linux下屏蔽intel cpu指令集

Linux系统调优 徐鑫 2024-09-12 发表

组网及说明

intel平台+linux系统

问题描述

由于某些应用程序特定的原因，需要将CPU指令集关闭。

如HLE与RTM两个指令集，会影响上层虚拟机迁移，需要关闭。

过程分析

绝大部分BIOS配置都不会改变CPU本身的功能，如没有确切选项的情况下，要从系统下屏蔽。

修改之前：

```
[root@localhost ~]# lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                20
On-line CPU(s) list:   0-19
Thread(s) per core:    2
Core(s) per socket:    10
Socket(s):             1
NUMA node(s):          1
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 85
Model name:            Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz
Stepping:              4
CPU MHz:               2200.000
BogoMIPS:              4400.00
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              1024K
L3 cache:              14080K
NUMA node0 CPU(s):     0-19
Flags:                 fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc art arch_perfmon pebs bts rep_good nopl xtopolog
y nonstop_tsc aperfmperf eagerfpu pni pclmulqdq dtes64 monitor ds_cpl vmx smx est tm2 ssse3 sdbg fma cx16 xtpr pdcm pcid dca sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_timer aes xsave avx f16c rdrand lahf_lm abm 3dnowprefetch epb
 cat_l3 cdp_l3 invpcid_single intel_ppin intel_pt ssbd mba ibrs ibpb stibp tpr_shadow vnmi flexpriority ept vpid fsgsbase tsc_adjust bmi1 hle avx2 smep bmi2 erms invpcid rtm cqm mpx rdt_a avx512f avx512dq rdseed adx smap clflushopt c
lwb avx512cd avx512bw avx512vl xsaveopt xsavec xgetbv1 cqm_llc cqm_occup_llc cqm_mbm_total cqm_mbm_local dtherm ida arat pln pts pku ospke md_clear spec_ctrl intel_stibp flush_l1d arch_capabilities
```

为了禁用它，我们需要从源代码中识别 x86 架构的功能/标志的位号，以及应用于引导 CMDLIN E 所需的内核参数。

我们找到 clearcpuid 参数，其描述如下：

```
clearcpuid=BITNUM[,BITNUM...] [X86]
                        Disable CPUID feature X for the kernel. See
                        arch/x86/include/asm/cpufeatures.h for the valid bit
                        numbers. Note the Linux specific bits are not necessarily
                        stable over kernel options, but the vendor specific
                        ones should be.
                        Also note that user programs calling CPUID directly
                        or using the feature without checking anything
                        will still see it. This just prevents it from
                        being used by the kernel or shown in /proc/cpuinfo.
                        Also note the kernel might malfunction if you disable
                        some critical bits.
```

我们需要找到此功能的有效位号，可以在下边源代码文件中找到该位号。

https://github.com/torvalds/linux/blob/master/arch/x86/include/asm/cpufeatures.h

```
#define X86_FEATURE_BMI1                ( 9*32+ 3) /* "bmi1" 1st group bit manipulation extensions */
#define X86_FEATURE_HLE                 ( 9*32+ 4) /* "hle" Hardware Lock Elision */
#define X86_FEATURE_AVX2                ( 9*32+ 5) /* "avx2" AVX2 instructions */
#define X86_FEATURE_FDP_EXCPTN_ONLY     ( 9*32+ 6) /* FPU data pointer updated only on x87 exceptions */
#define X86_FEATURE_SMEP                ( 9*32+ 7) /* "smep" Supervisor Mode Execution Protection */
#define X86_FEATURE_BMI2                ( 9*32+ 8) /* "bmi2" 2nd group bit manipulation extensions */
#define X86_FEATURE_ERMS                ( 9*32+ 9) /* "erms" Enhanced REP MOVSB/STOSB instructions */
#define X86_FEATURE_INVPCID             ( 9*32+10) /* "invpcid" Invalidate Processor Context ID */
#define X86_FEATURE_RTM                 ( 9*32+11) /* "rtm" Restricted Transactional Memory */
#define X86_FEATURE_CQM                 ( 9*32+12) /* "cqm" Cache QoS Monitoring */
```

HLE 功能与位号9 × 32 + 4 = 292，RTM9 × 32 + 11 = 299

修改/etc/default/grub，添加clearcpuid=292 clearcpuid=299

```
[root@localhost ~]# cat /etc/default/grub
GRUB_TIMEOUT=5
GRUB_DISTRIBUTOR="$(sed 's, release .*$,,g' /etc/system-release)"
GRUB_DEFAULT=saved
GRUB_DISABLE_SUBMENU=true
GRUB_TERMINAL_OUTPUT="console"
GRUB_CMDLINE_LINUX="crashkernel=auto spectre_v2=retpoline rd.lvm.lv=centos/root rd.lvm.lv=centos/swap rhgb quiet clearcpuid=292 clearcpuid=299"
GRUB_DISABLE_RECOVERY="true"
```

用grub2-mkconfig -o /boot/efi/EFI/centos/grub.cfg 重新构建 grub 配置文件即可禁用。

重启后功能禁用：

```
[root@localhost ~]# lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                20
On-line CPU(s) list:   0-19
Thread(s) per core:    2
Core(s) per socket:    10
Socket(s):             1
NUMA node(s):          1
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 85
Model name:            Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz
Stepping:              4
CPU MHz:               2200.000
BogoMIPS:              4400.00
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              1024K
L3 cache:              14080K
NUMA node0 CPU(s):     0-19
Flags:                 fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc art arch_perfmon pebs bts rep_good nopl xtopolog
y nonstop_tsc aperfmperf eagerfpu pni pclmulqdq dtes64 monitor ds_cpl vmx smx est tm2 ssse3 sdbg fma cx16 xtpr pdcm pcid dca sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_timer aes xsave avx f16c rdrand lahf_lm abm 3dnowprefetch epb
 cat_l3 cdp_l3 invpcid_single intel_ppin intel_pt ssbd mba ibrs ibpb stibp tpr_shadow vnmi flexpriority ept vpid fsgsbase tsc_adjust bmi1 avx2 smep bmi2 erms invpcid cqm mpx rdt_a avx512f avx512dq rdseed adx smap clflushopt clwb avx5
12cd avx512bw avx512vl xsaveopt xsavec xgetbv1 cqm_llc cqm_occup_llc cqm_mbm_total cqm_mbm_local dtherm ida arat pln pts pku ospke md_clear spec_ctrl intel_stibp flush_l1d arch_capabilities
```

解决方法

方法如上