# 一线运维故障排查经验分享

李成栋(brytonlee01@gmail.com)

2014-05-09

# 进程和线程

## 1.查看进程和线程(ps, top,/proc/)

```
top - 09:27:53 up 552 days, 21:08,  1 user,  load average: 1.99, 2.19, 1.79
Tasks: 1333 total,   4 running, 1329 sleeping,   0 stopped,   0 zombie
Cpu(s):  7.9%us,  3.8%sy,  0.0%ni, 87.4%id,  0.2%wa,  0.0%hi,  0.4%si,  0.3%st
Mem:   5242880k total,  5179544k used,    63336k free,   146248k buffers
Swap:  2097144k total,     3484k used,  2093660k free,   784032k cached

  PID USER      PR  NI  VIRT  RES  SHR S %CPU %MEM    TIME+  COMMAND
29102 admin     16   0 1672m 1.3g  17m S  2.0 25.1   1:43.36 java
29516 admin     16   0 1659m 1.3g  17m S  1.6 25.5   0:40.34 java
28397 admin     15   0 1672m 1.3g  17m S  1.3 25.1   9:13.57 java
28187 admin     15   0 1656m 1.3g  17m S  1.3 25.2   9:48.82 java
28401 admin     15   0 1656m 1.3g  17m S  1.3 25.2   9:11.71 java
28396 admin     15   0 1659m 1.3g  17m S  1.3 25.5   8:45.77 java
29559 admin     15   0 1659m 1.3g  17m S  1.3 25.5   9:26.07 java
23344 chengdon  15   0 13672 2052  816 R  1.0  0.0   0:00.63 top
28985 admin     16   0 1672m 1.3g  17m S  1.0 25.1   9:37.38 java
```

# 进程和线程(cont.1)

## 2.理解进程的状态(uptime, vmstat)

D Uninterruptible sleep (usually IO)
R Running or runnable (on run queue)
S Interruptible sleep (waiting for an event to complete)
T Stopped, either by a job control signal or because it is being traced.
W paging (not valid since the 2.6.xx kernel)
X dead (should never be seen)
Z Defunct ("zombie") process, terminated but not reaped by its parent.

< high-priority (not nice to other users)
N low-priority (nice to other users)
L has pages locked into memory (for real-time and custom IO)
s is a session leader
l is multi-threaded (using CLONE_THREAD, like NPTL pthreads do)
+ is in the foreground process group

# 进程和线程(cont.2)

## 3.进程和线程的区别

进程是资源分配单位。

线程(kernel-thread)是调度单位。

## 4.理解load和CPU利用率

load = D + R. load vs load average (vmstat)

CPU利用率的多核性(多硬件执行线程)

# 进程和线程(cont.3)

5.进程里面有什么？(pmap, ldd)

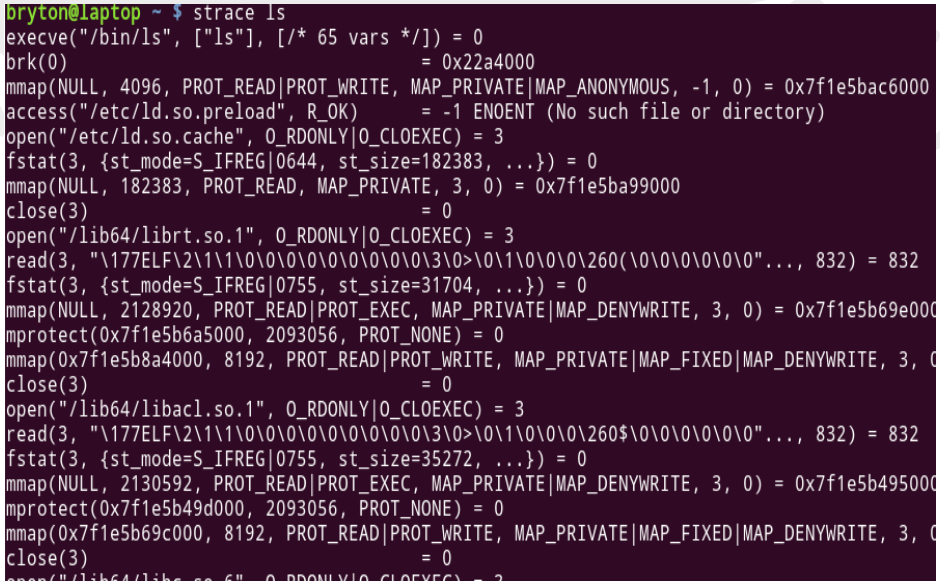
6.理解堆和栈(gdb bt,perf)。

```
[chengdong.licd@membercenter-service19 ~]$ pmap $$
23307:   -bash
0000000000400000    712K r-x--  /bin/bash
00000000006b2000     40K rw---  /bin/bash
00000000006bc000     20K rw---    [ anon ]
00000000008bb000     32K rw---  /bin/bash
0000000005024000    264K rw---    [ anon ]
0000003bc7a00000    112K r-x--  /lib64/ld-2.5.so
0000003bc7c1b000      4K r----  /lib64/ld-2.5.so
0000003bc7c1c000      4K rw---  /lib64/ld-2.5.so
0000003bc7e00000   1328K r-x--  /lib64/libc-2.5.so
0000003bc7f4c000   2048K -----  /lib64/libc-2.5.so
0000003bc814c000     16K r----  /lib64/libc-2.5.so
0000003bc8150000      4K rw---  /lib64/libc-2.5.so
0000003bc8151000     20K rw---    [ anon ]
0000003bc8200000      8K r-x--  /lib64/libdl-2.5.so
0000003bc8202000   2048K -----  /lib64/libdl-2.5.so
0000003bc8402000      4K r----  /lib64/libdl-2.5.so
0000003bc8403000      4K rw---  /lib64/libdl-2.5.so
0000003bc9600000     12K r-x--  /lib64/libtermcap.so.2.0.8
0000003bc9603000   2044K -----  /lib64/libtermcap.so.2.0.8
0000003bc9802000      4K rw---  /lib64/libtermcap.so.2.0.8
00002abd79330000      4K rw---    [ anon ]
00002abd79343000     12K rw---    [ anon ]
00002abd79346000     40K r-x--  /lib64/libnss_files-2.5.so
00002abd79350000   2044K -----  /lib64/libnss_files-2.5.so
00002abd7954f000      4K r----  /lib64/libnss_files-2.5.so
00002abd79550000      4K rw---  /lib64/libnss_files-2.5.so
00002abd79551000  55120K r----  /usr/lib/locale/locale-archive
00002abd7cb25000     28K r--s-  /usr/lib64/gconv/gconv-modules.cache
00002abd7cb2c000      8K rw---    [ anon ]
00007fff31765000     84K rw---    [ stack ]
ffffffffff600000   8192K -----    [ anon ]
 total             74268K
```

```
Breakpoint 1, ngx_http_stylecombine_body_filter (r=0x1222900, in=0x1223bf8)
    at /home/bryton/code/styleCombine3/src/nginx//ngx_http_stylecombine_filter_module.c:402
402     {
(gdb) bt
#0  ngx_http_stylecombine_body_filter (r=0x1222900, in=0x1223bf8)
    at /home/bryton/code/styleCombine3/src/nginx//ngx_http_stylecombine_filter_module.c:402
#1  0x0000000000415ea7 in ngx_output_chain (ctx=ctx@entry=0x1223b20, in=in@entry=0x7fffe4c34
#2  0x0000000000448f30 in ngx_http_copy_filter (r=0x1222900, in=0x7fffe4c34210) at src/http/
#3  0x000000000045a429 in ngx_http_range_body_filter (r=0x1222900, in=<optimized out>)
    at src/http/modules/ngx_http_range_filter_module.c:559
#4  0x000000000043f3ee in ngx_http_output_filter (r=r@entry=0x1222900, in=in@entry=0x7fffe4c
    at src/http/ngx_http_core_module.c:1967
#5  0x000000000045957d in ngx_http_static_handler (r=0x1222900) at src/http/modules/ngx_http
#6  0x000000000043f882 in ngx_http_core_content_phase (r=0x1222900, ph=0x1237830) at src/htt
#7  0x000000000043a653 in ngx_http_core_run_phases (r=r@entry=0x1222900) at src/http/ngx_htt
#8  0x000000000043a755 in ngx_http_handler (r=r@entry=0x1222900) at src/http/ngx_http_core_m
#9  0x0000000000445290 in ngx_http_process_request (r=r@entry=0x1222900) at src/http/ngx_htt
#10 0x000000000044586d in ngx_http_process_request_headers (rev=rev@entry=0x123c7b0) at src/
#11 0x0000000000445d8b in ngx_http_process_request_line (rev=0x123c7b0) at src/http/ngx_http
#12 0x0000000000443575 in ngx_http_init_request (rev=0x123c7b0) at src/http/ngx_http_request
#13 0x000000000430c3d in ngx_epoll_process_events (cycle=<optimized out>, timer=<optimized
    at src/event/modules/ngx_epoll_module.c:683
#14 0x0000000000428301 in ngx_process_events_and_timers (cycle=cycle@entry=0x12138f0) at src
#15 0x000000000042f6a4 in ngx_worker_process_cycle (cycle=cycle@entry=0x12138f0, data=data@e
    at src/os/unix/ngx_process_cycle.c:853
#16 0x000000000042cd0c in ngx_spawn_process (cycle=cycle@entry=0x12138f0, proc=proc@entry=0x
    data=data@entry=0x0, name=name@entry=0x4944af "worker process", respawn=respawn@entry=-3
#17 0x000000000042eaec in ngx_start_worker_processes (cycle=cycle@entry=0x12138f0, n=1, type
    at src/os/unix/ngx_process_cycle.c:392
#18 0x000000000042fd9f in ngx_master_process_cycle (cycle=cycle@entry=0x12138f0) at src/os/u
#19 0x0000000000041242f in main (argc=<optimized out>, argv=<optimized out>) at src/core/ngin
(gdb)
```

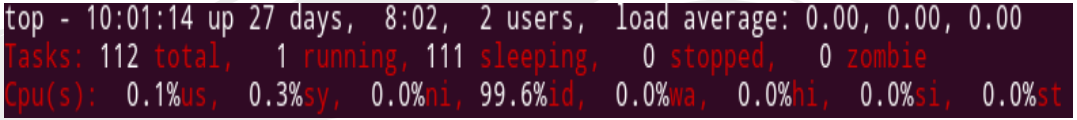
# 进程和线程(cont.4)

## 7.理解进程的内核态和用户态(strace)



# 进程和线程(cont.5)

8.理解进程时间。




- us: time running un-niced user processes.
- sy: time running kernel process.
- ni: time running niced user process.
- wa: time waiting for I/O completion.
- hi: time spent servicing hardware interrupts.
- si: time spent servicing software interrupts.
- st: time stolen from this vm by the hypervisor (better explanation: involuntary wait)

# 进程和线程(cont.6)

## 9.理解进程I/O (lsof)。

```
nginx   24970 admin    0u   CHR                1,3                 1057 /dev/null
nginx   24970 admin    1u   CHR                1,3                 1057 /dev/null
nginx   24970 admin    2w  FIFO                0,6          109015102 pipe
nginx   24970 admin    3u  unix 0xffff8801cfb9e140          109015107 socket
nginx   24970 admin    4u  unix 0xffff8801cfb9f1c0          109015106 socket
nginx   24970 admin    5w  FIFO                0,6          109015103 pipe
nginx   24970 admin    6r  0000               0,11        0 109015109 eventpoll
nginx   24970 admin    7u  IPv4          109015091                TCP *:http (LISTEN)
nginx   24970 admin    8u  unix 0xffff8801cfb9e400          109015110 socket
nginx   24970 admin    9u  unix 0xffff88015df4cc80          109015094 socket
nginx   24970 admin   10u  unix 0xffff88015df4c180          109015095 socket
nginx   24970 admin   11u  unix 0xffff88015df4cf40          109015096 socket
nginx   24970 admin   12u  unix 0xffff88015df4dd00          109015097 socket
nginx   24970 admin   13u  unix 0xffff88015df4d780          109015098 socket
nginx   24970 admin   14u  unix 0xffff88015df4c440          109015099 socket
nginx   24970 admin   15u  unix 0xffff8801cfb9f480          109015100 socket
nginx   24970 admin   16w  FIFO                0,6          109015102 pipe
nginx   24970 admin   17w  FIFO                0,6          109015101 pipe
nginx   24970 admin   18u  IPv4          109015116                TCP 10.147.96.105:60592->10.97.126.70:etlservicemgr (ESTABLISHED)
nginx   24970 admin   19u  unix 0xffff8801cfb9fa00          109015113 socket
nginx   24970 admin   20u  unix 0xffff8801d4a32340          109015120 socket
nginx   24970 admin   21u  unix 0xffff8801d4a32600          109015123 socket
nginx   24970 admin   22u  unix 0xffff88015df4d200          109015093 socket
nginx   24970 admin   23r   REG                0,3        0 4026531842 /proc/meminfo
```

# 进程和线程实战(gdb)

1.调试进程(attach),查看线程数(info threads)。
2.查看线程执行栈(backtrace)。
3.查看堆栈变量(print)和当前执行代码(list)。

# 进程和线程实战(jvm)

1.查看jvm里面线程(jstack)

2.查看jvm里面某个线程当前栈(jstack)

# 内存

1.系统内存状况。(free)

2.某个进程内存使用状况。(top)

3.理解虚拟内存,交换空间。

```
[chengdong.licd@cm10-vm-346 admin]$ free -m
             total       used       free     shared    buffers     cached
Mem:          7500       6429       1070          0        527       3880
-/+ buffers/cache:       2021       5478
Swap:         1961         41       1919
```

# 内存实战

1.应用程序泄漏。

a.一般进程查看堆增长(pmap,top)。

b.java进程(jstat -gcutil, jmap)。

2.内核内存泄漏(cat /proc/meminfo)

```
MemTotal:       7680000 kB
MemFree:        1096432 kB
Buffers:         540444 kB
Cached:         3973212 kB
SwapCached:           0 kB
Active:         2820160 kB
Inactive:       2990296 kB
HighTotal:            0 kB
HighFree:             0 kB
LowTotal:       7680000 kB
LowFree:        1096432 kB
SwapTotal:      2008116 kB
SwapFree:       1965524 kB
Dirty:              432 kB
Writeback:            0 kB
AnonPages:      1296792 kB
Mapped:           47928 kB
Slab:            579924 kB
PageTables:        8988 kB
NFS_Unstable:         0 kB
Bounce:               0 kB
CommitLimit:    5848116 kB
Committed_AS:   3891032 kB
VmallocTotal: 34359738367 kB
VmallocUsed:       1224 kB
VmallocChunk: 34359737143 kB
```

# I/O - 网络

1.网络I/O的特点。

a.远程, 双方遵循相同的通信协议(TCP/IP)。

b.主动和被动关系。

c.可靠数据传输协议基于连接。

d.网络超时和空闲。

# TCP链接

1.链接状态建立(netstat/ss)。

| | Time | Source | Destination | Protocol | Length | Info |
|---|---|---|---|---|---|---|
| 1 | 0.000000 | 172.22.2.85 | 10.147.148.3 | TCP | 74 | 34298 > http [SYN] Seq=0 Win=5840 Len=0 MSS=1460 SACK_PERM=1 TSval=3848636735 TSecr=0 WS=128 |
| 2 | 0.001042 | 10.147.148.3 | 172.22.2.85 | TCP | 74 | http > 34298 [SYN, ACK] Seq=0 Ack=1 Win=14480 Len=0 MSS=1460 SACK_PERM=1 TSval=1093321886 TSecr=3848636 |
| 3 | 0.001060 | 172.22.2.85 | 10.147.148.3 | TCP | 66 | 34298 > http [ACK] Seq=1 Ack=1 Win=5888 Len=0 TSval=3848636736 TSecr=1093321886 |
| 4 | 0.001118 | 172.22.2.85 | 10.147.148.3 | HTTP | 388 | GET /tigo/traceparam/httpreq?fields=stat_date,trace_param_value&filter=stat_date:[2013-12-20%20TO%20201 |
| 5 | 0.002190 | 10.147.148.3 | 172.22.2.85 | TCP | 66 | http > 34298 [ACK] Seq=1 Ack=323 Win=15616 Len=0 TSval=1093321887 TSecr=3848636736 |

2.链接状态可能出现的问题。

a. syn-flood攻击。

b.应用程序不响应。

```
bryton@laptop ~/code $ netstat -lntp
(Not all processes could be identified, non-owned process info
 will not be shown, you would have to be root to see it all.)
Active Internet connections (only servers)
Proto Recv-Q Send-Q Local Address           Foreign Address         State       PID/Program name
tcp        2      0 0.0.0.0:8000            0.0.0.0:*               LISTEN      7838/./server
```

# TCP链接(cont.1)

| Time | ---cpu-- util | ---mem-- util | ---tcp- retran |
|---|---|---|---|
| 09/05/14-09:45 | 57.70 | 29.54 | 0.73 |
| 09/05/14-09:50 | 59.11 | 29.66 | 0.70 |
| 09/05/14-09:55 | 59.37 | 29.64 | 0.74 |
| 09/05/14-10:00 | 60.15 | 29.73 | 0.77 |
| 09/05/14-10:05 | 60.93 | 29.71 | 0.74 |
| 09/05/14-10:10 | 61.48 | 29.81 | 0.72 |
| 09/05/14-10:15 | 61.57 | 29.76 | 0.79 |
| 09/05/14-10:20 | 62.45 | 29.86 | 0.78 |
| 09/05/14-10:25 | 62.64 | 29.81 | 0.76 |
| 09/05/14-10:30 | 63.27 | 29.93 | 0.76 |
| 09/05/14-10:35 | 63.32 | 29.89 | 0.80 |
| 09/05/14-10:40 | 63.37 | 29.98 | 0.77 |
| 09/05/14-10:45 | 63.65 | 29.93 | 0.75 |
| 09/05/14-10:50 | 65.04 | 30.05 | 0.76 |
| 09/05/14-10:55 | 64.35 | 30.00 | 0.77 |
| 09/05/14-11:00 | 64.30 | 30.10 | 0.80 |
| 09/05/14-11:05 | 65.14 | 30.05 | 0.78 |
| 09/05/14-11:10 | 64.92 | 30.14 | 0.76 |
| | | 08 | 0.80 |

3.链接建立之后可能的问题。

a.超时(tsar)。

b.链接保持(TCP KeepAlive)。

| | Time | Source | Destination | Protocol | Length | Info |
|---|---|---|---|---|---|---|
| 1 | 0.000000 | 172.22.2.85 | 10.147.208.49 | TCP | 74 | 58936 > http [SYN] Seq=0 Win=5840 Len=0 MSS=1460 SACK_PERM=1 TSval=38492 |
| 2 | 0.000979 | 10.147.208.49 | 172.22.2.85 | TCP | 74 | http > 58936 [SYN, ACK] Seq=0 Ack=1 Win=5840 Len=0 MSS=1452 SACK_PERM=1 |
| 3 | 0.001013 | 172.22.2.85 | 10.147.208.49 | TCP | 54 | 58936 > http [ACK] Seq=1 Ack=1 Win=5840 Len=0 |
| 4 | 0.001065 | 172.22.2.85 | 10.147.208.49 | HTTP | 376 | GET /tigo/traceparam/httpreq?fields=stat_date,trace_param_value&filter=s |
| 5 | 0.002436 | 10.147.208.49 | 172.22.2.85 | TCP | 60 | http > 58936 [ACK] Seq=1 Ack=323 Win=15544 Len=0 |
| 6 | 90.001903 | 10.147.208.49 | 172.22.2.85 | TCP | 60 | http > 58936 [RST] Seq=1 Win=0 Len=0 |

c.速度上不去(尤其是同机房)。

net.core.rmem_default = 212992  net.core.rmem_max = 212992

net.core.wmem_default = 212992  net.core.wmem_max = 212992

net.ipv4.tcp_rmem = 4096  87380  6291456  net.ipv4.tcp_wmem = 4096 16384 4194304

# 网络问题实战(tsar, tcpdump)

1.超时问题(tsar)。

2.嗅探之王(tcpdump)。

tcpdump -D

tcpdump -i eth0 -n tcp and port 80 and host x.x.x.x

tcpdump -i eth0 -n -XX tcp and port 80 and host x.x.x.x

tcpdump -i eth0 -s 0 -w file.cap tcp and port 80 and host x.x.x.x

# Thanks!

# QA