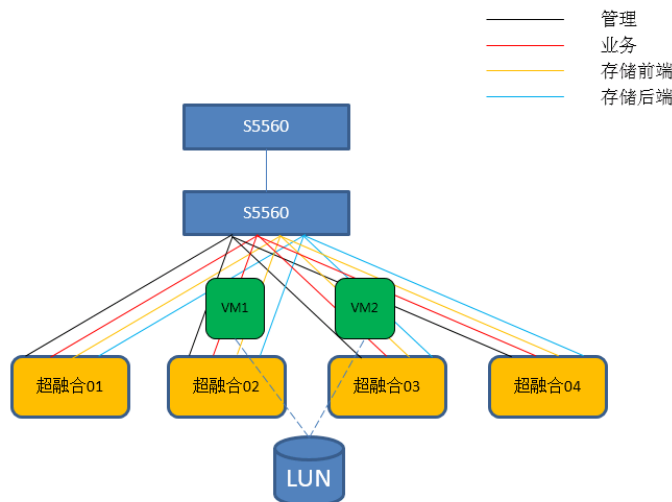


知 存储异常触发linux系统bug导致虚拟机内部操作系统部分卡住，鼠标点击无反应

厉梦如 2018-09-11 发表

组网及说明

4台超融合服务器，4个网口分别规划为管理网、业务网、存储前端网、存储后端网。4个网卡连接到同一台交换机S5560。2台Centos 7.3 64bit虚拟机内部直接通过ISCSI连接ONESTor的1T的lun。



问题描述

客户反馈有3台虚拟机访问控制台卡住，通过鼠标点击无反应。尝试重启其中一台虚拟机，虚拟机就恢复正常。

过程分析

1. 确认问题现象：只有linux虚拟机存在卡住的情况，windows虚拟机控制台访问和操作正常。且linux虚拟机一部分业务运行正常。

2. 登陆CVM管理页面查看CAS平台主机状态和虚拟机状态正常

3. Ssh到异常虚拟机所在CVK主机，查看syslog日志，当前时间点没有异常记录

4. 尝试ssh到异常虚拟机分析操作系统内部日志：

通过ssh到虚拟机内部操作系统查看message日志发现都有部分cpu卡住等记录，同时查看到在该时间点都有虚拟机/CVK访问存储超时的情况（有虚拟机直接内部操作系统挂载ONESTor存储）。

详细查看三台虚拟机记录见下：

1) 虚拟机HJDJK3-GATE：

查看虚拟机内部操作系统message日志，读到部分卡住时间在8月6日11点53左右

Aug 6 11:53:44 HJDJK3-GATE kernel: NMI watchdog: BUG: soft lockup - CPU#4 stuck for 22s! [java:132156]

对应该时间点，虚拟机所在CVK主机上有记录此时访问存储超时的记录（该虚拟机的磁盘文件就存储在dm-2上的）

Aug 6 03:51:31 cvknode1 kernel: [3340102.097774] hpilo 0000:01:00.2: Open could not dequeue a packet

Aug 6 11:55:11 cvknode1 kernel: [3369150.221490] (o2hb-DADF02EF65,578588,18):o2hb_thread:1139 o2hb_do_disk_heartbeat used 38318 msecs on device(dm-2), ret = 0.

Aug 6 11:55:20 cvknode1 kernel: [3369159.062911] (o2hb-DADF02EF65,578588,18):o2hb_thread:1139 o2hb_do_disk_heartbeat used 4823 msecs on device(dm-2), ret = 0.

2) 虚拟机HJDJK3-DB2：

查看虚拟机内部操作系统message日志，读到在8月17日8点12分出现系统卡住的情况，在系统卡住之前可以看到日志打印存储访问iscsi连接中断（该虚拟机单独通过iscsi挂载了存储卷）。

Aug 17 08:12:09 HJDJK3-DB2 iscsid: Kernel reported iSCSI connection 1:0 error (1022 - Invalid or unknown error code) state (3)

Aug 17 08:12:27 HJDJK3-DB2 kernel: NMI watchdog: BUG: soft lockup - CPU#9 stuck for 23s! [hasvr d:2764]

Aug 17 08:12:27 HJDJK3-DB2 kernel: Modules linked in: fuse xt_CHECKSUM iptable_mangle ipt_MASQUERADE nf_nat_masquerade_ipv4 iptable_nat nf_nat_ipv4 nf_nat nf_conntrack_ipv4

```
nf_defrag_ipv4 xt_conntrack nf_conntrack ipt_REJECT nf_reject_ipv4 tun bridge stp llc ebtable_filter
ebtables ip6table_filter ip6_tables iptable_filter sd_mod crc_t10dif crct10dif_generic crct10dif_commo
n iscsi_tcp libiscsi_tcp libiscsi scsi_transport_iscsi intel_powerclamp ppdev parport_pc i2c_piix4 parp
ort bttrfs sg virtio_balloon zlib_deflate pcspkr raid6_pq xor vfat fat ext4 mbcache jbd2 replicator(OE) nf
sd_auth_rpcgss nfs_acl lockd grace sunrpc binfmt_misc ip_tables xfs libcrc32c sr_mod cdrom ata_ge
neric pata_acpi virtio_net virtio_console virtio_scsi virtio_blk cirrus drm_kms_helper syscopyarea sysf
illrect sysimgblt fb_sys_fops ttm drm ata_piix
```

3) 虚拟机HJDJK3-QX

查看虚拟机内部操作系统message日志，读到在8月6日11点55分出现系统卡住的情况。

```
Aug 6 11:55:23 HJDJK3-QX kernel: NMI watchdog: BUG: soft lockup - CPU#4 stuck for 23s! [java:51080]
```

```
Aug 6 11:55:23 HJDJK3-QX kernel: Modules linked in: tcp_lp arc4 md4 nls_utf8 cifs dns_resolver fus
e xt_CHECKSUM iptable_mangle ipt_MASQUERADE nf_nat_masquerade_ipv4 iptable_nat
nf_nat_ipv4 nf_nat nf_conntrack_ipv4 nf_defrag_ipv4 xt_conntrack nf_conntrack ipt_REJECT nf_reje
ct_ipv4 tun bridge stp llc ebtable_filter ebtables ip6table_filter ip6_tables iptable_filter ppdev sg
parport_pc parport virtio_balloon i2c_piix4 pcspkr intel_powerclamp nfsd_auth_rpcgss nfs_acl lockd g
race sunrpc binfmt_misc ip_tables xfs libcrc32c sr_mod cdrom virtio_scsi virtio_console virtio_blk virti
o_net ata_generic pata_acpi cirrus drm_kms_helper syscopyarea sysfillrect sysimgblt fb_sys_fops tt
m drm ata_piix serio_raw libata virtio_pci virtio_ring i2c_core virtio floppy dm_mirror dm_region_hash
dm_log dm_mod
```

```
Aug 6 11:55:23 HJDJK3-QX kernel: CPU: 4 PID: 51080 Comm: java Tainted: G          L -----
3.10.0-514.el7.x86_64 #1
```

```
Aug 6 11:55:23 HJDJK3-QX kernel: Hardware name: QEMU Standard PC (i440FX + PIIX, 1996), BI
OS rel-1.7.5-0-ge51488c-20140602_164612-nilsson.home.kraxel.org 04/01/2014
```

```
Aug 6 11:55:23 HJDJK3-QX kernel: task: ffff8800b3273ec0 ti: ffff8801661cc000 task.ti: ffff8801661cc
000
```

```
Aug 6 11:55:23 HJDJK3-QX kernel: RIP: 0010:[<ffffffff810f98e2>] [<ffffffff810f98e2>]
smp_call_function_many+0x202/0x260
```

同样，对应该时间点，虚拟机所在CVK主机cvknode1上有记录此时访问存储超时的记录（该虚拟机的磁盘文件就存储在dm-2上的）

```
Aug 6 03:51:31 cvknode1 kernel: [3340102.097774] hpilo 0000:01:00.2: Open could not dequeue a p
acket
```

```
Aug 6 11:55:11 cvknode1 kernel: [3369150.221490] (o2hb-
DADF02EF65,578588,18):o2hb_thread:1139 o2hb_do_disk_heartbeat used 38318 msecs on device(d
m-2), ret = 0.
```

```
Aug 6 11:55:20 cvknode1 kernel: [3369159.062911] (o2hb-
DADF02EF65,578588,18):o2hb_thread:1139 o2hb_do_disk_heartbeat used 4823 msecs on device(d
m-2), ret = 0.
```

5.分析soft lockup相关资料说明

虚拟机卡住都是在报soft lockup，且虚拟机都是centos7.3的版本。对应内核版本：Linux version 3.10.0-514.el7.x86_64 (mockbuild@x86-039.build.eng.bos.redhat.com) (gcc version 4.8.5 20150623 (Red Hat 4.8.5-11) (GCC)) #1 SMP Wed Oct 19 11:24:13 EDT 2016

查找相关资料确认该内核版本有一个已知问题，存储不稳定或访问高延迟情况下触发linux虚拟机内核smp模块lock BUG（Linux version 3.10.0 内核是2013年6月30发布的，已经停止更新了），导致虚拟机系统部分异常。linux kernel对于smp模块lock使用已经有优化，现场虚拟机是发布版本，内核应该是没有更新过。

所以问题就定位在8月6日11:55和8月7日11:12出现存储访问异常触发虚拟机内部操作系统centos7.3的已知内核bug，导致操作系统部分异常。

解决方法

问题定位在存储异常导致触发虚拟机操作系统bug，所以可以从以下三方面进行处理：

- 1、可以将虚拟机重新启动暂时规避。
- 2、请客户更新虚拟机内部操作系统内核，解决操作系统已知内核bug。
- 3、存储不稳定或高延迟情况：请排查存储设备在8月6日11:55和8月7日11:12时间点异常。