At the system level: Building blocks

I want to take some time to talk about the fundamental toolset that most programs that system administrators work with are built over.

The most important of these are the system calls. When we run `strace` to see exactly what a process is doing, we are watching this fundamental interaction between a program and its requests to the operating system, usually for access to resources controlled by

the operating system.

Building blocks for Unix power tools

A Unix system call is a direct request to the kernel regarding a system resource. It might be a request for a file descriptor to manipulate a file, it might be a request to write to a file descriptor, or any of hundreds of possible operations.

These are exactly the tools that every Unix program is built upon.

File descriptor and file descriptor operations

In some sense, the mainstay operations are those on the file system.

File descriptor and file descriptor operations

Unlike many other resources which are just artifacts of the operating system and disappear at each reboot, changing a file system generally is an operation that has some permanence. Of course it is possible and even common to create “RAM” disk filesystems since they are quite fast and for items that are meant to be temporary, they are quite acceptable. (For instance, as

you might have done when setting up `MailScanner`, for instance, in `/var/spool/incoming`.)

Important file descriptor calls

A file descriptor is an `int`. It provides stateful access to an i/o resource such as a file on a filesystem, a pseudo-terminal, or a socket to a tcp session.

```
open()    -- create a new file descriptor to access a file
close()   -- deallocate a file descriptor
```

Important file descriptor calls

```
dup()      -- duplicate a file descriptor
dup2()     -- improved way to duplicate a file descriptor
```

Important file descriptor calls

```
fchmod()  -- change the permissions of a file associated with a file
          -- descriptor
fchown()  -- change the ownership of a file associated with a file
fchdir()  -- change the working directory for a process via fd
```

Important file descriptor calls

```
fcntl()   -- miscellaneous manipulation of file descriptors: dup(), set
          -- close on exec(), set to non-blocking, set to asynchronous
          -- mode, locks, signals
ioctl()   -- manipulate the underlying ``device'' parameters for a file
          -- descriptor
```

Important file descriptor calls

```
flock()   -- lock a file associated with a file descriptor
```

Important file descriptor calls

```
pipe()     -- create a one-way association between two file
           -- descriptors so that output from
           -- one goes to the input of the other
```

Important file descriptor calls

```
select()  -- multiplex on pending i/o to or from a set of file descriptor
```

Important file descriptor calls

```
read()    -- send data to a file descriptor
write()   -- take data from a file descriptor
fsync()   -- forces a flush for a file descriptor
```

Important file descriptor calls

`readdir() -- raw read of directory entry from a file descriptor`

Important file descriptor calls

```
fstat()    -- return information about a file associated with a fd: inode,
              perms, hard links, uid, gid, size, modtimes
fstatfs()  -- return the mount information for the filesystem that the fil
           -- descriptor is associated with
```

Important filesystem operations

In addition to using the indirect means of file descriptors, Unix also offers a number of direct functions on files.

```
access()   -- returns a value indicating if a file is accessible
chmod()    -- changes the permissions on a file in a filesystem
chown()    -- changes the ownership of a file in a filesystem
```

Important filesystem operations

```
link()    -- create a hard link to a file
symlink() -- create a soft link to a file
```

Important filesystem operations

```
mkdir()    -- create a new directory
rmdir()    -- remove a directory
```

Important filesystem operations

```
stat()    -- return information about a file associated with a fd: inode,
             perms, hard links, uid, gid, size, modtimes
statfs()  -- return the mount information for the filesystem that the fil
          -- descriptor is associated with
```

Signals

```
alarm       -- set an alarm clock for a SIGALRM to be sent to a process
            -- time measured in seconds
getitimer   -- set an alarm clock in fractions of a second to deliver eit
            -- SIGALRM, SIGVTALRM, SIGPROF
```

Signals

```
kill        -- send an arbitrary signal to an arbitrary process
killpg      -- send an arbitrary signal to all processes in a process gro
```

Signals

```
sigaction   -- interpose a signal handler (can include special ``default''
            -- ``ignore'' handlers)
sigprocmask -- change the list of blocked signals
```

Signals

```
wait        -- check for a signal (can be blocking or non-blocking) or ch
waitpid     -- check for a signal from a child process (can be general or
```

Modifying the current process's state

```
chdir       -- change the working directory for a process to dirname
chroot      -- change the root filesystem for a process
```

Modifying the current process's state

```
execve      -- execute another binary in this current process
fork        -- create a new child process running the same binary
clone       -- allows the child to share execution context (unlike fork(2
exit        -- terminate the current process
```

Modifying the current process's state

```
getdtablesize  -- report how many file descriptors this process can have
               -- active simultaneously
```

Modifying the current process's state

```
getgid      -- return the group id of this process
getuid      -- return the user id of this process
getpgid     -- return process group id of this process
getpgrp     -- return process group's group of this process
```

Modifying the current process's state

```
getpid      -- return the process id of this process
getppid     -- return parent process id of this process
getrlimit   -- set a resource limit on this process (core size, cpu time,
            -- data size, stack size, and others)
getrusage   -- find amount of resource usage by this process
```

Modifying the current process's state

```
nice()          -- change the calling process's priority
setpriority()   -- arbitrarily change any process's (or group or user) pr
setpriority()   -- get any process's priorities
```

Communications and Networking

```
socket      -- create a file descriptor (can be either network or local)

bind        -- bind a file descriptor to an address, such a tcp port
listen      -- specify willingness for some number of connections to be
            -- blocked waiting on accept()
accept      -- tell a file descriptor block until there is a new connecti

connect     -- actively connect to listen()ing socket

setsockopt  -- set options on a given socket associated with fd, such out
            -- data, keep-alive information, congestion notification, fin
            -- and so forth (see man tcp(7))
getsockopt  -- retrieve information about options enabled for a given con
```

```
getpeername -- retrieve information about other side of a connection from
getsockname -- retrieve information this side of a connection from fd
```

Others

```
brk          -- allocate memory for the data segment for the
             -- current process

gethostname  -- gets a ``canonical hostname'' for the machine
gettimeofday -- gets the time of day for the whole machine
settimeofday -- sets the time of day for the whole machine
mount        -- attaches a filesystem to a directory and makes it availab
sync         -- flushes all filesystem buffers, forcing changed blocks to
             -- ``drives'' and updates superblocks
futex        -- raw locking (lets a process block waiting on a change
                to a specific memory location)
sysinfo      -- provides direct access from the kernel to:
                    load average
                    total ram for system
```

```
available ram
amount of shared memory existing
amount of memory used by buffers
total swap space
swap space available
number of processes currently in proctable
```

SYS V IPC

```
msgctl      -- SYS V messaging control (uid, gid, perms, size)
msgget      -- SYS V message queue creation/access
msgrcv      -- receive a SYS V message
msgsnd      -- send a SYS V message

shmat       -- attach memory location to SYS V shared memory segment
shmctl      -- SYS V shared memory contrl (uid, gid, perms, size, etc)
shmget      -- SYS V shared memory creation/access
shmdt       -- detach from SYS V shared memory segment
```