with the ELF header's `e_phentsize` and *e_phnum* members [see "ELF Header" in Chapter 1].

Figure 2-1. Program Header

```
typedef struct {
          Elf32_Word    p_type;
          Elf32_Off     p_offset;
          Elf32_Addr    p_vaddr;
          Elf32_Addr    p_paddr;
          Elf32_Word    p_filesz;
          Elf32_Word    p_memsz;
          Elf32_Word    p_flags;
          Elf32_Word    p_align;
} Elf32_Phdr;
```

`p_type` This member tells what kind of segment this array element describes or how to interpret the array element's information. Type values and their meanings appear below.

`p_offset` This member gives the offset from the beginning of the file at which the first byte of the segment resides.

`p_vaddr` This member gives the virtual address at which the first byte of the segment resides in memory.

`p_paddr` On systems for which physical addressing is relevant, this member is reserved for the segment's physical address. This member requires operating system specific information, which is described in the appendix at the end of Book III.

`p_filesz` This member gives the number of bytes in the file image of the segment; it may be zero.

`p_memsz` This member gives the number of bytes in the memory image of the segment; it may be zero.

`p_flags` This member gives flags relevant to the segment. Defined flag values appear below.

`p_align` Loadable process segments must have congruent values for `p_vaddr` and `p_offset`, modulo the page size.This member gives the value to which the segments are aligned in memory and in the file. Values 0 and 1 mean that no alignment is required. Otherwise, `p_align` should be a positive, integral power of 2, and `p_addr` should equal `p_offset`, modulo `p_align`.

Some entries describe process segments; others give supplementary information and do not contribute to the process image.

Figure 2-2. Segment Types, `p_type`

Name	Value
PT_NULL	0
PT_LOAD	1
PT_DYNAMIC	2
PT_INTERP	3
PT_NOTE	4
PT_SHLIB	5
PT_PHDR	6
PT_LOPROC	0x70000000
PT_HIPROC	0x7fffffff

PT_NULL The array element is unused; other members' values are undefined. This type lets the program header table have ignored entries.

PT_LOAD The array element specifies a loadable segment, described by `p_filesz` and `p_memsz`. The bytes from the file are mapped to the beginning of the memory segment. If the segment's memory size (`p_memsz`) is larger than the file size (`p_filesz`), the "extra" bytes are defined to hold the value 0 and to follow the segment's initialized area. The file size may not be larger than the memory size. Loadable segment entries in the program header table appear in ascending order, sorted on the `p_vaddr` member.

PT_DYNAMIC The array element specifies dynamic linking information. See Book III.

PT_INTERP The array element specifies the location and size of a null-terminated path name to invoke as an interpreter. See Book III.

PT_NOTE The array element specifies the location and size of auxiliary information.

PT_SHLIB This segment type is reserved but has unspecified semantics. See Book III.

PT_PHDR The array element, if present, specifies the location and size of the program header table itself, both in the file and in the memory image of the program. This segment type may not occur more than once in a file. Moreover, it may occur only if the program header table is part of the memory image of the program. If it is present, it must precede any loadable segment entry. See "Program Interpreter" in the appendix at the end of Book III for further information.

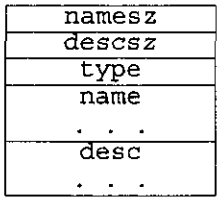
PT_LOPROC through PT_HIPROC	Values in this inclusive range are reserved for processor-specific semantics.

NOTE. Unless specifically required elsewhere, all program header segment types are optional. That is, a file's program header table may contain only those elements relevant to its contents.

Note Section

Sometimes a vendor or system builder needs to mark an object file with special information that other programs will check for conformance, compatibility, etc. Sections of type SHT_NOTE and program header elements of type PT_NOTE can be used for this purpose. The note information in sections and program header elements holds any number of entries, each of which is an array of 4-byte words in the format of the target processor. Labels appear below to help explain note information organization, but they are not part of the specification.

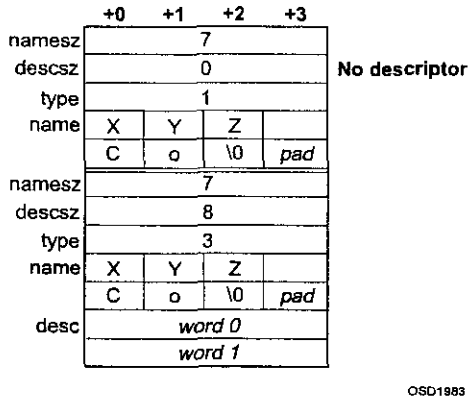
Figure 2-3. Note Information



namesz and name	The first namesz bytes in name contain a null-terminated character representation of the entry's owner or originator. There is no formal mechanism for avoiding name conflicts. By convention, vendors use their own name, such as "XYZ Computer Company," as the identifier. If no name is present, namesz contains 0. Padding is present, if necessary, to ensure 4-byte alignment for the descriptor. Such padding is not included in namesz.
descsz and desc	The first descsz bytes in desc hold the note descriptor. ELF places no constraints on a descriptor's contents. If no descriptor is present, descsz contains 0. Padding is present, if necessary, to ensure 4-byte alignment for the next note entry. Such padding is not included in descsz.
type	This word gives the interpretation of the descriptor. Each originator controls its own types; multiple interpretations of a single type value may exist. Thus, a program must recognize both the name and the type to "understand" a descriptor. Types currently must be non-negative. ELF does not define what descriptors mean.

To illustrate, the following note segment holds two entries.

Figure 2-4. Example Note Segment



NOTE. The system reserves note information with no name (`namesz==0`*) and with a zero-length name (*`name[0]=='\0'`*) but currently defines no types. All other names must have at least one non-null character.*

NOTE. Note information is optional. The presence of note information does not affect a program's TIS conformance, provided the information does not affect the program's execution behavior. Otherwise, the program does not conform to the TIS ELF specification and has undefined behavior.

Program Loading

Program loading is the process by which the operating system creates or augments a process image. The manner in which this process is accomplished and how the page management functions for the process are handled are dictated by the operating system and processor. See the appendix at the end of Book III for more details.

Dynamic Linking

The dynamic linking process resolves references either at process initialization time and/or at execution time. Some basic mechanisms need to be set up for a particular linkage model to work, and there are ELF sections and header elements reserved for this purpose. The actual definition of the linkage model is determined by the operating system and implementation. Therefore, the contents of these sections are both operating system and processor specific. (See the appendix at the end of Book III.)

CERTIFICATE OF SERVICE

I hereby certify that on the 20th day of June, 2005, a true and correct copy of the foregoing was sent by U.S. Mail, postage prepaid, to the following:

Brent O. Hatch
Mark F. James
HATCH, JAMES & DODGE, P.C.
10 West Broadway, Suite 400
Salt Lake City, Utah 84101

Stephen N. Zack
Mark J. Heise
BOIES, SCHILLER & FLEXNER LLP
100 Southeast Second Street, Suite 2800
Miami, Florida 33131

Robert Silver
Edward Normand
Sean Eskovitz
BOIES, SCHILLER & FLEXNER LLP
333 Main Street
Armonk, NY 10504

Amy F. Sorenson