Tips and Tricks with the proc filesystem

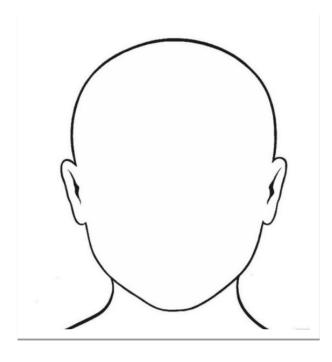


INTRODUCTION

Julius Schwartzenberg

jschwartzenberg@luxoft.com

http://luxoft-training.ru http://luxoft-training.com



TRAINING ROADMAP: STRUCTURE & PREPARATION

- single 1-2 hour session
- Practical, please open a terminal to a GNU/Linux system
- Use the lower right area of the screen
- Example daemon, mini_httpd: https://github.com/jschwartzenberg/mini httpd
- Example media files: https://media.xiph.org/





Preparation

Get a copy of mini_httpdGit clone URLs:

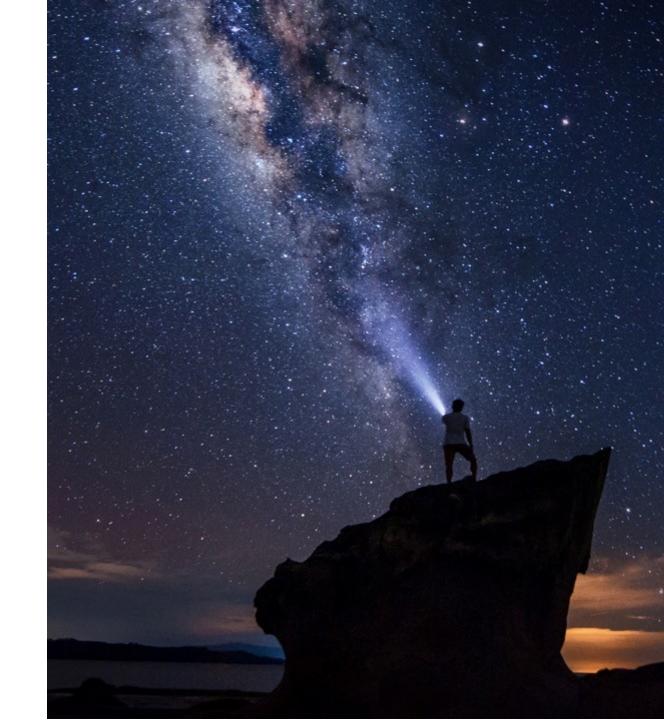
- git@github.com:jschwartzenberg/ mini_httpd.git
- https://github.com/jschwartzenberg/ mini_httpd.git

Should build with a simple make



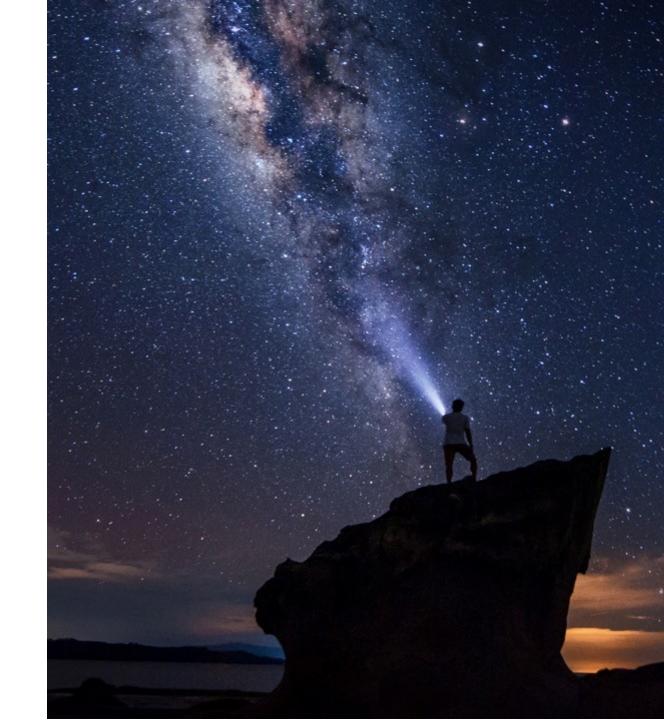
Introduction

- Introduced with UNIX (8th edition),
 further enhanced in Plan 9, BSD 4.4
- Essential part of Solaris and Linux
- Continues on the principle"everything is a file"
- Information about the system and individual processes



Problem 1: Is my process running? (1/2)

- Daemon crashes sometimes
- Daemon should be started only once



Problem 1: Is my process running? (2/2)

- Every running process is represented below /proc
- Daemons should create a pid file on start-up

```
#!/bin/bash
PID_FILE="/tmp/httpd.pid"
PID=$(<$PID_FILE)
echo The Process ID is $PID
if [ ! -e $PID_FILE ] || [ ! -d "/proc/$PID" ];
then
  echo "Not running!"
else
  echo "Everything is fine!"
```

Problem 2: How was my process started? (1/2)

- You are looking at an unknown system
- Analyse its processes
- Discover its configuration



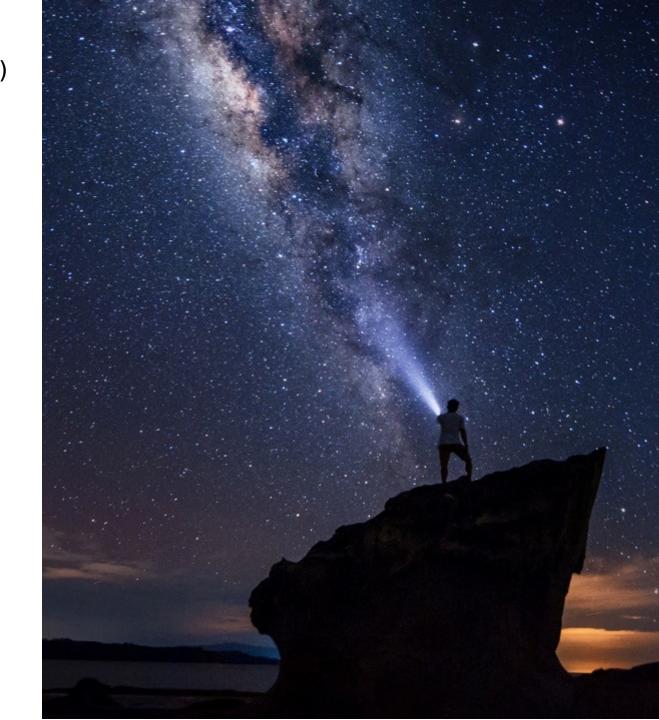
Problem 2: How was my process started? (2/2)

- Every running process has its own environment
- It can be read through /proc
- Individual variables are terminated with a NUL
- The command line used to start a process is also available

```
julius@pols:~$ ps ax|grep dm
  973 ?
              Ssl 0:00 /usr/bin/sddm
 1008 tty1 Ssl+ 8:34 /usr/lib/xorg/Xorg -noli
-a1f2-4e6b7d5d6814} -background none -noreset -disp
 1199 ?
                     0:00 /usr/lib/x86_64-linux-gr
cb-c5ba-40ea-af6f-7f2b5065e216 --id 1 --start /usr/
16332 pts/3 S+ 0:00 grep --color=auto dm
julius@pols:~$ sudo strings /proc/973/environ
LANG=nl NL.UTF-8
PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/
INVOCATION_ID=2c89e4f65d6b4b15b6722bea7499d11d
JOURNAL_STREAM=9:31538
julius@pols:~$
```

Problem 3: Where is that process logging? (1/2)

- You are looking at an unknown system
- No idea if any logging is in place



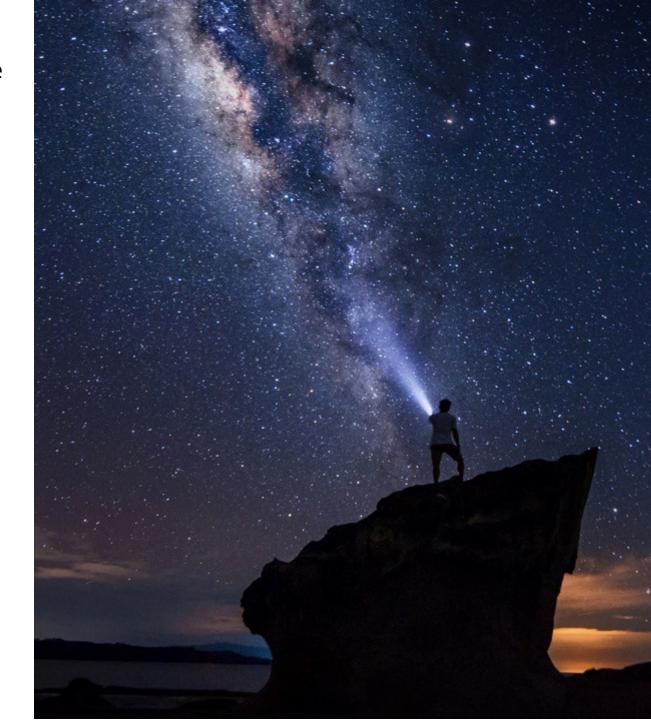
Problem 3: Where is that process logging? (2/2)

- Every running process has stdio: stdin, stdout, stderr
- Logging to stdout & stderr
- File descriptors are visible below /proc
- stdin: 0; stdout: 1; stderr: 2
- ss tool to investigate sockets

```
julius@pols:~$ ps ax|grep cron
  850 ? Ss
                     0:00 /usr/sbin/cron -f
20870 pts/3 S+
                     0:00 grep --color=auto cron
julius@pols:~$ sudo ls -o /proc/850/fd
totaal 0
lr-x----- 1 root 64 okt 28 12:09 0 -> /dev/null
lrwx----- 1 root 64 okt 28 12:09 1 -> 'socket:[26413]'
lrwx----- 1 root 64 okt 28 12:09 2 -> 'socket:[26413]'
lrwx----- 1 root 64 okt 28 12:09 3 -> /run/crond.pid
julius@pols:~$ sudo ss|grep 26413
u str
                ESTAB
             /run/systemd/journal/stdout 26901
u str
                ESTAB
          * 26901
julius@pols:~$
```

Problem 4: Does the running process match the executable on disk? (1/3)

- Updates were installed
- Version on disk might be different from the running one
- The previous version is not referenced anymore from the filesystem
- It is still somewhere



Problem 4: Does the running process match the executable on disk? (2/3)

- There is a pointer to the executable that is running below /proc
- Multiple references can point to the same filesystem object
- An object is only removed from the filesystem when there are no references left



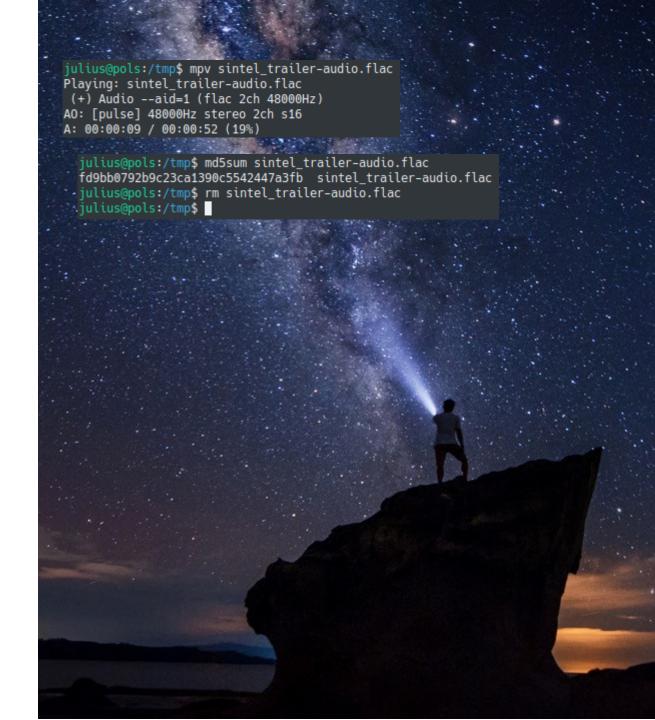
Problem 4: Does the running process match the executable on disk? (3/3)

- Start mini_httpd
- Find it below /proc
- What happens with mv?
- What happens with rm?
- What happens when replacing the executable?
- Executing from inside /proc?

```
julius@pols:/tmp/mini_httpd$ ./mini_httpd -i /tmp/httpd.pid -p 8080
bind: Address already in use
julius@pols:/tmp/mini_httpd$ ls -o /proc/$(</tmp/httpd.pid)/exe</pre>
lrwxrwxrwx 1 julius 0 okt 29 09:51 /proc/31936/exe -> /tmp/mini_httpd/mini_httpd/
ulius@pols:/tmp/mini_httpd$ mv mini_httpd mini_httpd-moved
 ulius@pols:/tmp/mini_httpd$ ls -o /proc/$(</tmp/httpd.pid)/exe</pre>
lrwxrwxrwx 1 julius 0 okt 29 09:51 /proc/31936/exe -> /tmp/mini_httpd/mini_httpd-moved
 ulius@pols:/tmp/mini_httpd$ rm mini_httpd-moved
 ulius@pols:/tmp/mini_httpd$ ls -o /proc/$(</tmp/httpd.pid)/exe</pre>
lrwxrwxrwx 1 julius 0 okt 29 09:51
 ulius@pols:/tmp/mini_httpd$ /proc/31936/exe -h
usage: /proc/31936/exe [-C configfile] [-D] [-S] [-E certfile] [-Y cipher] [-p port] [-d dir] [-d
d data_dir] [-c cgipat] [-u user] [-h hostname] [-r] [-v] [-l logfile] [-i pidfile] [-T charset] [
-P P3P] [-M maxage]
 ulius@pols:/tmp/mini_httpd$ /proc/31936/exe -i /tmp/httpd2.pid -p 8081
bind: Address already in use
ulius@pols:/tmp/mini_httpd$ cat /tmp/httpd2.pid
32125
ulius@pols:/tmp/mini_httpd$
```

Problem 5: File recovery for open files (1/2)

- You accidentally deleted a song.
- But it is still playing!
- How we get to it?



Problem 5: File recovery for open files (2/2)

- Whenever a process opens a file, this is handled through a file descriptor
- If your process might complete, you might want to suspend it: Ctrl+Z

```
julius@pols:/tmp$ mpv sintel_trailer-audio.flac
Playing: sintel_trailer-audio.flac
(+) Audio --aid=1 (flac 2ch 48000Hz)
AO: [pulse] 48000Hz stereo 2ch s16
A: 00:00:09 / 00:00:52 (19%)
                             mpv sintel_trailer-audio.flac
[1]+ Stopped
 julius@pols:/tmp$ ps ax|grep mpv
                               sintel_trailer-audio.flac
   608 pts/2
              Τl
                       0:00
  1286 pts/10 S+
                       0:00 grep --color=auto
 julius@pols:/tmp$ ls -o /proc/608/fd
 totaal 0
 lrwx----- 1 julius 64 okt 29 10:03 0 -> /dev/pts/2
 lrwx----- 1 julius 64 okt 29 10:03
                                      -> /dev/pts/2
 lr-x---- 1 julius 64 okt 29 10:03
 l-wx----- 1 julius 64 okt 29 10:03
 lrwx----- 1 julius 64 okt 29 10:03
 lrwx----- 1 julius 64 okt 29 10:03
 lrwx----- 1 julius 64 okt 29 10:03
 lr-x---- 1 julius 64 okt 29 10:03
  l-wx----- 1 julius 64 okt 29 10:03
 lrwx----- 1 julius 64 okt 29 10:03
 lr-x---- 1 julius 64 okt 29 10:03
 l-wx----- 1 julius 64 okt 29 10:03
  lr-x----- 1 julius 64 okt 29 10:03
 l-wx----- 1 julius 64 okt 29 10:03
 lr-x---- 1 julius 64 okt 29 10:03
 l-wx----- 1 julius 64 okt 29 10:03
 lr-x---- 1 julius 64 okt 29 10:03
 julius@pols:/tmp$ cp /proc/608/fd/9 ./recovered.flac
  julius@pols:/tmp$ md5sum ./recovered.flac
  fd9bb0792b9c23ca1390c5542447a3fb
```

Problem 6: Read stdin when a file is expected (1/2)

- Your want to embed a file in your shell script and pass it as input to your program
- Many programs can read stdin, but not all
- Example: wget --cacertificate=<file>



Problem 6: Read stdin when a file is expected (2/2)

stdin in is represented as a file

#!/bin/bash

CERT="----BEGIN CERTIFICATE----

MIIDeTCCAmGgAwIBAgIJAPziuikCTox4MA0GCSqGSIb3DQEBCwUAMGIxCzAJBgNVBAYTAlVTMRMwEQYDVQQIVQQHDA1TYW4gRnJhbmNpc2NvMQ8wDQYDVQQKDAZCYWRTU0wxFTATBgNVBAMMDCouYmFkc3NsLmNvbTAeFw0xMDgyMzQxNTJaMGIxCzAJBgNVBAYTAlVTMRMwEQYDVQQIDApDYWxpZm9ybmlhMRYwFAYDVQQHDA1TYW4gRnJhYWRTU0wxFTATBgNVBAMMDCouYmFkc3NsLmNvbTCCASIwDQYJKoZIhvcNAQEBBQADggEPADCCAQoCggEBAMIEaEmwIrX5lZ6xKyx2PmzAS2BMTOqytMAPgLaw+XLJhgL5XEFdEyt/

ccRLvOmULlA3pmccYYz2QULFRtMWhyefdOsKnRFSJiFzbIRMeVXk0WvoBj1IFVKtsyjbqv9u/2CVSndrOfEkq71FdIzSOciccfCFHpsKOo3St/qbLVytH5aohbcabFXRNsKEqveww9HdFxBIuGa+RuT5q0iBikusbpJHAwnndAcgCskgjZjFeEU4EFy+b+a1SYQCeFxxC7c3DvaRhBB0VVfPlkPz0sw61865MaTIbRyoUCAwEAAaMyMDAwCQAaggwqLmJhZHNzbC5jb22CCmJhZHNzbC5jb20wDQYJKoZIhvcNAQELBQADggEBAGlwCdbPxflZfYOaukZGCa95w7fChXvP3YkE3UYUE7mupZ0eg4ZILr/A0e7JQDsgIu/

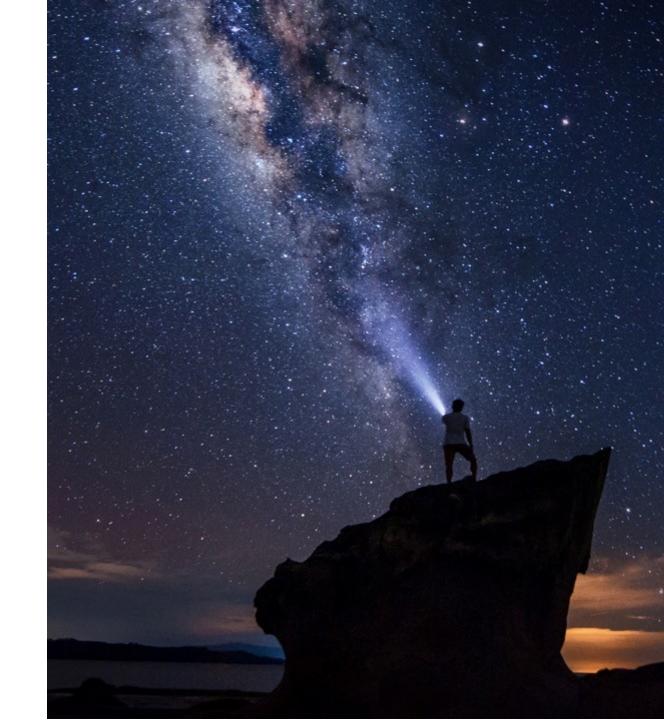
SRTUE0domCKgPZ8v99k3Avka4LpLK51jHJJK7EFgo3ca2nldd97GM0MU41xHFk8qaK1tWJkfrrfcGwDJ4GQP TlRX1CLd8ufWhhiwW0W75Va5AEnJuqpQrKwl3KQVewGj67WWRgLfSr+4QG1mNvCZb2CkjZWmxkGPuoP40/ y7Yu50FqxP5tAjj4YixCYTWEVA0pmzIzqBq+JIe3PdRy27T0asqQW/F4TY61Yk=----END CERTIFICATE-

wget --ca-certificate=/dev/stdin https://self-signed.badssl.com/ -0 - <<< \$CERT



Problem 7: Write stdout/stderr when a file is expected (1/2)

- You want to see the output in your terminal
- mini_httpd -l <logfile>



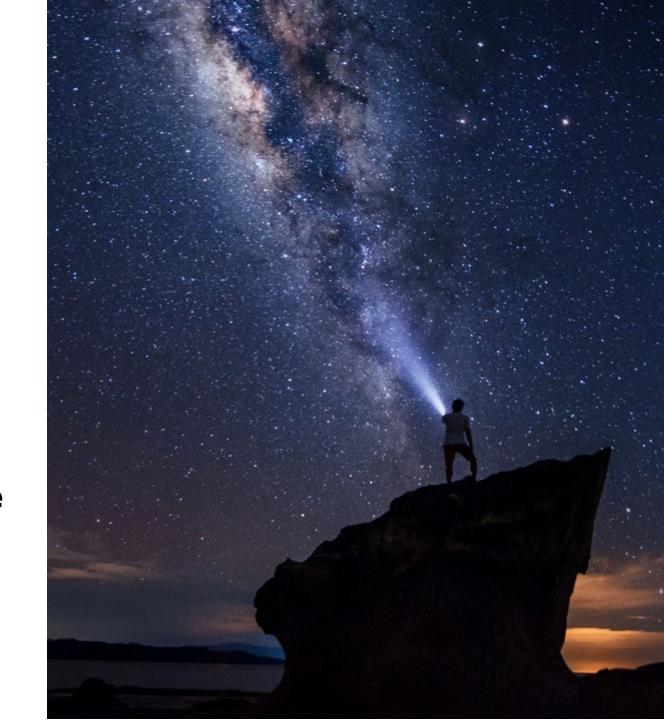
Problem 7: Write stdout/stderr when a file is expected (2/2)

Solution is similar to the one from problem 6

```
ulius@pols:/tmp/mini_httpd$ ./mini_httpd -h
usage: ./mini_httpd [-C configfile] [-D] [-S] [-E certfile] [-Y cipher] [-p port] [-d ata_dir] [-c cgipat] [-u user] [-h hostname] [-r] [-v] [-l logfile] [-i pidfile] [-T cha
P3P] [-M maxage]
julius@pols:/tmp/mini_httpd$ ./mini_httpd -i /tmp/httpd.pid -l /dev/stderr -p 8080
bind: Address already in use
julius@pols:/tmp/mini_httpd$ curl localhost:8080/wrong > /dev/null 2>&1
127.0.0.1 - - [29/0ct/2020:10:20:32 +0100] "GET /wrong HTTP/1.1" 404 - "" "curl/7.58.0"
 julius@pols:/tmp/mini httpd$
```

Problem 8: Determine progress of a running program (1/2)

- A process is running over a large file that will take a long time
- You have no idea how much time is left
- Usually processes read a file from start to end



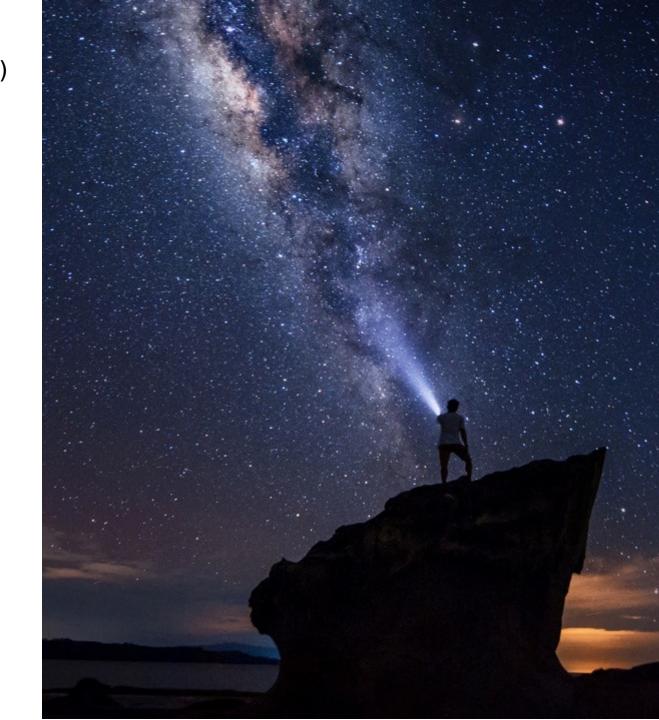
Problem 8: Determine progress of a running program (2/2)

- In addition to
 /proc/<pid>/fd, there is
 also /proc/<pid>/fdinfo
- Look for the value of "pos"

```
julius@pols:/fritz.nas/gemu$ ls -o
totaal 6519244
-rw-r--r-- 1 1019 6675685734 okt 18 10:51 rhel5.gcow2.gz
julius@pols:/fritz.nas/qemu$ md5sum rhel5.qcow2.gz &
julius@pols:/fritz.nas/gemu$ ls -o /proc/6450/fd/
lrwx----- 1 julius 64 okt 29 10:36 0 -> /dev/pts/10
lrwx----- 1 julius 64 okt 29 10:36 1 -> /dev/pts/10
lrwx----- 1 julius 64 okt 29 10:36 2 -> /dev/pts/10
lr-x---- 1 julius 64 okt 29 10:36 3 ->
julius@pols:/fritz.nas/qemu$ cat /proc/6450/fdinfo/3
flags: 0100000
julius@pols:/fritz.nas/qemu$ bc -l <<< "243531776/6675685734*100"
julius@pols:/fritz.nas/gemu$ cat /proc/6450/fdinfo/3
flags: 0100000
mnt id: 467
julius@pols:/fritz.nas/gemu$ bc -l <<< "550043648/6675685734*100"
8.23950781862853941200
julius@pols:/fritz.nas/qemu$
```

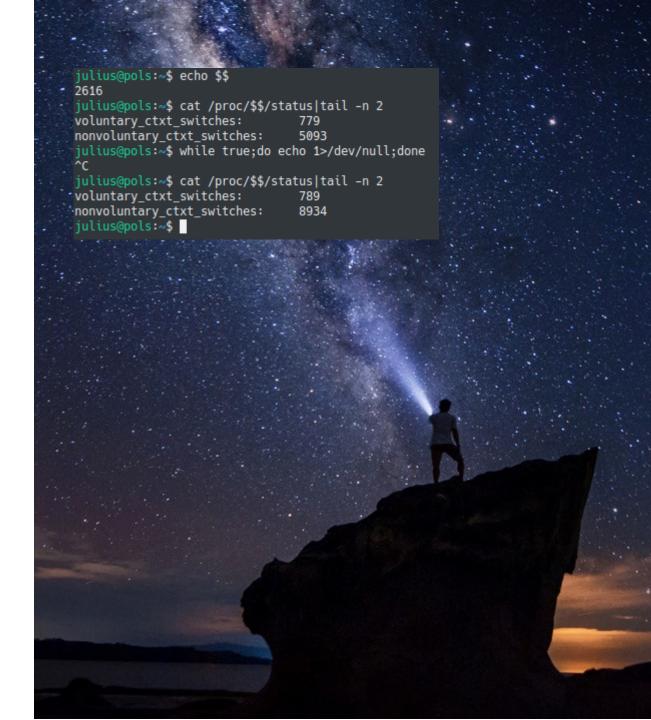
Problem 9: Is a process CPU- or IO-bound? (1/2)

- A process runs slow
- Where is the bottleneck?



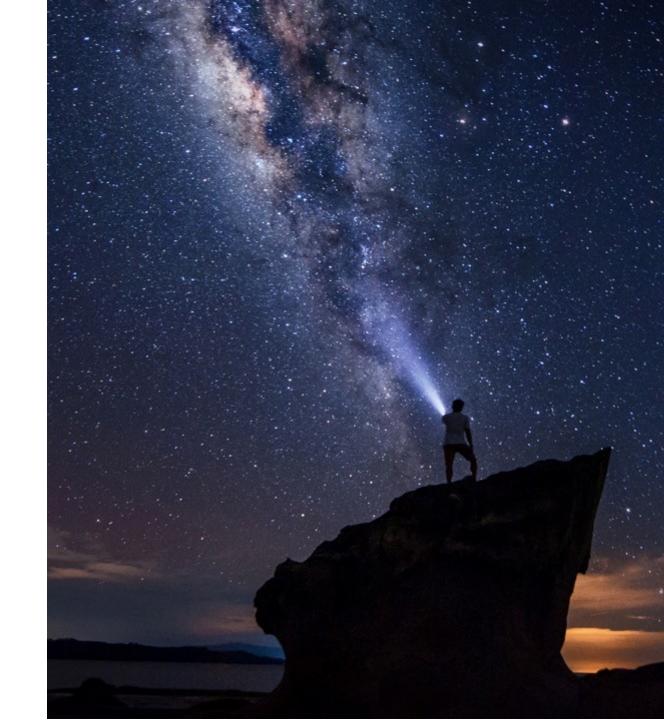
Problem 9: Is a process CPU- or IO-bound? (2/2)

- Processes that use IO intensively will make more system calls
- A system call implies a context switch
- Processes that do only calculation will not request a context switch
- Context switches are exposed via /proc/<pid>/status



Problem 10: Which libraries are loaded? (1/2)

- Why is the program crashing?
- Might be be loading incompatible libraries at runtime?



Problem 10: Which libraries are loaded? (2/2)

- /proc provides insight in the memory map
- It shows which libraries are loaded
- Take a look at /proc/<pid>/maps

```
root@fritz:/var/mod/root#
 774 root
                           [avm_debugd]
                          [avmnet workqueu]
 901 root
 931 root
                          [avmnet_timer]
 1021 root
                          [avm connect eve]
                          avmipcd
 2610 root
                3496 S
                          avm2fiber xdsld -d
 2633 root
                3848 S
                   0 RW< [avm dect thread]
 2659 root
                          /bin/avmnexusd
 2671 root
                6332 S
                          /usr/bin/avm/ctlmgr
 2829 root
               23452 S
                          /bin/avmike
 3084 root
                6164 S
                1300 S
19653 root
                          {busybox} grep avm
root@fritz:/var/mod/root# cat /proc/2633/maps
00400000-00404000 r-xp 00000000 07:00 2705
                                                 /usr/sbin/avm2fiber xdsld
00413000-00414000 rw-p 00003000 07:00 2705
                                                 /usr/sbin/avm2fiber xdsld
                                                 /SYSVfa00063f (deleted)
77a5e000-77a6c000 rw-s 00000000 00:04 0
                                                 /lib/libgcc_s.so.1
77a6c000-77a7b000 r-xp 00000000 07:00 1645
77a7b000-77a8a000 ---p 00000000 00:00 0
                                                 /lib/libgcc_s.so.1
77a8a000-77a8b000 rw-p 0000e000 07:00 1645
                                                 /lib/libdputil.so.2.0.0
77a8b000-77aa4000 r-xp 00000000 07:00 1594
77aa4000-77ab4000 ---p 00000000 00:00 0
                                                 /lib/libdputil.so.2.0.0
77ab4000-77ab5000 r--p 00019000 07:00 1594
                                                 /lib/libdputil.so.2.0.0
77ab5000-77ab8000 rw-p 0001a000 07:00 1594
77ab8000-77aba000 rw-p 00000000 00:00 0
                                                 /lib/libewnwlinux.so.2.0.0
```

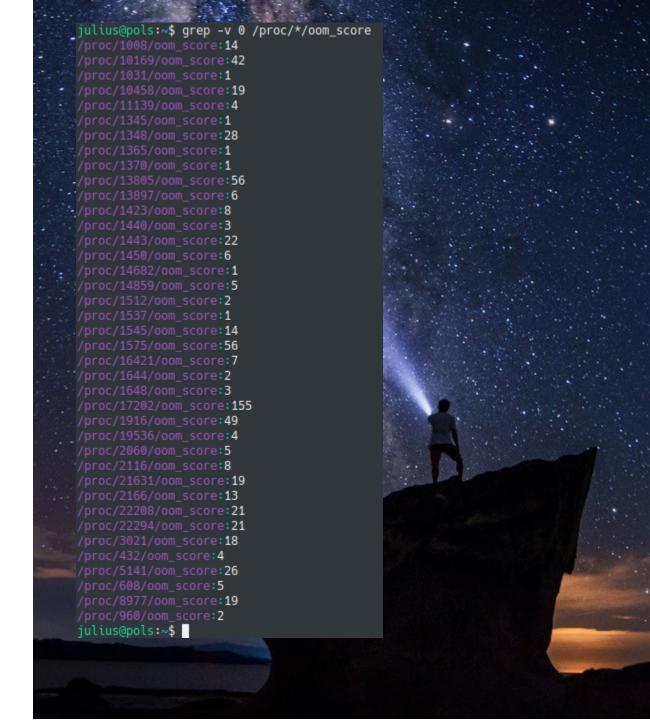
Problem 11: Which process is using so much memory? (1/2)

- Memory usage on a system is very high
- Which process is the most likely culprit?
- The Out-Of-Memory (OOM)
 killer might kick in when it is
 too late



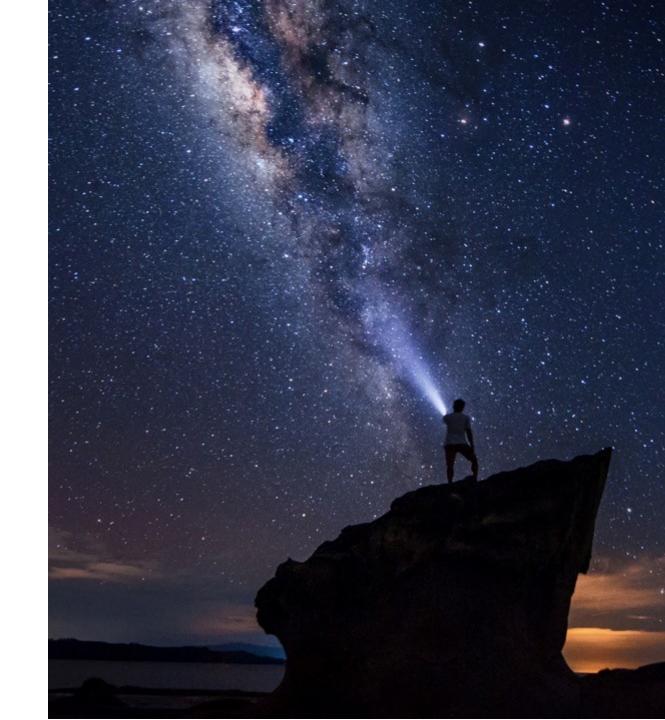
Problem 11: Which process is using so much memory? (2/2)

- The OOM killer keeps track
- The scores are available below /proc



Problem 12: Writing to other terminals (1/2)

- Can you hijack one terminal from another?
- Can one script output to multiple terminal windows?



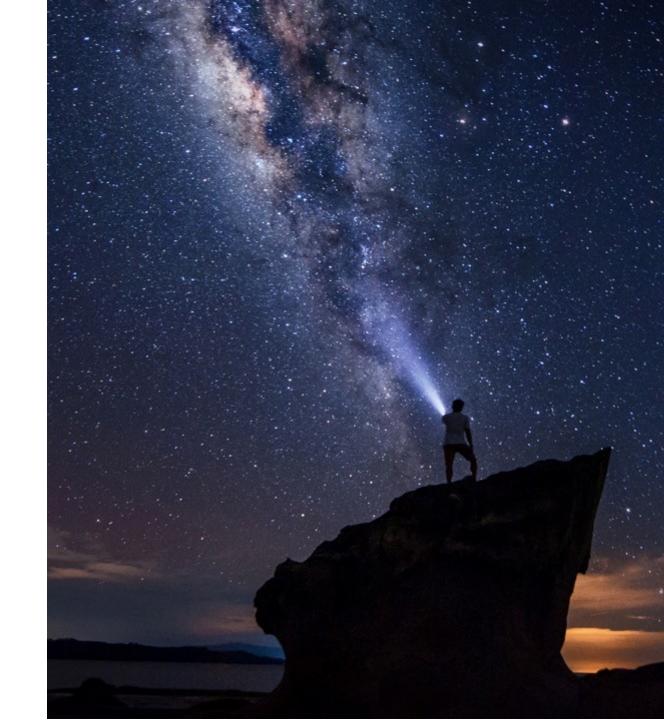
Problem 12: Writing to other terminals (2/2)

- A terminal is attached to a pseudo terminal device
- Look at its file descriptors
- What happens when you write into a pseudo terminal device?

```
#!/bin/bash
konsole &
KONSOLE PID 1=$!
sleep 2
for f in /proc/$KONSOLE PID 1/fd/*; do
 FBASENAME=$(basename "$f")
 if [ $FBASENAME != 0 ] && [ $FBASENAME != 1 ] && [ $FBASENAME != 2 ]; then
   if [[ $(readlink $f) = *"/dev/pts"* ]]; then
      KONSOLE1 FD PTS=$f
   fi
 fi
done
sleep 2
echo "Hi on Konsole One!!" >> $KONSOLE1_FD_PTS
sleep 2
konsole &
KONSOLE PID 2=$!
sleep 2
for f in /proc/$KONSOLE_PID_2/fd/*; do
 FBASENAME=$(basename "$f")
 if [ $FBASENAME != 0 ] && [ $FBASENAME != 1 ] && [ $FBASENAME != 2 ]; then
   if [[ $(readlink $f) = *"/dev/pts"* ]]; then
      KONSOLE2_FD_PTS=$f
   fi
 fi
done
echo "Hi on Konsole Two!!" >> $KONSOLE2 FD PTS
sleep 2
echo "Bye on Konsole One!!" >> $KONSOLE1_FD_PTS
sleep 2
kill $KONSOLE_PID_1
echo "Bye on Konsole Two!!" >> $KONSOLE2_FD_PTS
sleep 2
kill $KONSOLE PID 2
```

Problem 13: Mounting a remote /proc? (1/2)

Can you mount /proc remotely?



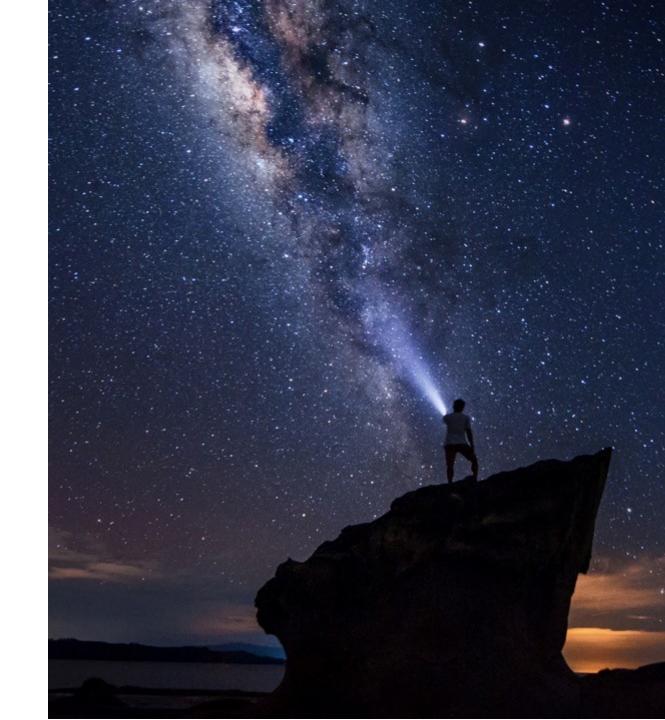
Problem 13: Mounting a remote /proc? (2/2)

- NFS does not work
- What about SSHFS?

```
[julius@hand ~]$ sshfs raspberrypi:/proc mnt
julius@raspberrypi's password:
[julius@hand ~]$ ls mnt
        1605
               24
                       48
                                   crypto
10819
       1686
               24021
                       49
                                   devices
        17
               24024
                       493
                                                   pagetypeinfo
11137
               241
                       494
                                   diskstats
                                                   partitions
11766
       18905
               243
                       541
                                   driver
                                                   sched_debug
11771
       18910
               253
                       551
                                                   schedstat
                                   execdomains
               27186
                                   fb
11999
       18924
                                   filesystems
        18925
               27187
                                                   slabinfo
12
12055
       18926
               27353
                       7193
                                   fs
                                                   softirgs
12290
       18927
               274
                                                   stat
                                   interrupts
       19
               277
                       815
12675
                                   iomem
                                                   swaps
12679
       19503
               289
                       816
                                   ioports
                                                   SVS
       19507
12680
                                   ira
                                                   sysrq-trigger
12702
       19893
               36
                       9318
                                   kallsyms
                                                   sysvipc
               37
                       9549
12727
                                   keys
                       9591
                                                   timer list
12733
                                   key-users
       211
               39
                       9763
12738
                                   kmsq
                                                   tty
       21173
                       9901
                                                   uptime
12740
                                   kpagecgroup
12742
       21190
                       asound
                                   kpagecount
                                                   version
       220
               414
                       buddyinfo
                                   kpageflags
                                                   vmallocinfo
13
14
       22252
               42
                                   latency stats
                       bus
                                                   vmstat
       226
               428
14137
                                   loadavg
                                                   zoneinfo
                       cgroups
       22723
               432
                       cmdline
                                   locks
14140
1441
       23
               439
                       consoles
                                   meminfo
       236
               44
                                  misc
15
                       cpu
       23984
                                   modules
16
               47
                       cpuinfo
[julius@hand ~]$
```

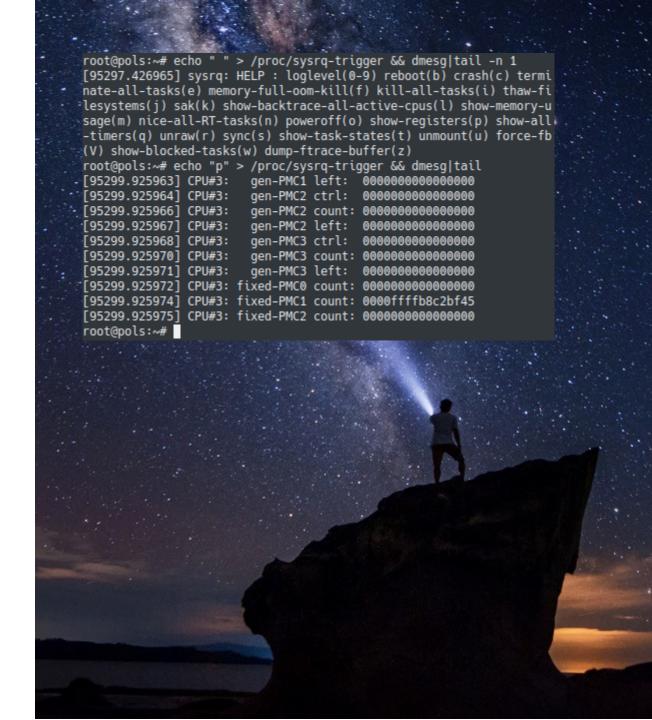
Problem 14: Rebooting or shutting down (1/2)

- The userland shutdown and reboot binaries are missing
- Logged on remotely (no SYSRQ key)



Problem 14: Rebooting or shutting down (2/2)

- Magic SysRq is also possible via /proc/sysrq-trigger
- Be careful!!
- Alt+SysRq+Space shows available triggers



The End: Questions?

