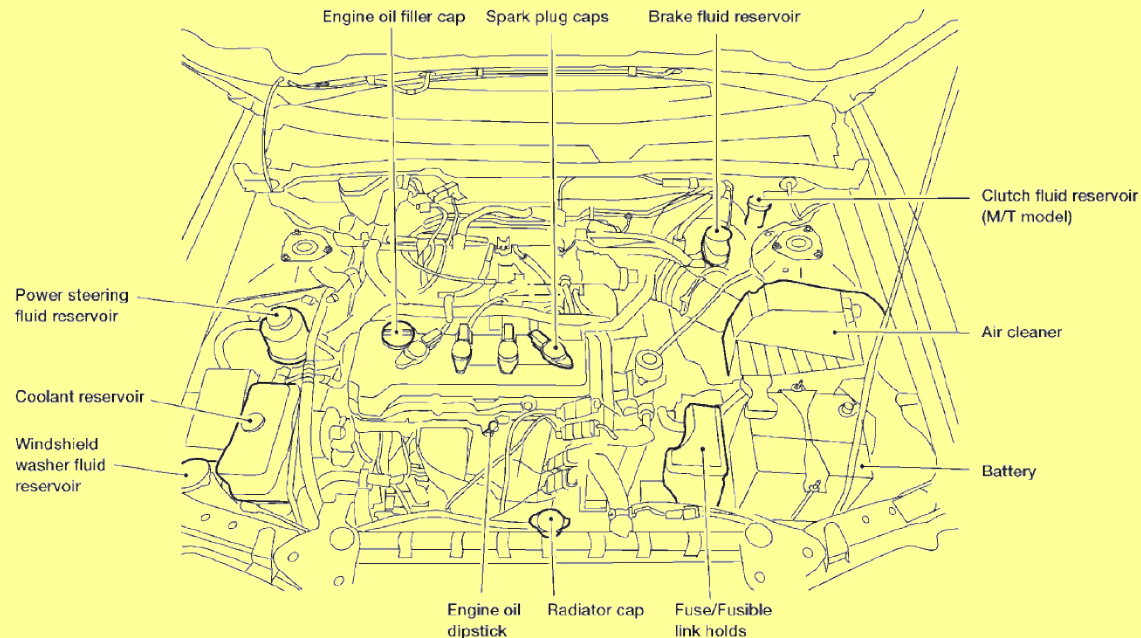


Looking Under The Hood.



2016-08-28, BarcampGR11

Adam Tauno Williams <awilliam@whitemice.org>

Copyright

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. You may obtain a copy of the GNU Free Documentation License from the Free Software Foundation by visiting their Web site or by writing to: Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.



If you find this document useful or further it's distribution we would appreciate you letting us know. I also maintain an Amazon wish list for the above e-mail address.

whoami

Adam Tauno Williams (awilliam@whitemice.org)

GPG: D95ED393

Systems & Network Administrator, Morrison Industries

Active Directory, PostgreSQL, Informix, RabbitMQ, AcuCOBOL,
WAN (Fiber, AT&T ASE, MPLS, Cisco IOS, IPSec), VM-ware,
ISC Bind DNS & ISC DHCP, ZenOSS, AIX, CUPS/IPP, Cyrus IMAP,
SASL, Postfix, ...

Meeting Coordinator, Michigan Association of Railroad Passengers

Developer (Python). OpenGroupware Coils

<http://www.whitemice.org>

<http://www.whitemiceconsulting.com>

<http://www.opengroupware.us>

What is happening?

Related to performance the most common answer is:

nothing

But where is that nothing happening?

Possibly something is choking, somewhere else.

pidstat -d s# / -p pid#

```
awilliam@linux-nysu:~> pidstat -d 1
Linux 3.4.11-2.16-desktop (linux-nysu.site) 02/25/2013
_x86_64_ (8 CPU)
01:50:24 PM          PID    kB_rd/s    kB_wr/s kB_ccwr/s  Command
01:50:25 PM        4395         0.00         4.00         0.00  gnome-
terminal
01:50:27 PM          PID    kB_rd/s    kB_wr/s kB_ccwr/s  Command
01:50:28 PM        3077         0.00        20.00         0.00  telepathy-
gabb1
01:50:28 PM          PID    kB_rd/s    kB_wr/s kB_ccwr/s  Command
01:50:29 PM        4395         0.00         4.00         0.00  gnome-
terminal
```

ccwr = cancelled I/O

Report, every second, the disk I/O of every active process.

pidstat -r for VM stats

```
awilliam@linux-nysu:~> pidstat -d 1 -r
Linux 3.4.11-2.16-desktop (linux-nysu.site) 02/25/2013 _x86_64_(8 CPU)
01:54:51 PM          PID  minflt/s  majflt/s     VSZ     RSS   %MEM  Command
01:54:52 PM        1003     1.96     0.00   24992     896   0.01  smpppd
01:54:52 PM        2425     5.88     0.00 2097388 318724   1.94  gnome-
shell
01:54:52 PM        3578     0.98     0.00 1462624 425412   2.59  firefox
01:54:52 PM       16091     0.98     0.00   14432    1392   0.01  slabtop
01:54:52 PM       16482   586.27     0.00    4900    1240   0.01  pidstat

01:54:51 PM          PID  kB_rd/s  kB_wr/s kB_ccwr/s  Command
01:54:52 PM          PID  minflt/s  majflt/s     VSZ     RSS   %MEM  Command
01:54:53 PM        1003     2.00     0.00   24992     896   0.01  smpppd
01:54:53 PM        2425     3.00     0.00 2097388 318724   1.94  gnome-
shell
01:54:53 PM       16482   602.00     0.00    4900    1280   0.01  pidstat
```

Also “-t” for threads, “-s” for stack, “-w” for context.

Use the man page!

Count the system calls

```
strace -c -p 7774
```

% time	seconds	usecs/call	calls	errors	syscall
63.44	0.019996	40	505		fsync
9.52	0.003000	27	112		fdatasync
9.52	0.003000	6	502		ftruncate
6.97	0.002198	0	9790		read
5.32	0.001676	1	2570	557	open
1.83	0.000576	24	24		brk
1.63	0.000514	0	5628		write
1.03	0.000325	36	9		munmap
0.32	0.000102	0	2121	2040	unlink`

getaddrinfo
select / poll
read / write / fsync
open / close

This example is connecting to a PostgreSQL worker performing a **VACUUM FULL**;

You can connect via PID and record stats until you hit your break key (Ctrl-C, usually)

Watching Systems Calls

```
awilliam@p105s6207:~> strace -p 3431
```

```
Process 3431 attached
```

```
select(7, [3 4], [], NULL, NULL) = 1 (in [3]) <0.004347>
```

```
Read(3, "\262\356\207\35H\315\37\314&I_\356\277\256\310\33I&\30J  
C\320nD!\236\321\33-\364\250\222"... , 8192) = 80 <0.000061>
```

```
select(7, [3 4], [5], NULL, NULL) = 1 (out [5]) <0.000031>
```

```
write(5, "\33[00m\33[00;34mcoils-code\33[00m\r\n\33["... , 33) = 33  
<0.000032>
```

```
select(7, [3 4], [], NULL, NULL) = 1 (in [3]) <0.080674>
```

```
read(3, "\344\377\3266\263\2759z\237\310m\260\336\24w\302\367\273  
k+\354>1\344\203\231\0\0:\212\316"... , 8192) = 176 <0.000034>
```

```
select(7, [3 4], [5], NULL, NULL) = 1 (out [5]) <0.000056>
```

```
write(5, "\33]0;awilliam@aleph:~\7[awilliam@a"... , 41) = 41  
<0.000044>
```


Open Files

(lsof)

```
lsof -p 2977 | cut -c23-
```

```
2w REG 253,0 98580 13500664 /home/awilliam/.config/banshee-1/log
4r CHR 1,9 0t0 1034 /dev/urandom
5u unix 0xffff8801f526ce00
6u 0000 0,9 0 3688 anon_inode
8u unix 0xffff88021f569c00
9u REG 253,0 26275840 13501376 /home/awilliam/.config/banshee-1/banshee.db
10u sock 0,7 0t0 139941 can't identify protocol
16w FIFO 0,8 0t0 141490 pipe
17u unix 0xffff8801e169c4c0 0t0 141493 socket
19u IPv4 134945 0t0 TCP localhost:8089 (LISTEN)
21u IPv4 958178 0t0 TCP 10.66.1.101:38333-
>174.37.70.140-static.reverse.softlayer.com:http (ESTABLISHED)
22u 0000 0,9 0 3688 anon_inode
```

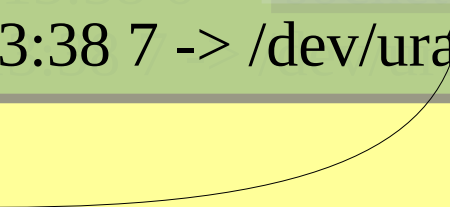
lsof lists all the file like objects open by the specified process.

Open Files

(/proc/{pid#}/fd/)

```
awilliam@linux-nysu:/proc/7253/fd> ls -l
total 0
lrwx----- 1 awilliam users 64 Feb 25 13:38 0 -> /dev/pts/3
lrwx----- 1 awilliam users 64 Feb 25 13:38 1 -> /dev/pts/3
lrwx----- 1 awilliam users 64 Feb 25 13:38 11 -> /tmp/ffidlzD0I
(deleted)
lrwx----- 1 awilliam users 64 Feb 25 11:33 2 -> /dev/pts/3
l-wx----- 1 awilliam users 64 Feb 25 13:38 3 -> /var/log/coins.log
lrwx----- 1 awilliam users 64 Feb 25 13:38 4 -> socket:[71828]
lr-x----- 1 awilliam users 64 Feb 25 13:38 5 -> /dev/urandom
lrwx----- 1 awilliam users 64 Feb 25 13:38 6 -> socket:[64321]
lr-x----- 1 awilliam users 64 Feb 25 13:38 7 -> /dev/urandom
```

inode



Where is the I/O going? strace it!

```
strace -T -e trace=read,write,open,close \  
/bin/cat /etc/passwd > /dev/null
```

```
open("/etc/passwd", O_RDONLY|O_LARGEFILE) = 3 <0.000031>  
read(3, "at:x:25:25:Batch jobs daemon:/va"...  
write(1, "at:x:25:25:Batch jobs daemon:/va"...  
read(3, "", 32768) = 0 <0.000027>
```

File Handle

Bytes Read/Written

Duration

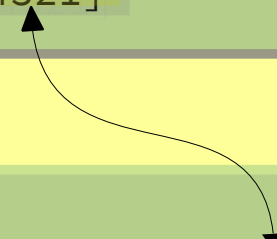
You can attach to a running process with strace by using the **-p *PID*** argument.

The **-T** argument will add the duration of each system call to the output,

Sockets Are Files Too

```
awilliam@linux-nysu:/proc/7253/fd> ls -l
total 0
lrwx----- 1 awilliam users 64 Feb 25 13:38 4 -> socket:[71828]
lrwx----- 1 awilliam users 64 Feb 25 13:38 6 -> socket:[64321]
```

```
awilliam@linux-nysu:~> netstat --tcp -e
Active Internet connections (w/o servers)
Proto Local Address          Foreign Address        State      User      Inode
tcp    localhost:48142        localhost:amqp         ESTABLISHED awilliam  71828
tcp    192.168.1.121:54304    hod-a.mormail.c:ndl-aas ESTABLISHED awilliam  27023
tcp    192.168.1.121:42253    tun.mormail:xmpp-client ESTABLISHED awilliam  26094
tcp    192.168.1.121:58031    moa.mormail.com:pdap-np ESTABLISHED awilliam  81018
...
```



This reveals the end-points of the sockets.

What about my socket?

My IRC clients local end-point is TCP/59475; lsof told me this.

```
ss --extended --processes --info '( sport == :59475 )'
```

```
State      Recv-Q  Send-Q  Local Address:Port      Peer Address:Port
ESTAB      0        0      10.66.1.101:59475      213.179.58.83:ircu
timer:(keepalive,74min,0) users:(("xchat",4085,9))
uid:1000 ino:47314 sk:ffff8800c543b100
ts sack cubic wscale:7,6 rto:352 rtt:149.875/1 ato:40 cwnd:5
send 386.5Kbps rcv_rtt:797.25 rcv_space:22626
```

Data You've Sent

ACK Time Out
Round-Trip Time

Retransmission Time Out

TCP Window Scale

Process I/O

```
cat /proc/2977/io
```

```
rchar: 24396601  
wchar: 65012571  
syscr: 23966  
syscw: 29911  
read_bytes: 31813632  
write_bytes: 56659968  
cancelled_write_bytes: 2625536
```

For profiling and performance do not neglect looking at I/O; everyone likes to look at CPU and memory, I/O is more likely to be the choke point.

System I/O

dstat

```
-----total-cpu-usage----- -dsk/total- -net/total- ----paging-- ----system--
usr  sys  idl  wai  hiq  siq| read  writ| recv  send|  in   out  | int  csw
  0   1  98   0   0   0| 40k  194k|   0    0 |   0   0  | 541  318
  0   0 100   0   0   0|   0    0 | 429B 429B|   0   0  |1027   44
  0   1  99   0   0   0|   0   32k| 126B 338B|   0   0  | 972   37 2
  0   7  92   1   0   0|   0  456k| 297B 461B|   0   0  | 992  642
```

Using swap (paging) space is not a performance problem; paging is.

Logging Detailed System Wide I/O

```
echo "1" > /proc/sys/vm/block_dump
```

```
[ 2032.934178] postmaster(11528): READ block 5058592 on dm-3 (16 sectors)
[ 2032.934200] postmaster(11528): READ block 5058624 on dm-3 (32 sectors)
[ 2032.934240] postmaster(11528): READ block 3172800 on dm-3 (16 sectors)
[ 2032.945328] banshee-1(11267): dirtied inode 1051864 (banshee.db-
journal) on dm-0
[ 2032.945336] banshee-1(11267): dirtied inode 1051864 (banshee.db-
journal) on dm-0
[ 2033.042671] python(11518): READ block 9017928 on dm-2 (32 sectors)
[ 2033.055771] python(11518): dirtied inode 267260 (expatbuilder.pyc) on
dm-2
[ 2033.055808] python(11518): READ block 9017960 on dm-2 (40 sectors)
[ 2033.412972] nautilus(11078): dirtied inode 410492
```

Logs to the kernel ring buffer.

```
echo "0" > /proc/sys/vm/block_dump
```


How many network connections?

```
sudo ss --summary
```

```
Total: 639 (kernel 707)
```

```
TCP: 46 (estab 18, closed 9, orphaned 0, synrecv 0, timewait 9/0),  
ports 40
```

Transport	Total	IP	IPv6
*	707	-	-
RAW	2	2	0
UDP	19	12	7
TCP	37	30	7
INET	58	44	14
FRAG	0	0	0

Detailed Network Statistics

```
netstat --interfaces
```

Iface	MTU	Met	RX-OK	RX-ERR	RX-DRP	RX-OVR	TX-OK	TXERR	TXDRP	TXOVR	Flg
eth0	1500	0	46425	0	0	0	38928	0	0	0	BMRU
lo	16436	0	474	0	0	0	474	0	0	0	LRU
Vmnet1	1500	0	0	0	0	0	36	0	0	0	BMRU
vmnet8	1500	0	0	0	0	0	36	0	0	0	BMRU

Boring! And not really useful; lack of interface errors does not mean your network is feeling well.

```
netstat --statistics
```

```
Tcp:  
 423 active connections openings  
 12 passive connection openings  
 1 failed connection attempts  
 10 connection resets received  
 8 connections established  
 24 segments retransmitted  
 0 bad segments received.  
 206 resets sent
```

Much more interesting,
far more useful.