PPC: KVM: Book3S HV: Fix conflicting use of HSTATE_HOST_R1

Bug #1867717 reported by Mike Ranweiler on 2020-03-17

This bug affects 1 person

278

Affects	Status	Importance	Assigned to	Milestone
The Ubuntu-power-systems project	Fix Released	High	Ubuntu on IBM Power Systems Bug Triage	
linux (Ubuntu)	Fix Released	Undecided	Ubuntu Security Team	
Bionic	Fix Released	Undecided	Ubuntu Security Team	

Bug Description

```
---Problem Description---
Currently a malicious user can craft a code to be executed in the guest
kernel space that puts CPU in TM suspended mode and call a hypercall (for
instance H_PUT_TERM_CHAR, token 0x58) leading to a kernel panic on host. I
was not able to reproduce it upstream, nonetheless it's reproducible on
most updated stock kernel for Ubuntu Bionic Beaver, i.e 4.15.0-76.86.
Guest kernel version is not meaningful unless TM facility is disabled (it
must be enabled).

---Steps to Reproduce---
 The following hypercall fuzzer I'll trigger it: https://github.
com/gromero/hinjector

$ git clone https://github.com/gromero/hinjector.git && cd hinjector
$ make
$ make insmod
$ sudo ./injector

Currently it's possible to crash a host from a guest by calling a
hypercall when
CPU is in TM suspended mode. Whilst on guest a TM Bad Thing is caught, on
host
the following traces are observed:

[ 618.563991] Oops: Exception in kernel mode, sig: 4 [#1]
[ 618.563994] LE SMP NR_CPUS=2048 NUMA PowerNV
[ 618.563999] Modules linked in: 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
xt_tcpudp bridge
stp llc ebtable_filter ebtables devlink ip6table_filter ip6_tables
iptable_filter
kvm_hv kvm vmx_crypto ipmi_powernv ipmi_devintf ipmi_msghandler
uio_pdrv_genirq
uio leds_powernv crct10dif_vpmsum ibmpowernv powernv_rng sch_fq_codel nfsd
auth_rpcgss
nfs_acl lockd grace sunrpc ip_tables x_tables autofs4 xfs btrfs
zstd_compress
raid10 raid456 async_raid6_recov async_memcpy async_pq async_xor async_tx
xor
raid6_pq libcrc32c raid1 raid0 multipath linear lpfc crc32c_vpmsum
nvmet_fc
nvmet nvme_fc nvme_fabrics nvme_core tg3 ipr scsi_transport_fc
[ 618.564064] CPU: 51 PID: 0 Comm: swapper/51 Not tainted 4.15.0-76-
generic #86-Ubuntu
[ 618.564066] NIP: 0000000000000000 LR: 0000000000000000 CTR:
d0000000072f0580
[ 618.564068] REGS: c00000003fd9bca0 TRAP: 0e40 Not tainted (4.15.0-76-
generic)
[ 618.564068] MSR: 9000000102883003 <SF,HV,VEC,VSX,FP,ME,RI,LE,TM[E]> CR:
28200222 XER: 20000000
[ 618.564077] CFAR: c000000000f53f0 SOFTE: 0
[ 618.564077] GPR00: 0000000000000000 c00000003fd9bf20 c00000000171c800
0000000000000000
[ 618.564077] GPR04: c000000ff4d10000 c0000000ff067400 000000000ad0cc9e
c0000000000fb4bc
[ 618.564077] GPR08: 804800000180f000 c000000dcabcbe80 0000000000000000
0000000020000000
[ 618.564077] GPR12: 0000000000000e80 c00000000faa3100 0000000000000000
0000000000000000
[ 618.564077] GPR16: 0000000000000000 0000000000000000 0000000000000000
0000000000000000
[ 618.564077] GPR20: 0000000000000000 0000000000000000 0000000000000000
0000000000000000
[ 618.564077] GPR24: 0000000000000000 d0000000072e0158 000000000000009b
000000000000009c
[ 618.564077] GPR28: 000000000000009c 0000000000000000 0000000000000000
0010000000000000
[ 618.564100] NIP [0000000000000000] (null)
[ 618.564101] LR [0000000000000000] (null)
[ 618.564101] Call Trace:
[ 618.564102] Instruction dump:
[ 618.564105] XXXXXXXX XXXXXXXX XXXXXXXX XXXXXXXX XXXXXXXX XXXXXXXX
XXXXXXXX XXXXXXXX
[ 618.564109] XXXXXXXX XXXXXXXX XXXXXXXX XXXXXXXX 0100421c f2820104
0000001b 00000132
[ 618.564118] ---[ end trace f0be3cc10ea6fc44 ]---
[ 618.569897]
[ 618.593555] KVM: CPU 51 seems to be stuck
[ 258.967652] Kernel panic - not syncing: Attempted to kill the idle task!
[ 258.967677] Unable to handle kernel paging request for data at address
0xc000001ff6c9d700
[ 618.596478] Faulting instruction address: 0xc000000000077cf0
[ 618.596479] Oops: Kernel access of bad area, sig: 11 [#2]
[ 618.596480] LE SMP NR_CPUS=2048 NUMA PowerNV
[ 618.596482] Modules linked in: 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
```

Report a bug

This report contains **Public Security** information

Everyone can see this security related information.

You are not directly subscribed to this bug's notifications.

Edit bug mail

Other bug subscribers

Subscribe someone else

Notified of all changes

- Ben Romer
- Brad Figg
- Frank Heimes
- Gustavo Romero
- Juerg Haefliger
- Leonardo Garcia
- Marcelo Cerri
- Mike Ranweiler
- Seth Forshee
- Terry Rudd
- Thadeu Lima de So...
- Ubuntu Security Team

May be notified

- Abalan
- Adriano Pangione
- Alejandro J. Alva...
- Andrew Cloke
- Andy Whitcroft
- Ashani Holland
- Bruno Garcia
- CRC
- Calub Viem
- Carles
- Charlie_Smotherman
- Christian Rüger
- Christina A Reitb...
- Cold
- Cyrus Lien
- Debian PTS
- Dmitriev Artem An...
- Doraann2
- Franko Fang
- Gavin Guo
- Guido
- H.P.J. Groeneweg
- HaySayCheese
- Hidagawa
- Hui Wang
- Jesse Jones
- Joel Robison
- John Jack
- Joseph Salisbury
- José Alfonso
- Kai-Heng Feng
- Keng-Yu Lin
- Kernel Packages
- MarcMiralles
- Marius Vlad
- Matija Marinic
- Matt j
- Michael Rowland H...
- Ming Lei
- Mr. MInhaj
- Name Changed
- PCTeacher012
- Paolo Topa
- Patricia Domingues
- Peter Bullert
- Punnsa
- Richard Seguin
- Richard Williams
- Rob Linc
- Solved
- Taihsiang Ho
- Tim Gardner
- Tom Weiss
- Tuxsimon

```
xt_tcpudp bridge stp llc ebtable_filter ebtables devlink ip6table_filter
ip6_tables iptable_filter kvm_hv kvm vmx_crypto ipmi_powernv ipmi_devintf
ipmi_msghandler uio_pdrv_genirq uio leds_powernv crct10dif_vpmsum
ibmpowernv
powernv_rng sch_fq_codel nfsd auth_rpcgss nfs_acl lockd grace sunrpc
ip_tables
x_tables autofs4 xfs btrfs zstd_compress raid10 raid456 async_raid6_recov
async_memcpy async_pq async_xor async_tx xor raid6_pq libcrc32c raid1
raid0
multipath linear lpfc crc32c_vpmsum nvmet_fc nvmet nvme_fc nvme_fabrics
nvme_core tg3 ipr scsi_transport_fc
[ 618.596521] CPU: 51 PID: 0 Comm: swapper/51 Tainted: G D 4.15.0-76-
generic #86-Ubuntu
[ 618.596522] NIP: c000000000077cf0 LR: c000000000080c84 CTR:
c000000000077c90
[ 618.596524] REGS: c00000003fd9b040 TRAP: 0300 Tainted: G D (4.15.0-76-
generic)
[ 618.596524] MSR: 9000000000001033 <SF,HV,ME,IR,DR,RI,LE> CR: 28244242
XER: 00000000
[ 618.596530] CFAR: c000000000080c80 DAR: c000001ff6c9d700 DSISR: 40000000
SOFTE: 0
[ 618.596530] GPR00: c000000000080c84 c00000003fd9b2c0 c00000000171c800
0000000006c9d700
[ 618.596530] GPR04: 00000000000001ac 0071d13aa0080040 0000000000000002
0000000000000002
[ 618.596530] GPR08: 0000000000000001 0000000000000002 00000e3a27540100
c000001ff6c9d700
[ 618.596530] GPR12: c000001ff0000000 c00000000faa3100 0000000000000000
0000000000000000
[ 618.596530] GPR16: 0000000000000004 0071d13aa0080040 00000000000001ac
c0000000018be858
[ 618.596530] GPR20: 800000000000000e d00038008004000c 0000000071d13aa
c0000000018be280
[ 618.596530] GPR24: 0000000000000001 0000000000000002 0000000000000300
0000000000000300
[ 618.596530] GPR28: 4000000000000000 0000000000000000 c0000000018be2d0
00000000000000b0
[ 618.596560] NIP [c000000000077cf0] native_hpte_updatepp+0x60/0x680
[ 618.596562] LR [c000000000080c84] __hash_page_64K+0x4c4/0x560
[ 618.596562] Call Trace:
[ 618.596563] Instruction dump:
[ 618.596565] 791cf046 3fc2001a 3bde1ad0 3d62001a 396b2188 91810008
f821ff71 7fbefa14
[ 618.596570] ebbd0048 e98b0000 7d4ae878 7d6c1a14 <7c0c1c28> 794a3e24
7f9c5378 48000018
[ 618.596576] ---[ end trace f0be3cc10ea6fc45 ]---
[ 618.602738]
[ 618.625946] KVM: CPU 51 seems to be stuck
[ 258.999498] Kernel panic - not syncing: Attempted to kill the idle task!
[ 618.653500] KVM: CPU 51 seems to be stuck
```

This is due to conflicting use of HSTATE_HOST_R1 to store r1 state in kvmppc_hv_entry plus in kvmppc_{save,restore}_tm leading to a stack corruption.

The commit that introduced such a conflict is
f024ee098476 ("KVM: PPC: Book3S HV: Pull out TM state save/restore into separate procedures")
but issue really appears when change
87a11bb6a7f7 ("KVM: PPC: Book3S HV: Work around XER[SO] bug in fake suspend mode")
is applied too because it creates a new stack to the two conflicting r1 stored to HSTATE_HOST_R1 are different.

The issue was fixed accidentally by
6f597c6b63b6 ("KVM: PPC: Book3S PR: Add guest MSR parameter for kvmppc_save_tm()/kvmppc_restore_tm()")
which is actually a change most related to Book3S PR.

This commit fixes the issue by backporting from 6f597c6b63b6 the part only responsible for storing r1 to a different memory location (HSTATE_SCRATCH2)
avoiding the conflict and so the stack corruption.

On Ubuntu Bionic, tag "Ubuntu-4.15.0-91.92" is affected.

Tags: ppc64el

CVE References

2020-8834

Mike Ranweiler (mranweil) wrote on 2020-03-17: #1

PPC: KVM: Book3S HV: Fix conflicting use of HSTATE_HOST_R1 (8.4 KiB, text/plain)

Mike Ranweiler (mranweil) wrote on 2020-03-17: #2

Xenial should not be affected - it doesn't have 87a11bb6a7f7. Since that's a power9 specific patch it's not something we would include.

There was no CVE for this right now - should we get one?

Frank Heimes (fheimes) on 2020-03-17

tags:added: ppc64el
Changed in ubuntu-power-systems:
assignee:nobody → Ubuntu Security Team (ubuntu-security)
Changed in linux (Ubuntu):
assignee:nobody → Ubuntu Security Team (ubuntu-security)
Changed in ubuntu-power-systems:
assignee:Ubuntu Security Team (ubuntu-security) → Ubuntu on IBM Power Systems Bug Triage (ubuntu-power-triage)

Seth Arnold (seth-arnold) wrote on 2020-03-18: #3

Ubuntu on IBM Pow...
Vasanth
Vic Parker
Wer Wie Was
Woodrow Shen
XxEarthxX
Yang Kun (YK)
You-Sheng Yang
Yujin.Wu
ahepas
basilisgabri
dsfkj dfjx
eoininmoran
ganesh
linuxgijs
majid hussain
nikonikic42
nobin
praveen reddy somu
projevie@hotmail.com
qadir
sankaran
scholl.w@bridge.com
van

Patches

PPC: KVM: Book3S HV: Fix conflicting use of HSTATE_HOST_R1

Add patch

Hello, I don't understand when TM is available (power8 vs power9, hardware
vs virtualized, powernv vs powervm guests, etc) -- is there a short
summary of which systems are affected, in which ways?

Please use CVE-2020-8834 for this issue.

Thanks

Gustavo Romero (gromero) wrote on 2020-03-19: #5

It seems the email reply didn't work, so pasting here again (sorry if it
yields a duplication later):

--

Hi Seth,

Well, it's a mess and confusing...

PowerVM doesn't share the same code base as KVM, so the bug doesn't affect
PowerVM, so it's KVM-specific.

POWER8 has TM supported both on baremetal (PowerNV or powernv) and on KVM
guests.
The fix involves a hypercall implemented by the KVM, so it affects the
POWER8 hosts running a KVM guest. In that case it's like the guest is
attacking the host and its (guest's kernel) kernel version is no relevant
to reproduce the issue.

POWER9 doesn't support TM on baremetal, only on KVM guests, but TM is
software assisted (due to a bug in the chip - that's the reason on the
other hand why it's not supported on baremetal, only on guests), so the
code path on the host when P9 guests use TM is a bit different.
But I haven't gone so far to able to explain why it doesn't affect P9
hosts, but it's probably be cause of the software assisted part. So, P9
hosts are not affected.

So, summing it up, it affects only POWER8 + KVM running Bionic 4.15
kernels.

BTW, I would be glad if credit could be attributed to me when filling up
the CVE details, when applicable :)

Kind regards,
Gustavo

Seth Arnold (seth-arnold) wrote on 2020-03-23: #6

Hello Gustavo, yes I can credit you with the discovery.

Thanks for the explanation of which systems are affected, it helps me a
lot.

Have you contacted other Linux distributions? IBM? Any other Power
vendors?

Is this effect of this issue still private? If so, have you already
coordinated a date with anyone else? If not, our kernel team may like to
propose a date and time that would fit nicely with currently in-progress
security issues.

Thanks

Gustavo Romero (gromero) wrote on 2020-03-24: #7

Hello Seth,

Thanks :)

No, I didn't contacted any other distro or Power vendor. IBM, well, I
think it's basically only me working with that issue at IBM. I thought of
talking to Michael Ellerman (PowerPC maintainer) but it's fixed upstream
on all stables and longterms afaics. Hence yes, effectively this issue is
still private in my understanding.

Looking upstream, I only can see that release v4.17 was affected (not
interesting anymore, right?):

f024ee098476 v4.8 -> conflict was introduced
87a11bb6a7f7 v4.17 -> commit necessary to trigger stack corruption (needs
f024ee098476)
6f597c6b63b6 v4.18 -> fixed accidentally

Thus, yeah, I think it's better to coordinate an embargo with other
distros on the closed security mailing just to let them at least try the
simple test-case on the releases they deem appropriate. I believe it will
also help Canonical to fit the fix nicely with currently in-progress
security issues and next SRUs.

HTH.

Thanks,
Gustavo

Thadeu Lima de Souza Cascardo (cascardo) wrote on 2020-03-24: #8

I looked into linux-ibm-gt as this one has the fix, but a complete
backport of the fixing commit. I backported the following 3 upstream
commits and they apply cleanly. I am suggesting that after proper testing,
we go with those 3 commits, as:

1) We end up with code more similar to upstream, making it easier to apply
any followup fixes in the future;
2) Both generic and ibm-gt end up with the same codebase, so we don't
maintain two very different codebases for 4.15.

The commits are:
7b0e827c6970e8ca77c60ae87592204c39e41245 KVM: PPC: Book3S HV: Factor fake-suspend handling out of kvmppc_save/restore_tm
009c872a8bc4d38f487a9bd62423d019e4322517 KVM: PPC: Book3S PR: Move kvmppc_save_tm/kvmppc_restore_tm to separate file
6f597c6b63b6f3675914b5ec8fcd008a58678650 KVM: PPC: Book3S PR: Add guest
MSR parameter for kvmppc_save_tm()/kvmppc_restore_tm()

Cascardo.

Steve Beattie (sbeattie) wrote on 2020-03-30: #9

The Ubuntu Security Team is proposing a CRD of Monday, April 6th, 16:00 UTC.

Gustavo Romero (gromero) wrote on 2020-03-30: #10

Hi,

I agreed on making the code base similar to the -gt kernel by applying the mentioned comments as Cascardo proposed.

Seth, btw, on the CVE credits, please add Paul Mackerras too.

Steve, thanks for update about the schedule for a fix.

Kind regards,
Gustavo

Gustavo Romero (gromero) wrote on 2020-03-30: #11

s/I agreed/I agree/
s/comments/commits/

sigh

Frank Heimes (fheimes) on 2020-03-30

```
Changed in ubuntu-power-systems:
    status:New → Triaged
```

Frank Heimes (fheimes) on 2020-04-06

```
Changed in ubuntu-power-systems:
  importance:Undecided → High
```

Steve Beattie (sbeattie) wrote on 2020-04-09: #12

Fixes for this issue were published in USN 4318-1 https://usn.ubuntu.com/4318-1/. Closing this issue on the Ubuntu side of things and making the report public.

Thanks for all your help!

```
Changed in linux (Ubuntu Bionic):
          status:New → Fix Released
        assignee:nobody → Ubuntu Security Team (ubuntu-security)
Changed in linux (Ubuntu):
          status:New → Fix Released
information type:Private Security → Public Security
```

Frank Heimes (fheimes) on 2020-04-09

```
Changed in ubuntu-power-systems:
 status:Triaged → Fix Released
```

See full activity log

To post a comment you must log in.