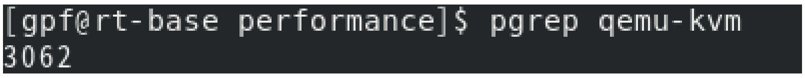
虚拟机退出次数测试

测试虚拟机退出(VMexit)的步骤

1. 首先需要找到qemu-kvm的进程

   ```
   pgrep qemu-kvm
   ```

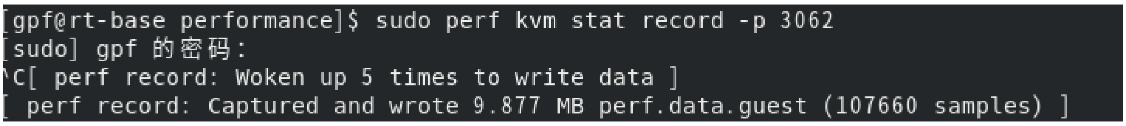
   ```
   [gpf@rt-base performance]$ pgrep qemu-kvm
   3062
   ```

   可知qemu-kvm的进程id为3062.

2. 使用perf命令开始记录虚拟机退出事件:

   ```
   sudo perf kvm stat record -p 3062
   ```

   该命令会开始记录虚拟机退出时间,同时不断地写入一个文件perf.data.guest,直到用ctrl+C发送信号停止记录,最后会生成一个文件,用来记录内容:

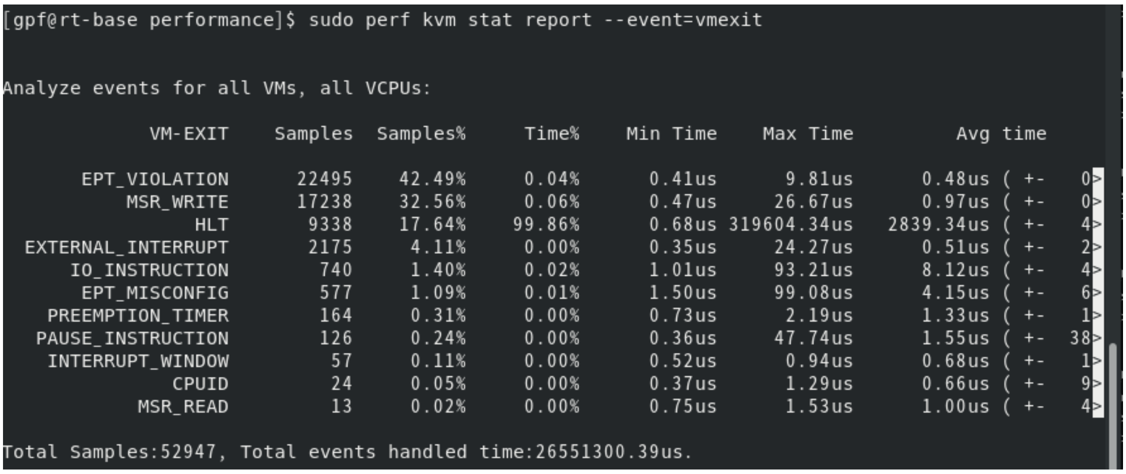
   

   ```
   [gpf@rt-base performance]$ sudo perf kvm stat record -p 3062
   [sudo] gpf 的密码:
   ^C[ perf record: Woken up 5 times to write data ]
   [ perf record: Captured and wrote 9.877 MB perf.data.guest (107660 samples) ]
   ```

3. 使用perf report命令读取上面生成的文件然后生成汇总信息,命令如下:

   ```
   sudo perf kvm stat report --event=vmexit
   ```

   生成的结果如下图:

   ```
   [gpf@rt-base performance]$ sudo perf kvm stat report --event=vmexit


   Analyze events for all VMs, all VCPUs:

                VM-EXIT    Samples  Samples%     Time%    Min Time     Max Time         Avg time

           EPT_VIOLATION      22495    42.49%     0.04%      0.41us       9.81us      0.48us ( +-    0>
               MSR_WRITE      17238    32.56%     0.06%      0.47us      26.67us      0.97us ( +-    0>
                     HLT       9338    17.64%    99.86%      0.68us 319604.34us   2839.34us ( +-    4>
      EXTERNAL_INTERRUPT       2175     4.11%     0.00%      0.35us      24.27us      0.51us ( +-    2>
          IO_INSTRUCTION        740     1.40%     0.02%      1.01us      93.21us      8.12us ( +-    4>
           EPT_MISCONFIG        577     1.09%     0.01%      1.50us      99.08us      4.15us ( +-    6>
        PREEMPTION_TIMER        164     0.31%     0.00%      0.73us       2.19us      1.33us ( +-    1>
       PAUSE_INSTRUCTION        126     0.24%     0.00%      0.36us      47.74us      1.55us ( +-   38>
        INTERRUPT_WINDOW         57     0.11%     0.00%      0.52us       0.94us      0.68us ( +-    1>
                   CPUID         24     0.05%     0.00%      0.37us       1.29us      0.66us ( +-    9>
                MSR_READ         13     0.02%     0.00%      0.75us       1.53us      1.00us ( +-    4>

   Total Samples:52947, Total events handled time:26551300.39us.
   ```

VMExit统计字段

字段	意义
EPT_VIOLATION	EPT表缺页
MSR_WRITE	写入MSR寄存器
HLT	停机时间
EXTERNAL_INTERRUPT	外部中断
IO_INSTRUCTION	IO指令
EPT_MISCONFIG	EPT表的重新配置?
PREEMPTION_TIMER	抢占定时器
PAUSE_INSTRUCTION	暂停指令
INTERRUPT_WINDOW	中断窗口
CPUID	检测CPU
MSR_READ	读取MSR寄存器

1. EPT是为了提升虚拟化内存映射的效率而提供的一项技术，打开EPT后，GuestOS运行时，通过页表转化处理的地址不再是真实的物理地址，而是被称作为guest-physical addressed,经过EPT的转化后才成为真实的物理地址。
2. MSR(Model Specific Register)指的是在x86架构处理器中，一系列用于控制CPU运行、功能开关、调试、跟踪程序运行、检测CPU性能方面的寄存器。
3. Preemption Timer是一种可以周期性使VM触发VMEXIT的一种机制。即设置了Preemption Timer之后，可以使得虚拟机在指定的TSC cycle之后产生一次VMEXIT并设置对应的exit_reason，trap到VMM中。

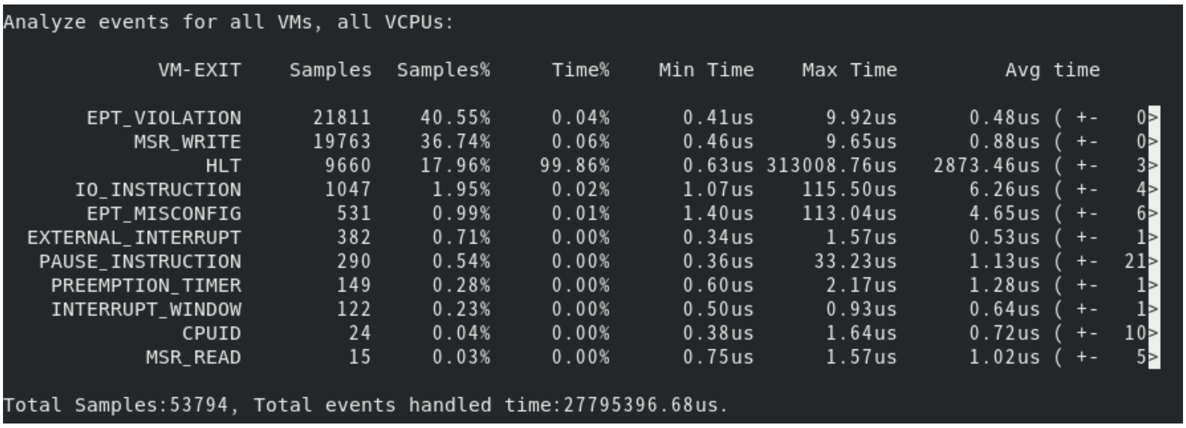
测试结果

该测试使用了两种组合方式，分别是GP-RT和GP-GP。且每种组合分别测试了是否绑定CPU的情况。一共四种，GuestOS执行的cyclictest指令如下：

```
sudo cyclictest -t1 -p 99 -i 10000 -l 1000
```

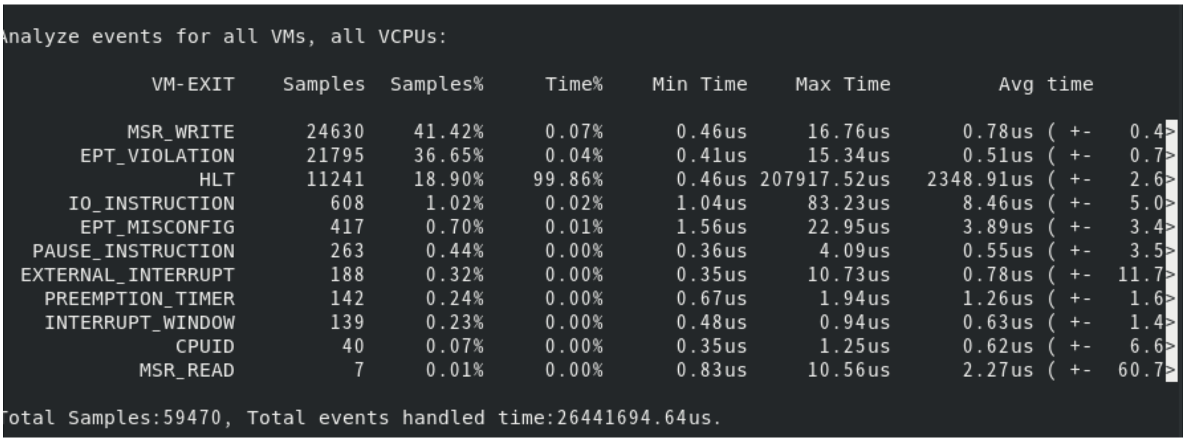
具体结果如下：

GP-GP-RAW:



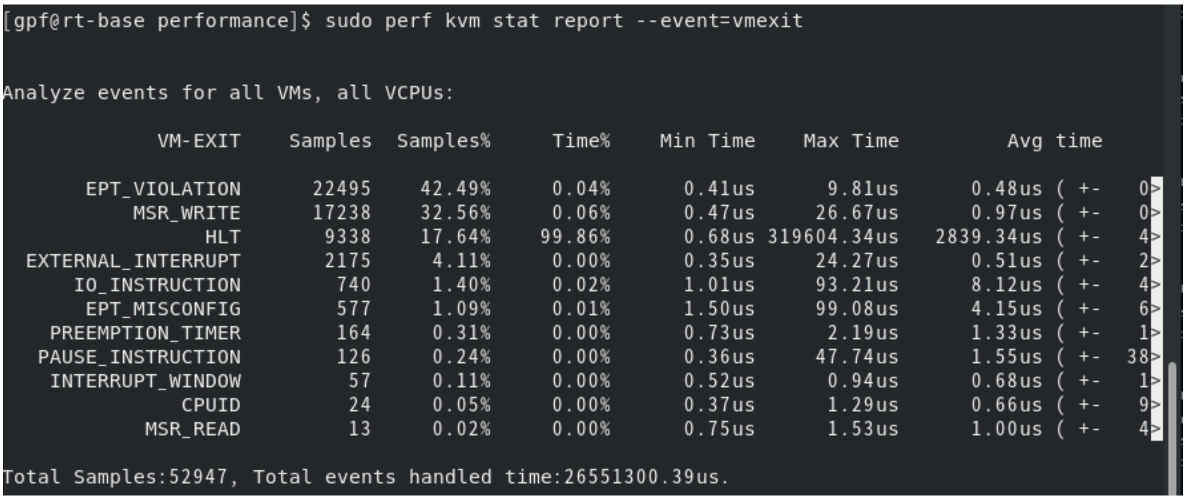
```
Analyze events for all VMs, all VCPUs:

             VM-EXIT    Samples  Samples%     Time%    Min Time     Max Time         Avg time

       EPT_VIOLATION      21811    40.55%     0.04%      0.41us       9.92us      0.48us ( +-   0>
           MSR_WRITE      19763    36.74%     0.06%      0.46us       9.65us      0.88us ( +-   0>
                 HLT       9660    17.96%    99.86%      0.63us  313008.76us   2873.46us ( +-   3>
      IO_INSTRUCTION       1047     1.95%     0.02%      1.07us     115.50us      6.26us ( +-   4>
       EPT_MISCONFIG        531     0.99%     0.01%      1.40us     113.04us      4.65us ( +-   6>
  EXTERNAL_INTERRUPT        382     0.71%     0.00%      0.34us       1.57us      0.53us ( +-   1>
   PAUSE_INSTRUCTION        290     0.54%     0.00%      0.36us      33.23us      1.13us ( +-  21>
    PREEMPTION_TIMER        149     0.28%     0.00%      0.60us       2.17us      1.28us ( +-   1>
    INTERRUPT_WINDOW        122     0.23%     0.00%      0.50us       0.93us      0.64us ( +-   1>
               CPUID         24     0.04%     0.00%      0.38us       1.64us      0.72us ( +-  10>
            MSR_READ         15     0.03%     0.00%      0.75us       1.57us      1.02us ( +-   5>

Total Samples:53794, Total events handled time:27795396.68us.
```

GP-GP-BIND:



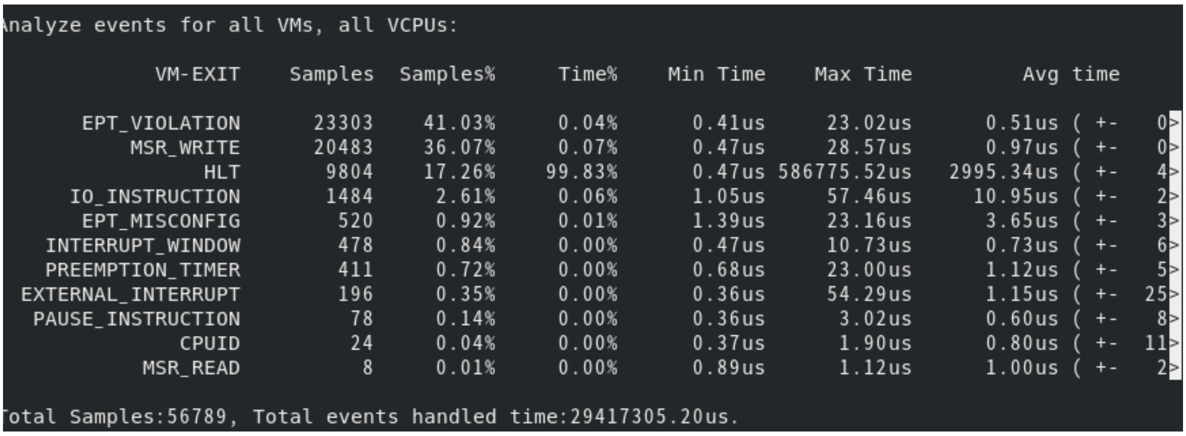
```
nalyze events for all VMs, all VCPUs:

             VM-EXIT    Samples  Samples%     Time%    Min Time    Max Time         Avg time

           MSR_WRITE      24630    41.42%     0.07%      0.46us     16.76us      0.78us ( +-   0.4>
       EPT_VIOLATION      21795    36.65%     0.04%      0.41us     15.34us      0.51us ( +-   0.7>
                 HLT      11241    18.90%    99.86%      0.46us 207917.52us   2348.91us ( +-   2.6>
      IO_INSTRUCTION        608     1.02%     0.02%      1.04us     83.23us      8.46us ( +-   5.0>
        EPT_MISCONFIG       417     0.70%     0.01%      1.56us     22.95us      3.89us ( +-   3.4>
   PAUSE_INSTRUCTION        263     0.44%     0.00%      0.36us      4.09us      0.55us ( +-   3.5>
  EXTERNAL_INTERRUPT        188     0.32%     0.00%      0.35us     10.73us      0.78us ( +-  11.7>
    PREEMPTION_TIMER        142     0.24%     0.00%      0.67us      1.94us      1.26us ( +-   1.6>
    INTERRUPT_WINDOW        139     0.23%     0.00%      0.48us      0.94us      0.63us ( +-   1.4>
               CPUID         40     0.07%     0.00%      0.35us      1.25us      0.62us ( +-   6.6>
            MSR_READ          7     0.01%     0.00%      0.83us     10.56us      2.27us ( +-  60.7>

otal Samples:59470, Total events handled time:26441694.64us.
```

GP-RT-RAW:



```
[gpf@rt-base performance]$ sudo perf kvm stat report --event=vmexit


Analyze events for all VMs, all VCPUs:

             VM-EXIT    Samples  Samples%     Time%    Min Time    Max Time         Avg time

       EPT_VIOLATION      22495    42.49%     0.04%      0.41us      9.81us      0.48us ( +-    0>
           MSR_WRITE      17238    32.56%     0.06%      0.47us     26.67us      0.97us ( +-    0>
                 HLT       9338    17.64%    99.86%      0.68us 319604.34us   2839.34us ( +-    4>
  EXTERNAL_INTERRUPT       2175     4.11%     0.00%      0.35us     24.27us      0.51us ( +-    2>
      IO_INSTRUCTION        740     1.40%     0.02%      1.01us     93.21us      8.12us ( +-    4>
       EPT_MISCONFIG        577     1.09%     0.01%      1.50us     99.08us      4.15us ( +-    6>
    PREEMPTION_TIMER        164     0.31%     0.00%      0.73us      2.19us      1.33us ( +-    1>
   PAUSE_INSTRUCTION        126     0.24%     0.00%      0.36us     47.74us      1.55us ( +-   38>
    INTERRUPT_WINDOW         57     0.11%     0.00%      0.52us      0.94us      0.68us ( +-    1>
               CPUID         24     0.05%     0.00%      0.37us      1.29us      0.66us ( +-    9>
            MSR_READ         13     0.02%     0.00%      0.75us      1.53us      1.00us ( +-    4>

Total Samples:52947, Total events handled time:26551300.39us.
```

GP-RT-BIND:



```
nalyze events for all VMs, all VCPUs:

             VM-EXIT    Samples  Samples%     Time%    Min Time    Max Time         Avg time

       EPT_VIOLATION      23303    41.03%     0.04%      0.41us     23.02us      0.51us ( +-    0>
           MSR_WRITE      20483    36.07%     0.07%      0.47us     28.57us      0.97us ( +-    0>
                 HLT       9804    17.26%    99.83%      0.47us 586775.52us   2995.34us ( +-    4>
      IO_INSTRUCTION       1484     2.61%     0.06%      1.05us     57.46us     10.95us ( +-    2>
       EPT_MISCONFIG        520     0.92%     0.01%      1.39us     23.16us      3.65us ( +-    3>
    INTERRUPT_WINDOW        478     0.84%     0.00%      0.47us     10.73us      0.73us ( +-    6>
    PREEMPTION_TIMER        411     0.72%     0.00%      0.68us     23.00us      1.12us ( +-    5>
  EXTERNAL_INTERRUPT        196     0.35%     0.00%      0.36us     54.29us      1.15us ( +-   25>
   PAUSE_INSTRUCTION         78     0.14%     0.00%      0.36us      3.02us      0.60us ( +-    8>
               CPUID         24     0.04%     0.00%      0.37us      1.90us      0.80us ( +-   11>
            MSR_READ          8     0.01%     0.00%      0.89us      1.12us      1.00us ( +-    2>

otal Samples:56789, Total events handled time:29417305.20us.
```

总结

1. 对于两种环境，绑定CPU后都会使IO_INSTRUCTION的处理次数变多，同时该事件的平均处理时间也会变长。
2. 绑定CPU会使MSR_WRITE的事件占比变多，但是处理时间基本不变。
3. GuestOS为RT时，绑定CPU后会使PREEMPTION_TIMER事件的占比变多(0.31% => 0.72%)。
4. GuestOS为RT时，绑定CPU后会使EXTERNAL_INTERRUPT事件的占比变少(4.11% => 0.35%)。

后续工作

1. 以上四种情况cyclictest测试效果最好的是第四种，即GP-RT加入CPU绑定后，之前的实验已经说明了这一点。根据上面的总结可以看出，提升性能的主要手段有减少外部中断(EXTERNAL_INTERRUPT)，提升PREEMPTTION_TIMER的次数，同时多进行写入MSR寄存器的操作。根据数据这里面最主要的因素是外部中断的发生次数。

EXTERNAL_INTERRUPT	PREEMPTION_TIMER	MSR_WRITE
↓	↑	↑

参考资料

1. EPT学习总结及KVM的处理
2. x86 CPU的MSR寄存器
3. Intel VT技术中的Preemption Timer
4. KVM: perf: kvm events analysis tool