Crash Dump Analysis
Command line tools

Jakub Jermář
Martin Děcký
Overview

- Solaris command line tools for monitoring system activity
  - Processor utilization, saturation and system load
  - Processor and process statistics
  - Monitoring system memory activity
  - Tracing syscalls
  - Displaying stack
  - Displaying process' address space
Literature

- Richard McDougall, Jim Mauro, Brendan Gregg: *Solaris Performance and Tools: DTrace and MDB Techniques for Solaris 10 and OpenSolaris*
Terminology

- **User time** is time spent by the processor executing in user space (i.e. instructions of the user program)
- **System time** is time spent by the processor executing in kernel context (i.e. syscalls and kernel threads)
- **Idle time** is the rest of the time when the processor is not executing any threads
Terminology (2)

- \(\%\text{usr}\) – percentage of user time
- \(\%\text{sys}\) – percentage of system time
- \(\%\text{idle}\) – percentage of idle time

\[\%\text{usr} + \%\text{sys} + \%\text{idle} = 100\]
Terminology (3)

- **Processor utilization**
  - Percentage of time spent by the processor(s) doing useful work
  - utilization = %sys + %usr

- **Processor saturation**
  - How much more work is there than the processor(s) can currently handle
  - One possible measure of processor saturation is the size of the run queue(s)
• Processor utilization
  ▪ Actually not a very good indicator of system health
  ▪ 100% is usually okay, gradual decline in performance
  ▪ <100% means wasted CPU cycles, but also a reserve for a sudden performance peak
  ▪ Measuring interval is important as utilization is averaged across the whole time
Drawing conclusions (2)

- Processor saturation
  - Gradual performance degradation if sustained non-zero processor saturation
  - Gives an estimate of a speedup if more processors were added to the system
  - Important role of the measuring interval because of short periods of saturation
    - Saturation is the total # of threads in RQ
    - Utilization is in %
      - Can give non-zero saturation and idle time
Observing load indicators

- `vmstat 5`

<table>
<thead>
<tr>
<th>kthr</th>
<th>memory</th>
<th>page</th>
<th>disk</th>
<th>faults</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>r b w swap free re mf pi po fr de sr s1 s2 -- --</td>
<td>in sy cs us sy id</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0 3540164 362860 0 3 0 0 0 0 1 2 -0 0 0 655 8803 819 1 28 70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0 3474752 298976 2 12 0 0 0 0 0 0 0 0 0 0 655 24510 835 4 38 58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0 3474652 298972 0 1 0 0 0 0 0 8 0 0 0 669 25881 902 4 38 57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

non-saturated, mostly idle system

- kthr:r showed 0 threads waiting for CPU
  - Total for all processors!!!
- Utilization between 30 – 40 %
Observing load indicators (2)

- **psrinfo**
  
  0  on-line  since 05/04/2009 12:21:06
  1  on-line  since 05/04/2009 12:21:09
  2  on-line  since 05/04/2009 12:21:09
  3  on-line  since 05/04/2009 12:21:09

- **uptime**
  
  2:10pm  up 6 days 1:48, 4 users, load average: 1.14, 1.18, 1.18

- Load averages for last 1, 5 and 15 minutes
  
  - ~ Average number of running and runnable threads on all processors
Observing load indicators (3)

- **sar -q 1 5**

SunOS zulu.ms.mff.cuni.cz 5.11 snv_101b
i86pc 05/10/2009

<table>
<thead>
<tr>
<th>Time</th>
<th>runq-sz</th>
<th>%runocc</th>
<th>swpq-sz</th>
<th>%swpocc</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:51:26</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15:51:31</td>
<td>1.0</td>
<td>0</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>15:51:36</td>
<td>1.5</td>
<td>0</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>15:51:41</td>
<td>1.3</td>
<td>0</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>15:51:46</td>
<td>1.0</td>
<td>0</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>15:51:51</td>
<td>1.2</td>
<td>36</td>
<td>0.0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **runq-sz**
  - Saturation

- **%runocc**
  - Run queue occupancy
  - Percentage of time when runq-sz was non-zero
100 Hz sampling

- Tools like `vmstat` work with data sampled at 100 Hz
  - Possibility to lose information if some activity starts and completes between two 100 Hz ticks
- High resolution timers are used internally in Solaris to support microstate accounting
Observing processors

- **mpstat 1**

<table>
<thead>
<tr>
<th>CPU</th>
<th>minf</th>
<th>mjf</th>
<th>xcal</th>
<th>intr</th>
<th>ithr</th>
<th>csw</th>
<th>icsw</th>
<th>migr</th>
<th>smtx</th>
<th>srw</th>
<th>syscl</th>
<th>usr</th>
<th>sys</th>
<th>wt</th>
<th>idl</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>152</td>
<td>0</td>
<td>7</td>
<td>416</td>
<td>202</td>
<td>138</td>
<td>10</td>
<td>13</td>
<td>7</td>
<td>3</td>
<td>2566</td>
<td>1</td>
<td>50</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>1</td>
<td>145</td>
<td>0</td>
<td>7</td>
<td>220</td>
<td>1</td>
<td>209</td>
<td>10</td>
<td>17</td>
<td>8</td>
<td>3</td>
<td>2514</td>
<td>2</td>
<td>29</td>
<td>0</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>203</td>
<td>0</td>
<td>10</td>
<td>19</td>
<td>3</td>
<td>271</td>
<td>10</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>1780</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>81</td>
</tr>
<tr>
<td>3</td>
<td>157</td>
<td>0</td>
<td>8</td>
<td>15</td>
<td>0</td>
<td>239</td>
<td>8</td>
<td>16</td>
<td>8</td>
<td>3</td>
<td>2214</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>81</td>
</tr>
</tbody>
</table>

- Per processor statistics for faults, interrupts, context switches, mutex and rwlock events, syscalls, user, system and idle times
Observing processes/threads

- *prstat*

<table>
<thead>
<tr>
<th>PID</th>
<th>USERNAME</th>
<th>SIZE</th>
<th>RSS</th>
<th>STATE</th>
<th>PRI</th>
<th>NICE</th>
<th>TIME</th>
<th>CPU</th>
<th>PROCESS/NLWP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1497</td>
<td>root</td>
<td>1118M</td>
<td>1098M</td>
<td>cpu3</td>
<td>10</td>
<td>0</td>
<td>147:01:03</td>
<td>25%</td>
<td>VirtualBox/14</td>
</tr>
<tr>
<td>11012</td>
<td>root</td>
<td>6828K</td>
<td>1972K</td>
<td>cpu2</td>
<td>0</td>
<td>0</td>
<td>0:12:56</td>
<td>7.0%</td>
<td>bash/1</td>
</tr>
<tr>
<td>18465</td>
<td>root</td>
<td>6828K</td>
<td>1972K</td>
<td>sleep</td>
<td>0</td>
<td>0</td>
<td>0:13:08</td>
<td>7.0%</td>
<td>bash/1</td>
</tr>
<tr>
<td>2901</td>
<td>root</td>
<td>6824K</td>
<td>1968K</td>
<td>cpu0</td>
<td>0</td>
<td>0</td>
<td>0:28:22</td>
<td>6.9%</td>
<td>bash/1</td>
</tr>
<tr>
<td>1587</td>
<td>root</td>
<td>1100M</td>
<td>1081M</td>
<td>sleep</td>
<td>59</td>
<td>0</td>
<td>27:18:05</td>
<td>5.0%</td>
<td>VirtualBox/14</td>
</tr>
<tr>
<td>11372</td>
<td>root</td>
<td>6948K</td>
<td>3492K</td>
<td>cpu1</td>
<td>59</td>
<td>0</td>
<td>0:00:00</td>
<td>0.0%</td>
<td>prstat/1</td>
</tr>
<tr>
<td>549</td>
<td>root</td>
<td>37M</td>
<td>19M</td>
<td>sleep</td>
<td>59</td>
<td>0</td>
<td>0:07:14</td>
<td>0.0%</td>
<td>Xorg/1</td>
</tr>
</tbody>
</table>

... 

Total: 149 processes, 352 lwps, load averages: 5.15, 5.18, 5.08
Observing processes/threads (2)

- `prstat -mL`

<table>
<thead>
<tr>
<th>PID</th>
<th>USERNAME</th>
<th>USR</th>
<th>SYS</th>
<th>TRP</th>
<th>TFL</th>
<th>DFL</th>
<th>LCK</th>
<th>SLP</th>
<th>LAT</th>
<th>VCX</th>
<th>ICX</th>
<th>SCL</th>
<th>SIG</th>
<th>PROCESS/LWPID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1497</td>
<td>root</td>
<td>1.2</td>
<td>98</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>218</td>
<td>269</td>
<td>14K</td>
<td>0</td>
<td>VirtualBox/6</td>
</tr>
<tr>
<td>18465</td>
<td>root</td>
<td>16</td>
<td>12</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>55</td>
<td>17</td>
<td>1K</td>
<td>1K</td>
<td>15K</td>
<td>130 bash/1</td>
</tr>
<tr>
<td>1587</td>
<td>root</td>
<td>6.2</td>
<td>21</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>72</td>
<td>1.0</td>
<td>457</td>
<td>98</td>
<td>42K</td>
<td>0 VirtualBox/6</td>
</tr>
</tbody>
</table>

- Microstates per thread
  - LAT – latency
    - How long the thread had to wait for CPU
    - Possible speedup estimate
Observing system memory

- `vmstat -p 1`

<table>
<thead>
<tr>
<th>memory</th>
<th>page</th>
<th>executable</th>
<th>anonymous</th>
<th>filesystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>swap</td>
<td>free</td>
<td>re</td>
<td>mf</td>
<td>fr</td>
</tr>
<tr>
<td>3538820</td>
<td>361752</td>
<td>130</td>
<td>1517</td>
<td>0</td>
</tr>
<tr>
<td>3498864</td>
<td>323956</td>
<td>11</td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td>3498760</td>
<td>323884</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3498760</td>
<td>323884</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **Scan rate (sr)**
  - Speed (in pages/second) the system is scanning memory
  - Can indicate **memory pressure**
Observing system memory (2)

- `vmstat -p 1`

<table>
<thead>
<tr>
<th>memory</th>
<th>page</th>
<th>executable</th>
<th>anonymous</th>
<th>filesystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>swap</td>
<td>free</td>
<td>re</td>
<td>mf</td>
<td>fr</td>
</tr>
<tr>
<td>3538820</td>
<td>361752</td>
<td>130</td>
<td>1517</td>
<td>0</td>
</tr>
<tr>
<td>3498864</td>
<td>323956</td>
<td>11</td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td>3498760</td>
<td>323884</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3498760</td>
<td>323884</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- Filesystem and executables paging: necessary
- Anonymous paging: memory shortage
Observing system memory (3)

- `vmstat -S 1`

```
kthr memory page disk faults cpu
r b w swap free si so pi po fr de sr s1 s2 -- -- in sy cs us sy id
0 0 7 2801904 38124 0 0 4 15099 15369 0 873355 283 0 0 0 15734 26722 45493
1 42 57
```

- `kthr:w` is the number of threads that were swapped-out
  - Indication of severe memory shortage
Observing system memory (4)

- **prstat -mL**

<table>
<thead>
<tr>
<th>PID</th>
<th>USERNAME</th>
<th>USR</th>
<th>SYS</th>
<th>TRP</th>
<th>TFL</th>
<th>DFL</th>
<th>LCK</th>
<th>SLP</th>
<th>LAT</th>
<th>VCX</th>
<th>ICX</th>
<th>SCL</th>
<th>SIG</th>
<th>PROCESS/LWPID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1497</td>
<td>root</td>
<td>1.2</td>
<td>98</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>218</td>
<td>269</td>
<td>14K</td>
<td>0</td>
<td>VirtualBox/6</td>
</tr>
<tr>
<td>18465</td>
<td>root</td>
<td>16</td>
<td>12</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>55</td>
<td>17</td>
<td>1K</td>
<td>1K</td>
<td>15K</td>
<td>130</td>
<td>bash/1</td>
</tr>
<tr>
<td>1587</td>
<td>root</td>
<td>6.2</td>
<td>21</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>72</td>
<td>1.0</td>
<td>457</td>
<td>98</td>
<td>42K</td>
<td>0</td>
<td>VirtualBox/6</td>
</tr>
</tbody>
</table>

- **TFL** – percentage of time the thread has spent processing instruction page faults
- **DFL** - percentage of time the thread has spent processing data page faults
Scan rate function

Increase of scan rate on a 8GiB machine

Pages scanned per second vs. Number of free pages

Scan rate

Number of free pages

Pages scanned per second

10000000
1000000
100000
10000
1000
100
10
1
0
32733
1047459
100
Scan rate function (2)

- \text{freemem} > \text{lotsfree} + \text{deficit}
  - No scanning activity
- \text{freemem} < \text{lotsfree} + \text{deficit}
  - Scanning at slowscan rate
- \text{freemem} = \text{lotsfree}
  - Still scanning at slowscan rate
- \text{freemem} = 0
  - Scanning at fastscan rate
Tracing syscalls

- `truss echo 'Hello world!'`

```
execve("/usr/bin/echo", 0x08047CF4, 0x08047D00) argc = 2
mmap(0x00000000, 4096, PROT_READ|PROT_WRITE|PROT_EXEC, MAP_PRIVATE|MAP_ANON, -1, 0) = 0xFEFB0000
resolvepath("/usr/lib/ld.so.1", "/lib/ld.so.1", 1023) = 12
resolvepath("/usr/bin/echo", "/usr/bin/echo", 1023) = 13
sysconfig(_CONFIG_PAGESIZE) = 4096
xstat(2, "/usr/bin/echo", 0x080479B8) = 0
open("/var/ld/ld.config", O_RDONLY) Err#2 ENOENT
...
fstat64(1, 0x08047E0) = 0
write(1, "Hello world!"..., 13) = 13
_exit(0)
```

Crash Dump Analysis – MFF UK – Command line tools
Displaying thread stack

• pstack 1587/10

1587: /opt/VirtualBox/amd64/VirtualBox --comment centos --startvm cc4605e0-a
----------------- lwp# 10 / thread# 10 -----------------
fffffd7fff0a234a sigtimedwait (fffffd7ffbcc2eb0, fffffd7ffbcc2c40, 0)
fffffd7fff08b84 sigwaitinfo () + c
fffffd7ffedf95fe _Z13rttimerThreadP11RTTHREADINTPv () + 38e
fffffd7ffedd3b3c rtThreadMain () + 2c
fffffd7ffedf8d2b _Z18rtThreadNativeMainPv () + 7b
fffffd7fff099de5 _thrp_setup () + 8d
fffffd7fff09a0a0 _lwp_start ()
Displaying address space

• pmap 8394

8394:  less /etc/passwd

08045000  12K rw---  [ stack ]
08050000  108K r-x--  /usr/bin/less
0807A000  24K rwx--  /usr/bin/less
08080000  32K rwx--  [ heap ]
FEA00000  2416K r-x--  /usr/lib/locale/en_US.UTF-8/en_US.UTF-8.so.3
FEC6B000  4K rwx--  /usr/lib/locale/en_US.UTF-8/en_US.UTF-8.so.3
FED9E000  4K rwxs-  [ anon ]
...
FEFFD000  4K rwx--  /lib/ld.so.1

total      4420K