batman-adv - Bug #223

Kernel Crash when using more than one interface in bat0

08/20/2015 01:08 PM - Simon Wunderlich

Status:	Closed	**Start date:**	08/20/2015
Priority:	Normal	**Due date:**	
Assignee:		**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:	2016.2		

Description

Adding this issue from the mailing list:

https://lists.open-mesh.org/mailman3/hyperkitty/list/b.a.t.m.a.n@lists.open-mesh.org/message/3PN6NDV6JXKKX7CXWMRM2O3N2NKFOE2D/

```
[  879.532837] BUG: unable to handle kernel paging request at 0000000100022d60
[  879.532863] IP: [<ffffffffa04beaa5>] batadv_frag_clear_chain+0x55/0x90 [batman_adv]
[  879.532891] PGD 0
[  879.532900] Oops: 0002 [#1] SMP
[  879.532911] Modules linked in: ipt_MASQUERADE iptable_nat nf_conntrack_ipv4 nf_defrag_ipv4 nf_n
at_ipv4 nf_nat nf_conntrack ip_tables x_tables tun bridge stp llc batman_adv(O) crc32c_generic lib
crc32c ip_gre ip_tunnel gre evdev kvm_amd amd64_edac_mod kvm edac_mce_amd tpm_infineon radeon ttm
drm_kms_helper pcspkr drm i2c_algo_bit edac_core k10temp shpchp sp5100_tco i2c_piix4 i2c_core tpm_
tis tpm button acpi_cpufreq processor thermal_sys autofs4 ext4 crc16 mbcache jbd2 sg sd_mod crc_t1
0dif crct10dif_generic crct10dif_common ata_generic ohci_pci pata_atiixp ahci libahci ehci_pci ohc
i_hcd ehci_hcd libata scsi_mod tg3 ptp pps_core libphy usbcore usb_common
[  879.533106] CPU: 1 PID: 4215 Comm: kworker/u8:0 Tainted: G           O  3.16.0-4-amd64 #1 Debia
n 3.16.7-ckt11-1
[  879.533122] Hardware name: HP ProLiant MicroServer, BIOS O41     07/29/2011
[  879.533143] Workqueue: bat_events batadv_purge_orig [batman_adv]
[  879.533155] task: ffff8800d2ff8ae0 ti: ffff8800d7980000 task.ti: ffff8800d7980000
[  879.533167] RIP: 0010:[<ffffffffa04beaa5>]  [<ffffffffa04beaa5>] batadv_frag_clear_chain+0x55/0
x90 [batman_adv]
[  879.533196] RSP: 0018:ffff8800d7983d78  EFLAGS: 00010206
[  879.533208] RAX: 0000000100022d60 RBX: ffff8800d2a43ce0 RCX: 0000000000000357
[  879.533221] RDX: 0000000100023599 RSI: ffffffffa04c4440 RDI: ffff8800d2f92ce8
[  879.533234] RBP: 00005e1c0e1f72c6 R08: 00000000000000c3 R09: 0000000000000101
[  879.533247] R10: 0000000000002b67 R11: 03fffffffe062e74 R12: ffff8800d2f92ce8
[  879.533260] R13: ffffffffa04c4440 R14: 0000000000000000 R15: ffff8800d3601940
[  879.533274] FS:  00007f4c7795b700(0000) GS:ffff88011fc80000(0000) knlGS:0000000000000000
[  879.533290] CS:  0010 DS: 0000 ES: 0000 CR0: 000000008005003b
[  879.533302] CR2: 0000000100022d60 CR3: 00000000d2de7000 CR4: 00000000000007e0
[  879.533314] Stack:
[  879.533322]  ffff8800d2f92cf0 0000000000000008 ffffffffa04beb23 0000000000000000
[  879.533344]  ffff8800d3601968 ffff8800d2f92c00 0000000000000001 ffffffffa04c5144
[  879.533366]  0000000000000000 ffff8800d2f92cb0 ffff8800d2fec8c0 0000033501000000
[  879.533387] Call Trace:
[  879.533407]  [<ffffffffa04beb23>] ? batadv_frag_purge_orig+0x43/0x70 [batman_adv]
[  879.533433]  [<ffffffffa04c5144>] ? _batadv_purge_orig+0x294/0x470 [batman_adv]
[  879.533458]  [<ffffffffa04c5335>] ? batadv_purge_orig+0x15/0x40 [batman_adv]
[  879.533475]  [<ffffffff81081692>] ? process_one_work+0x172/0x420
[  879.533490]  [<ffffffff81081d23>] ? worker_thread+0x113/0x4f0
[  879.533505]  [<ffffffff8150d921>] ? __schedule+0x2b1/0x710
[  879.533519]  [<ffffffff81081c10>] ? rescuer_thread+0x2d0/0x2d0
[  879.533534]  [<ffffffff81087fad>] ? kthread+0xbd/0xe0
[  879.533550]  [<ffffffff81087ef0>] ? kthread_create_on_node+0x180/0x180
[  879.533565]  [<ffffffff81511518>] ? ret_from_fork+0x58/0x90
[  879.533580]  [<ffffffff81087ef0>] ? kthread_create_on_node+0x180/0x180
[  879.533592] Code: 48 89 03 48 b8 00 02 20 00 00 00 ad de 48 89 43 08 e8 f0 e6 f4 e0 48 89 df 48
 89 eb e8 75 f7 cc e0 48 8b 2b 48 8b 43 08 48 85 ed <48> 89 28 75 be 48 8b 7b 10 48 b8 00 01 10 00
 00 00 ad de 48 89
[  879.533729] RIP  [<ffffffffa04beaa5>] batadv_frag_clear_chain+0x55/0x90 [batman_adv]
```

```
[  879.533754]  RSP <ffff8800d7983d78>
[  879.533763] CR2: 0000000100022d60
```

It looks like there are problems in the fragmentation implementation

Related issues:		
Related to batman-adv - Bug #217: Oops: "Unable to handle kernel paging reque...	**Closed**	**06/04/2015**
Related to batman-adv - Bug #228: Workqueue: bat_events batadv_send_outstandi...	**Closed**	**01/09/2016**

History

#1 - 08/20/2015 02:58 PM - Sven Eckelmann

It looks like the crash happens during the hlist_del in batadv_frag_clear_chain. To be a little bit more specific: during the statment *pprev = next;

hlist_del tries to remove n from the hlist and thus has to modify the items prev and next

```
prev -> n -> next
```

- prev is the previous hlist_*
- n is the hlist_node which gets deleted
- next is the hlist_node which comes after the one which gets delete

And pprev is the pointer to prev->next (when prev is from type hlist_node) or prev->first (if prev is from type hlist_head as in this situation). Therefore, this statement should have set the new first element of the fragments list to next.

Now it is unknown whether the pprev pointer is invalid (list corruption even when the list is always locked with a specific chain lock?) or somebody free'd the originator below our feets. The first one should be detectable when always checking in batadv_frag_clear_chain mainloop right before hlist_del if entry->list.pprev == &head->first and create a warning when this statement is false.

#2 - 08/20/2015 06:25 PM - Sven Eckelmann

- File 0001-batman-adv-TEST-if-batadv_frag_clear_chain-is-valid.patch added

The mentioned test patch is attached. It should help to find out if the crash is caused by the fragmention code or is related to something else like the originator cleanup code.

#3 - 08/22/2015 02:18 PM - Bjoern Franke

What should be the output if it's caused by the fragmention code?

#4 - 08/22/2015 03:30 PM - Bjoern Franke

We got now the following crash:

```
[  164.525377] ------------[ cut here ]------------
[  164.525475] kernel BUG at /var/lib/dkms/batman-adv/2015.1/build/net/batman-adv/fragmentation.c:55!
[  164.525629] invalid opcode: 0000 [#1] SMP
[  164.525719] Modules linked in: ipt_MASQUERADE iptable_nat nf_conntrack_ipv4 nf_defrag_ipv4 nf_nat_ipv4 nf_n
at nf_conntrack ip_tables x_tables tun bridge stp llc batman_adv(O) crc32c_generic libcrc32c ip_gre ip_tunnel
```

```
gre radeon ttm drm_kms_helper kvm_amd drm kvm evdev k10temp i2c_algo_bit sp5100_tco i2c_piix4 pcspkr i2c_core
amd64_edac_mod edac_mce_amd edac_core tpm_infineon shpchp tpm_tis tpm button acpi_cpufreq processor thermal_sy
s autofs4 ext4 crc16 mbcache jbd2 sg sd_mod crc_t10dif crct10dif_generic crct10dif_common ata_generic ohci_pci
 ahci libahci pata_atiixp tg3 ptp pps_core libphy libata ohci_hcd ehci_pci ehci_hcd scsi_mod usbcore usb_commo
n
[  164.527233] CPU: 0 PID: 749 Comm: fastd Tainted: G           O  3.16.0-4-amd64 #1 Debian 3.16.7-ckt11-1+deb
8u3
[  164.527404] Hardware name: HP ProLiant MicroServer, BIOS O41     07/29/2011
[  164.527524] task: ffff88011af52ca0 ti: ffff8800d20f4000 task.ti: ffff8800d20f4000
[  164.527651] RIP: 0010:[<ffffffffa0338af2>]  [<ffffffffa0338af2>] batadv_frag_clear_chain+0xa2/0xb0 [batman_
adv]
[  164.527836] RSP: 0018:ffff88011fc03db8  EFLAGS: 00010217
[  164.527929] RAX: 0000000000000ec3 RBX: 000096dcb084edc6 RCX: 000000000000008e
[  164.528051] RDX: 000000000000008e RSI: 0000000000000140 RDI: ffff8800d2ad4148
[  164.528172] RBP: ffff8800d1750b20 R08: 0000000000000000 R09: 0000000000000900
[  164.528292] R10: ffff8800d0d93e00 R11: 0000000000000000 R12: 00000000ffff7b81
[  164.528413] R13: ffff88011fc03e30 R14: ffff8800d0d93e00 R15: ffff8800d2ad4140
[  164.528535] FS:  00007f97ecbe0700(0000) GS:ffff88011fc00000(0000) knlGS:0000000000000000
[  164.528673] CS:  0010 DS: 0000 ES: 0000 CR0: 0000000080050033
[  164.528771] CR2: 00007f5e4330a000 CR3: 00000000d2a22000 CR4: 00000000000007f0
[  164.528892] Stack:
[  164.528931]  ffff8800d2ad4148 ffff8800d1750b60 ffff8800d20bc84e ffffffffa0338c3c
[  164.529084]  0ec3000000000000 ffff8800d2ad4150 0000000000000000 ffff8800d2ad4000
[  164.529238]  ffff8800d20884c0 ffff8800d7582000 ffff8800d75828c0 ffff8800d2d48000
[  164.529392] Call Trace:
[  164.529437]  <IRQ>
[  164.529477]
[  164.529520]  [<ffffffffa0338c3c>] ? batadv_frag_skb_buffer+0xcc/0x430 [batman_adv]
[  164.529636]  [<ffffffffa03417af>] ? batadv_recv_frag_packet+0x1cf/0x240 [batman_adv]
[  164.529774]  [<ffffffffa033cc6b>] ? batadv_batman_skb_recv+0xcb/0x110 [batman_adv]
[  164.529909]  [<ffffffff8141cd23>] ? __netif_receive_skb_core+0x533/0x750
[  164.530026]  [<ffffffff8141db65>] ? process_backlog+0x95/0x160
[  164.530128]  [<ffffffff8141d350>] ? net_rx_action+0x140/0x240
[  164.530229]  [<ffffffff8106c611>] ? __do_softirq+0xf1/0x290
[  164.530328]  [<ffffffff815131bc>] ? do_softirq_own_stack+0x1c/0x30
[  164.530434]  <EOI>
[  164.530472]
[  164.530509]  [<ffffffff8106c84d>] ? do_softirq+0x4d/0x60
[  164.530584]  [<ffffffff8141a290>] ? netif_rx_ni+0x30/0x70
[  164.530680]  [<ffffffffa038c6a7>] ? tun_get_user+0x437/0x8d0 [tun]
[  164.530792]  [<ffffffffa038cc3b>] ? tun_chr_aio_write+0x7b/0xa0 [tun]
[  164.530905]  [<ffffffff811a794c>] ? do_sync_write+0x5c/0x90
[  164.531003]  [<ffffffff811a8252>] ? vfs_write+0xb2/0x1f0
[  164.531097]  [<ffffffff811a8d92>] ? SyS_write+0x42/0xa0
[  164.531189]  [<ffffffff8151158d>] ? system_call_fast_compare_end+0x10/0x15
[  164.535717] Code: 00 01 10 00 00 00 ad de 48 89 45 00 48 b8 00 02 20 00 00 00 ad de 48 89 45 08 e8 6a 46 0d
 e1 5b 48 89 ef 5d 41 5c e9 ee 56 e5 e0 <0f> 0b 5b 5d 41 5c c3 0f 1f 80 00 00 00 00 66 66 66 66 90 41 55
[  164.545694] RIP  [<ffffffffa0338af2>] batadv_frag_clear_chain+0xa2/0xb0 [batman_adv]
[  164.550508]  RSP <ffff88011fc03db8>
```

#5 - 08/22/2015 08:09 PM - Sven Eckelmann

If it is a problem in the fragmentation list (not necessarily the fragmentation code) then you should get the output you've posted. I personally expected a slightly different backlog but the "kernel BUG at /var/lib/dkms/batman-adv/2015.1/build/net/batman-adv/fragmentation.c:55!" is the line with the new BUG_ON

Now I am a little bit confused because I've said to Simon that the fragmentation list is properly locked everywhere. Maybe someone else could check it to make sure that I haven't overlooked something.

Btw. just to make sure: You are using batman-adv 2015.1 and the only extra modification is the patch I gave you for this test, right?

#6 - 08/22/2015 08:46 PM - Bjoern Franke

Yes, it's 2015.1 and the only patch applied is the one from above.

#7 - 08/23/2015 04:11 AM - Marek Lindner

If I remember correctly, Bjoern reported that same issue on IRC 2-3 weeks ago ?

Back then the kernel crashes appeared to be random and did not only affect fragmentation. Is that still the case ?

Also, there was a second seemingly identical server setup that did not crash ?

Have you tried a never kernel version as suggested on the mailing list ?

#8 - 08/23/2015 04:35 PM - Bjoern Franke

On IRC? Hm, maybe, but I discussed it at https://forum.freifunk.net/t/kommunikation-zwischen-den-supernodes/6888
some weeks ago. In this thread someone states that he had issues with kvm servers which were gone with vmware. In contrast to this, we have a kvm system with 2015.1 which crashes rarely, the more affected systems are dedicated servers.
On one system the crashes seem to be not only batman related, but 75 percent of them. We tried kernel 4.1 from jessie-backports, which had no positive effect.

#9 - 08/24/2015 10:18 AM - Marek Lindner

I quickly checked the forum link but couldn't find technical info that would allow me to help you. If there is info you think is valuable, please provide pointers to them or add the info directly to this ticket.

> In contrast to this, we have a kvm system with 2015.1 which crashes rarely, the more affected systems are dedicated servers.
> On one system the crashes seem to be not only batman related, but 75 percent of them. We tried kernel 4.1 from jessie-backports, which had no positive effect.

It is kind of difficult to help you if you don't provide all the information you can give us. Like: Are the kernel crash backtraces on 4.1 exactly the same ? If not, what do they look like ? Did you try Sven's patch on 4.1 ? Was the result identical ? What are the backtraces of the seemingly batman unrelated crashes ? What architecture are you running on ? What is the difference between the dedicated systems and kvm setups ?

Before I asked you did not mention there were other crashes or that some of your systems are stable. Is there anything else you can tell us, so that we can help you track this down ? Everything matters no matter how unrelated it might be.

FYI, multi-interface setups are rather common and yet you are the first to report these crashes (to the best of my knowledge). So, we are looking for something fairly unique in your setup. Can we replicate the crash without running your entire 'supernode' setup ? A simple 2 interface setup does not crash for me.

Does it still crash on your end if you turn off fragmentation ?

#10 - 08/24/2015 10:39 AM - Marek Lindner

Another question: When did the kernel crashes start ? Upon upgrading to batman-adv 2013.4/2015.1 ? Or did you activate/deactivate/change any configuration you might remember ?

#11 - 08/25/2015 12:07 PM - Matthias Fehl

Hi, i was just watching this and hoping for a resolution :D
Our multi-community setup with 6 supernodes and 3 domains is also affected by this behavior.
The problem is the same as Bjoern´s ..some of the nodes are rebooting every 20-40minutes
while some others have uptime around some hours. (sometimes they dont even reboot with kernel.panic=1)
We tested the 3.16 and the 4.1 kernel both on Debian 8 with everything (batman,batctl,alfred...) beeing 2015.1
I´ve still some crashdumps on my workstation at home, but i could also produce some more if
they are needed.

#12 - 08/25/2015 01:01 PM - Marek Lindner

Matthias, providing crash dumps is a good idea because we need to establish whether or not your issue is identical to the discussed issue.

Furthermore, it would be helpful if you answered all of the above questions to help narrow down the problem. Thanks!

#13 - 08/25/2015 01:40 PM - Bjoern Franke

Sorry for not providing all information because I thought it's unrelated.

All servers run Debian Jessie with 3.16 on 64 bit. The 2014.3 module is from the kernel, the 2015.1 are an own build dkms which uses the code from downloads.open-mesh.org. (with the patch from above).
I will check for the backtraces with 4.1 and if it also occurs when fragmentation is off. Sven's patch was not used for 4.1. The crashes started with the uprade from 2013.4 to 2014.3/2015.1.
The KVM setup has only one core, the dedicated setups are dualcores.

Maybe you can reproduce it if more than 2 nodes and 2 interfaces are used, e.g. 3 nodes which have a tunnel to each other.

A complete dump from a fragmentation related crash can be downloaded here: http://dev.ffnw.de/dumps/dump.201508242334.tar.gz

Here is one backtrace which is seemingly batman unrelated:
[1771.040934] skbuff: skb_over_panic: text:ffffffff81462934 len:1424 put:1424 head:ffff8800d207c800 data:ffff8800d207c930 tail:0x6c0 end:0x2c0 dev:<NULL>
[1771.041198] ------------[cut here]------------
[1771.041277] kernel BUG at /build/linux-ELRFVQ/linux-3.16.7-ckt11/net/core/skbuff.c:100!
[1771.041407] invalid opcode: 0000 [#1] SMP
[1771.041485] Modules linked in: tun xt_nat iptable_nat nf_conntrack_ipv4 nf_defrag_ipv4 nf_nat_ipv4 nf_nat nf_conntrack xt_TCPMSS xt_tcpudp iptable_filter ip_tables x_tables bridge stp llc ip_gre ip_tunnel gre kvm_amd radeon kvm edac_mce_amd ttm drm_kms_helper drm evdev k10temp pcspkr edac_core sp5100_tco i2c_piix4 i2c_algo_bit shpchp button i2c_core tpm_infineon tpm_tis tpm acpi_cpufreq processor thermal_sys batman_adv(O) crc32c_generic libcrc32c loop autofs4 ext4 crc16 mbcache jbd2 sg sd_mod crc_t10dif crct10dif_generic crct10dif_common ata_generic ohci_pci tg3 ptp pps_core libphy pata_atiixp ahci libahci ehci_pci ohci_hcd ehci_hcd libata scsi_mod usbcore usb_common
[1771.042800] CPU: 1 PID: 7540 Comm: ssh Tainted: G   O 3.16.0-4-amd64 #1 Debian 3.16.7-ckt11-1+deb8u3
[1771.042959] Hardware name: HP ProLiant MicroServer, BIOS O41   07/29/2011
[1771.043072] task: ffff88011a539670 ti: ffff8800d7bcc000 task.ti: ffff8800d7bcc000
[1771.043192] RIP: 0010:[<ffffffff8150ca5e>] [<ffffffff8150ca5e>] skb_panic+0x5f/0x61
[1771.043328] RSP: 0018:ffff8800d7bcfcf8 EFLAGS: 00010282
[1771.043416] RAX: 000000000000008b RBX: 0000000000000b28 RCX: 0000000000000000
[1771.043531] RDX: ffff88011fc8eda0 RSI: ffff88011fc8d478 RDI: 0000000000000246
[1771.043645] RBP: 0000000000000590 R08: 0000000000000000 R09: ffff8800d21a4500
[1771.043760] R10: 0000000000000b43 R11: 0000000000000000 R12: ffff8800d10a3e00
[1771.043875] R13: 0000000000003624 R14: ffff8800d1c9a040 R15: 00000000000009f8
[1771.043990] FS: 00007fc50b2a1800(0000) GS:ffff88011fc80000(0000) knlGS:0000000018b1e030
[1771.044121] CS: 0010 DS: 0000 ES: 0000 CR0: 0000000080050033
[1771.044214] CR2: 00007fe0ad41600f CR3: 00000000d1d5b000 CR4: 00000000000007e0
[1771.044329] Stack:
[1771.044364] ffff8800d207c930 00000000000006c0 00000000000002c0 ffffffff81736982
[1771.044501] ffffffff8140c752 ffffffff81462934 00007fc50cb70e30 0000059400000040

[1771.044637] ffff8800d1c9a160 ffff880100000000 0000000000000000 0000000000000500
[1771.044773] Call Trace:
[1771.044818] [<ffffffff8140c752>] ? skb_put+0x42/0x50
[1771.044903] [<ffffffff81462934>] ? tcp_sendmsg+0x3c4/0xd20
[1771.044996] [<ffffffff814030b6>] ? sock_aio_write+0xf6/0x120
[1771.045090] [<ffffffff814031de>] ? sock_aio_read.part.7+0xfe/0x120
[1771.045196] [<ffffffff811a794c>] ? do_sync_write+0x5c/0x90
[1771.045288] [<ffffffff811a8335>] ? vfs_write+0x195/0x1f0
[1771.045377] [<ffffffff811a811d>] ? vfs_read+0xed/0x170
[1771.045463] [<ffffffff811a8d92>] ? SyS_write+0x42/0xa0
[1771.045551] [<ffffffff8151158d>] ? system_call_fast_compare_end+0x10/0x15
[1771.045664] Code: 00 00 48 89 44 24 10 8b 87 c8 00 00 00 48 89 44 24 08 48 8b 87 d8 00 00 00 48 c7 c7 88 60 77 81 48 89 04 24 31 c0 e8 7a be ff ff <0f> 0b 48 8b 47 30 48 8b 17 80 e6 80 48 0f 44 c7 c3 0f 0b 0f 0b
[1771.046274] RIP [<ffffffff8150ca5e>] skb_panic+0x5f/0x61
[1771.046370] RSP <ffff8800d7bcfcf8>

#14 - 08/26/2015 06:17 AM - Marek Lindner

> The crashes started with the uprade from 2013.4 to 2014.3/2015.1.

That kind of makes sense because with 2014.0 the fragmentation code was entirely replaced. But also a lot of other things ..

> Maybe you can reproduce it if more than 2 nodes and 2 interfaces are used, e.g. 3 nodes which have a tunnel to each other.

What makes you think 2 or more nodes with multiple interfaces are needed ? I am asking as it sounds like you do have dual interface setups that don't crash at all ? Do you mind supplying a detailed description of your setup ? Which nodes are connected how using which interfaces ? A diagram might help (if you have one available).

Generally, it makes more sense to work on a crashing system to simplify the setup to the point that it stops crashing in order to isolate the cause. I could spend weeks trying to replicate your crashes without any progress if I am missing that piece of instability we are looking for. That is why I suggested to deactivate the fragmentation as it would help to confirm that the fragmentation is somehow involved. It could be the backtrace is bogus.

Are the KVM setups single core ? Meaning, did you assign one CPU to the KVM guests ? These KVM guests (non-crashing hosts) are identical in every aspect to the dedicated servers (crashing hosts) ? Same config, same amount of traffic, same interfaces with tunnels, etc ?

Thanks for the kernel dumps. During the weekend I'll take a closer look.

#15 - 08/26/2015 04:49 PM - Matthias Fehl

Ok i hope i got everything :)

Our setup is Debian 7.8 / 8.1 with 3.16 and 4.1 Kernel running on VMWare (hostsystem cpu=AMD)
and the six nodes are distributed on two hostsystems in different locations (of course different hosters too)
The routers are using Gluon 2015.1.2 with batman 2015.1 and the gateways are also on 2015.1
I´ve deactivated fragmentation on all gateways, but it does not make a difference on crashings.
The connecting routers using fastd with mtu of 1364 and the gateways use l2tp tunnels for the
local backbone with an mtu of 1488, the uplink to our outrouting backbone is a gre-link with mtu 1400.

the most crashing nodes are the first ones where the routers connect (node01,node02,node03...),
the other ones running around 10h - 1day till they crash and reboot.
if i shutdown the first nodes .. the last nodes (node04,5,6) are going to crash like the others before.

```
https://air-raid.de/crash/dump.tar
-------------------------------------------------------------------------------
[  343.677499] general protection fault: 0000 [#1] SMP
[  343.677544] Modules linked in: tun macvlan ebt_ip6 ebtable_filter ebtables iptable_mangle xt_mark br_netfil
ter bridge stp llc ip6t_REJECT nf_reject_ipv6 nf_log_ipv6 ip6table_filter ip6_tables ipt_REJECT nf_reject_ipv4
 nf_log_ipv4 nf_log_common xt_LOG xt_limit xt_tcpudp iptable_filter ip_tables x_tables ip_gre ip_tunnel gre io
sf_mbi coretemp crct10dif_pclmul crc32_pclmul ghash_clmulni_intel aesni_intel aes_x86_64 evdev pcspkr vmw_ball
oon lrw gf128mul glue_helper ablk_helper ppdev cryptd psmouse serio_raw acpi_cpufreq parport_pc 8250_fintek pa
rport vmwgfx ttm drm_kms_helper drm processor thermal_sys shpchp i2c_piix4 vmw_vmci battery ac button l2tp_ip
l2tp_debugfs l2tp_eth l2tp_netlink l2tp_core ip6_udp_tunnel udp_tunnel batman_adv(O) libcrc32c autofs4 ext4 cr
c16 mbcache jbd2 vmw_pvscsi sg sr_mod cdrom
[  343.677979]  sd_mod ata_generic crc32c_intel vmxnet3 mptspi scsi_transport_spi mptscsih mptbase floppy ata_
piix libata scsi_mod
[  343.678023] CPU: 0 PID: 2671 Comm: apt-get Tainted: G           O    4.1.0-1-amd64 #1 Debian 4.1.3-1
[  343.678045] Hardware name: VMware, Inc. VMware Virtual Platform/440BX Desktop Reference Platform, BIOS 6.00
 04/14/2014
[  343.678069] task: ffff8800392b4350 ti: ffff8800304a0000 task.ti: ffff8800304a0000
[  343.678086] RIP: 0010:[<ffffffff813a4346>]  [<ffffffff813a4346>] tty_poll+0x56/0x90
[  343.678122] RSP: 0018:ffff8800304a3978  EFLAGS: 00010282
[  343.678135] RAX: 6e2d737574617473 RBX: ffff88003e8ebc00 RCX: 0000000000000000
[  343.678151] RDX: 0000000080000000 RSI: 7fffffffffffffff RDI: ffff880039a0ec28
[  343.678166] RBP: ffff880039a0ec00 R08: ffff8800304a0000 R09: ffff88003e8ebc00
[  343.678182] R10: 0000000000000000 R11: 0000000000000104 R12: 0000000000000000
[  343.678204] R13: ffff8800304a3a98 R14: ffff880030b32d40 R15: 0000000000000017
[  343.678224] FS:  00007f1997673740(0000) GS:ffff88003fc00000(0000) knlGS:0000000000000000
[  343.678243] CS:  0010 DS: 0000 ES: 0000 CR0: 000000008005003b
[  343.678936] CR2: 00000000016697f8 CR3: 000000003adcf000 CR4: 00000000000407f0
[  343.679624] Stack:
[  343.680328]  0000000002800001 0000000000000040 0000000000800000 ffff88003e8ebc00
[  343.680911]  000000000000001a ffffffff811de753 0000000000000000 ffffffff81164265
[  343.681691]  0000000002800001 0000000000000000 0000000000e76ae0 ffff88003e8ebc00
[  343.682267] Call Trace:
[  343.682897]  [<ffffffff811de753>] ? do_select+0x343/0x780
[  343.683462]  [<ffffffff81164265>] ? __rmqueue+0x95/0x490
[  343.683986]  [<ffffffff811954ab>] ? __insert_vmap_area+0x7b/0xc0
[  343.684525]  [<ffffffff811de180>] ? poll_select_copy_remaining+0x140/0x140
[  343.685088]  [<ffffffff811de180>] ? poll_select_copy_remaining+0x140/0x140
[  343.685679]  [<ffffffff811de180>] ? poll_select_copy_remaining+0x140/0x140
[  343.686185]  [<ffffffff812dabdf>] ? cpumask_any_but+0x2f/0x40
[  343.686696]  [<ffffffff81067998>] ? flush_tlb_page+0x38/0x90
[  343.687261]  [<ffffffff8119910c>] ? ptep_clear_flush+0x4c/0x60
[  343.687743]  [<ffffffff81193b51>] ? page_add_new_anon_rmap+0x71/0xa0
[  343.688192]  [<ffffffff81185d86>] ? wp_page_copy.isra.53+0x2d6/0x500
[  343.688696]  [<ffffffff811ded19>] ? core_sys_select+0x189/0x2a0
[  343.689167]  [<ffffffff81073f8b>] ? wait_consider_task+0x8b/0xbf0
[  343.689660]  [<ffffffff81189c2b>] ? handle_mm_fault+0xd5b/0x1640
[  343.690164]  [<ffffffff812f7ec9>] ? list_del+0x9/0x30
[  343.690721]  [<ffffffff810b0df0>] ? remove_wait_queue+0x20/0x30
[  343.691267]  [<ffffffff8107df77>] ? recalc_sigpending+0x17/0x50
[  343.691865]  [<ffffffff8107ea3d>] ? __set_task_blocked+0x2d/0x70
[  343.692362]  [<ffffffff81081485>] ? __set_current_blocked+0x35/0x90
[  343.692849]  [<ffffffff811df044>] ? SyS_pselect6+0x124/0x250
[  343.693340]  [<ffffffff81576132>] ? system_call_fast_compare_end+0xc/0x6b
[  343.693818] Code: 48 89 ef e8 cd f3 ff ff 85 c0 74 11 5b 44 89 e0 5d 41 5c 41 5d 41 5e c3 0f 1f 44 00 00 48
 89 ef e8 10 81 00 00 49 89 c6 48 8b 00 <48> 8b 40 60 48 85 c0 74 0e 4c 89 ea 48 89 de 48 89 ef ff d0 41
[  343.694932] RIP  [<ffffffff813a4346>] tty_poll+0x56/0x90
[  343.695289]  RSP <ffff8800304a3978>
---------------------------------------
```

Hopefully you got what you need now, meanwhile i´m going to rent an ape that resets my vm´s every hour ;D

#16 - 08/27/2015 12:22 AM - Bjoern Franke

With fragmentation turned off, we get such crashes:

```
[19546.528525] general protection fault: 0000 [#1] SMP
[19546.528643] Modules linked in: ipt_MASQUERADE iptable_nat nf_conntrack_ipv4 nf_defrag_ipv4 nf_nat_ipv4 nf_n
at nf_conntrack ip_tables x_tables tun bridge stp llc batman_adv(O) crc32c_generic libcrc32c ip_gre ip_tunnel
 gre radeon evdev amd64_edac_mod sp5100_tco pcspkr ttm edac_mce_amd drm_kms_helper drm edac_core kvm_amd i2c_al
go_bit kvm i2c_piix4 i2c_core k10temp tpm_infineon tpm_tis shpchp tpm button acpi_cpufreq processor thermal_sy
s autofs4 ext4 crc16 mbcache jbd2 sg sd_mod crc_t10dif crct10dif_generic crct10dif_common ata_generic ohci_pci
 tg3 ptp pps_core libphy pata_atiixp ahci libahci ohci_hcd ehci_pci ehci_hcd libata scsi_mod usbcore usb_commo
n
[19546.530023] CPU: 0 PID: 2850 Comm: kworker/0:2 Tainted: G           O  3.16.0-4-amd64 #1 Debian 3.16.7-ckt1
1-1+deb8u3
[19546.530207] Hardware name: HP ProLiant MicroServer, BIOS O41     07/29/2011
[19546.530335] Workqueue: events cache_reap
[19546.530410] task: ffff88011a51a390 ti: ffff88011a740000 task.ti: ffff88011a740000
[19546.530540] RIP: 0010:[<ffffffff8118d053>]  [<ffffffff8118d053>] free_block+0xe3/0x1c0
[19546.530687] RSP: 0018:ffff88011a743d78  EFLAGS: 00010046
[19546.530781] RAX: ffff8800d2551000 RBX: ffff88011b3c4118 RCX: dead000000100100
[19546.530904] RDX: dead000000200200 RSI: ffffea0002e029b8 RDI: dead000000100100
[19546.531027] RBP: ffff88011ac71dc0 R08: ffff88011ac13a40 R09: 0000000000000000
[19546.531150] R10: ffff8800d282e700 R11: ffff8800d282e600 R12: 0000000080000000
[19546.531273] R13: ffffea0000000000 R14: 000077ff80000000 R15: ffff88011b3c4128
[19546.531397] FS:  00007fb3f10c2700(0000) GS:ffff88011fc00000(0000) knlGS:0000000000000000
[19546.531538] CS:  0010 DS: 0000 ES: 0000 CR0: 000000008005003b
[19546.531638] CR2: 00007fb0cb0d3000 CR3: 00000000d2d64000 CR4: 00000000000007f0
[19546.531761] Stack:
[19546.531799]  0000000000000000 ffff88011b3c4100 ffff88011ac13a40 0000000000000002
[19546.531946]  ffff88011ac71dc0 0000000000000002 ffff88011b3c4118 ffffffff8118d322
[19546.532093]  000000001af28290 ffff88011ac13a40 ffff88011ac71dc0 ffff88011fc12700
[19546.532239] Call Trace:
[19546.532288]  [<ffffffff8118d322>] ? drain_array+0xc2/0x120
[19546.532388]  [<ffffffff8118d532>] ? cache_reap+0x82/0x230
[19546.532484]  [<ffffffff81081662>] ? process_one_work+0x172/0x420
[19546.532591]  [<ffffffff81081cf3>] ? worker_thread+0x113/0x4f0
[19546.532694]  [<ffffffff8150d8c1>] ? __schedule+0x2b1/0x710
[19546.532792]  [<ffffffff81081be0>] ? rescuer_thread+0x2d0/0x2d0
[19546.532898]  [<ffffffff81087f7d>] ? kthread+0xbd/0xe0
[19546.532990]  [<ffffffff81087ec0>] ? kthread_create_on_node+0x180/0x180
[19546.533106]  [<ffffffff815114d8>] ? ret_from_fork+0x58/0x90
[19546.533206]  [<ffffffff81087ec0>] ? kthread_create_on_node+0x180/0x180
[19546.533318] Code: 01 ea 48 8b 0a 80 e5 80 0f 85 d9 00 00 00 48 89 d6 48 8b 55 78 48 8b 4e 20 48 bf 00 01 10
 00 00 00 ad de 4e 8b 04 0a 48 8b 56 28 <48> 89 51 08 48 89 0a 48 89 7e 20 48 2b 46 08 48 bf 00 02 20 00
[19546.533996] RIP  [<ffffffff8118d053>] free_block+0xe3/0x1c0
[19546.534100]  RSP <ffff88011a743d78>
```

With turning off one core on the dualcore supernodes (and enabled fragmenation), there are no crashes so far. I have no information about the kvm host, it's a rented "Vserver", but only one core is visible in the installed OS.

The supernodes are not completely identical. They have different CPUs, different amount of RAM, different network cards. Some of them have ipgre-tunnels with bird running on them, some have openvpn-tunnels. The traffic is also different, it differs from 5 to 20 TB/month.

But the setup in which batman is involved is identical.
We have 5 supernodes, each are connected to each other via a gretap tunnel. Previously we used tinc and had the same issues. I'll make a diagram tomorrow.

Matthias, do you have fragmentation enabled in your setup?

#17 - 08/27/2015 03:01 AM - Marek Lindner

Matthias Fehl wrote:

> I´ve deactivated fragmentation on all gateways, but it does not make a difference on crashings.
> The connecting routers using fastd with mtu of 1364 and the gateways use l2tp tunnels for the
> local backbone with an mtu of 1488, the uplink to our outrouting backbone is a gre-link with mtu 1400.

If I understand your setup correctly the smallest MTU you have is on each node (fastd 1364) and not on the servers that are crashing ? In that case the fragmentation on the servers is irrelevant because the packets already arrive fragmented and are forwarded as-is. You could run 'batctl td' or wireshark to confirm that.

> the most crashing nodes are the first ones where the routers connect (node01,node02,node03...),
> the other ones running around 10h - 1day till they crash and reboot.
> if i shutdown the first nodes .. the last nodes (node04,5,6) are going to crash like the others before.

That is an interesting observation. Do you have any theory as to why that would be the case ? Is there any preference in the firmware on the nodes towards the first 3 gateways ? Maybe somehow linked to your tunneling software ?

> [343.682267] Call Trace:
> [343.682897] [<ffffffff811de753>] ? do_select+0x343/0x780
> [343.683462] [<ffffffff81164265>] ? *__rmqueue+0x95/0x490*
> *[343.683986] [<ffffffff811954ab>] ? __insert_vmap_area+0x7b/0xc0*
> *[343.684525] [<ffffffff811de180>] ? poll_select_copy_remaining+0x140/0x140*
> *[343.685088] [<ffffffff811de180>] ? poll_select_copy_remaining+0x140/0x140*
> *[343.685679] [<ffffffff811de180>] ? poll_select_copy_remaining+0x140/0x140*
> *[343.686185] [<ffffffff812dabdf>] ? cpumask_any_but+0x2f/0x40*
> *[343.686696] [<ffffffff81067998>] ? flush_tlb_page+0x38/0x90*
> *[343.687261] [<ffffffff8119910c>] ? ptep_clear_flush+0x4c/0x60*
> *[343.687743] [<ffffffff81193b51>] ? page_add_new_anon_rmap+0x71/0xa0*
> *[343.688192] [<ffffffff81185d86>] ? wp_page_copy.isra.53+0x2d6/0x500*
> *[343.688696] [<ffffffff811ded19>] ? core_sys_select+0x189/0x2a0*
> *[343.689167] [<ffffffff81073f8b>] ? wait_consider_task+0x8b/0xbf0*
> *[343.689660] [<ffffffff81189c2b>] ? handle_mm_fault+0xd5b/0x1640*
> *[343.690164] [<ffffffff812f7ec9>] ? list_del+0x9/0x30*
> *[343.690721] [<ffffffff810b0df0>] ? remove_wait_queue+0x20/0x30*
> *[343.691267] [<ffffffff8107df77>] ? recalc_sigpending+0x17/0x50*
> *[343.691865] [<ffffffff8107ea3d>] ? __set_task_blocked+0x2d/0x70*
> *[343.692362] [<ffffffff81081485>] ? __set_current_blocked+0x35/0x90*
> *[343.692849] [<ffffffff811df044>] ? SyS_pselect6+0x124/0x250*
> *[343.693340] [<ffffffff81576132>] ? system_call_fast_compare*

That crash does not really look like batman-adv related. What made you believe your crashes are somehow linked to batman-adv ? Did they start appearing after you enabled some feature or after upgrading to a newer batman-adv version ?

You could apply Sven's debug patch to see whether batman-adv is somehow involved but since the crashes continued with fragmentation turned off I have little hope it'll do.

#18 - 08/27/2015 03:31 AM - Marek Lindner

Bjoern Franke wrote:

> With fragmentation turned off, we get such crashes:
> [...]

Odd, didn't know kthread_create() could crash that way. Did the crash followed the usual behavior ? Meaning, did it take longer to crash or was it faster ?

Is your smallest MTU also in the routers and not on the gateway itself ? Or did you have to re-configure something to make the test with turned off fragmentation work ?

> With disabling one core on the dualcore supernodes (and enabled fragmenation), there are no crashes so far.

Ok, good! That is a valuable piece of information.

> The supernodes are not completely identical. They have different CPUs, different amount of RAM, different network cards. Some of them have ipgre-tunnels with bird running on them, some have openvpn-tunnels. The traffic is also different, it differs from 5 to 20 TB/month.
>
> But the setup in which batman is involved is identical.
> We have 5 supernodes, each are connected to each other via a gretap tunnel. Previously we used tinc and had the same issues.

Ok.

> I'll make a diagram tomorrow.

Thanks!

#19 - 08/27/2015 08:44 AM - Matthias Fehl

Good morning booth of you :)

I´ve tried with and without fragmentation, but it seems unrelevant (the clients are also forced to 1280 via dhcp).
We think it must have to do something with compat 15, because this vm´s are just clones from other communitys that are still running compat 14 and this machines had uptimes about 2-3 weeks before.
Right after the switch to compat 15 the crashs started, but we wanted to look into everything else before wanted to bother you with this issue.
We also did a fresh and clean install of debian 8.1 with everything build from scratch, because we where first thinking that maybe gretap had an issue (saw smth on the kernel mailinglist).
So we tried fastd for routers and gretap to link the gateways, then we switched to fastd for the gateways..and now we link them with l2tp, but in every setup it keeps crashing.
I could also make a diagramm of our setup like Björn and include the complete configuration if you like to :)

```
[18366.088278] general protection fault: 0000 [#1] SMP
[18366.088312] Modules linked in: tun macvlan ebt_ip6 ebtable_filter ebtables iptable_mangle xt_mark br_netfil
ter bridge stp llc ip6t_REJECT nf_reject_ipv6 $
[18366.088625]  sd_mod ata_generic crc32c_intel floppy vmxnet3 ata_piix mptspi scsi_transport_spi mptscsih lib
ata mptbase scsi_mod
[18366.088684] CPU: 1 PID: 5609 Comm: kworker/u4:2 Tainted: G           O    4.1.0-1-amd64 #1 Debian 4.1.3-1
[18366.088744] Hardware name: VMware, Inc. VMware Virtual Platform/440BX Desktop Reference Platform, BIOS 6.00
 04/14/2014
[18366.088803] Workqueue: bat_events batadv_purge_orig [batman_adv]
[18366.088829] task: ffff88003cf18be0 ti: ffff88003b07c000 task.ti: ffff88003b07c000
[18366.088853] RIP: 0010:[<ffffffffa01e4115>]  [<ffffffffa01e4115>] batadv_frag_clear_chain+0x55/0xa0 [batman_
adv]
[18366.088897] RSP: 0018:ffff88003b07fd28  EFLAGS: 00010206
[18366.088914] RAX: f907c19900200180 RBX: ffff88003c0759a0 RCX: 0000000000000c176
[18366.088940] RDX: 0000000010045069e RSI: 0000000000000200 RDI: ffff88003c471508
[18366.088966] RBP: 5e2c97a802000100 R08: 00000000000094a6 R09: 0000000000000009
[18366.088992] R10: ffff88003d420000 R11: ffff88003cbf6c30 R12: ffffffffa01ea410
[18366.089018] R13: ffff88003c4715f0 R14: ffff88003c8eb840 R15: ffff880036ce9368
[18366.089908] FS:  0000000000000000(0000) GS:ffff88003fd00000(0000) knlGS:0000000000000000
[18366.090489] CS:  0010 DS: 0000 ES: 0000 CR0: 000000008005003b
[18366.091079] CR2: 00007f0030684000 CR3: 000000003c838000 CR4: 00000000000407e0
[18366.091867] Stack:
[18366.092705]  ffff88003b7380c0 ffff88003c471510 ffff88003c471508 ffffffffa01e41a3
[18366.093419]  ffff88003b7380c0 0000000000000000 0000000000000000 0000000000000001
[18366.094255]  ffff88003c471400 ffffffffa01eb1dd ffff880000000000 0000000000000000
[18366.095014] Call Trace:
[18366.096582]  [<ffffffffa01e41a3>] ? batadv_frag_purge_orig+0x43/0x70 [batman_adv]
[18366.097537]  [<ffffffffa01eb1dd>] ? _batadv_purge_orig+0x2ad/0x5b0 [batman_adv]
[18366.098535]  [<ffffffffa01eb4f5>] ? batadv_purge_orig+0x15/0x40 [batman_adv]
[18366.099494]  [<ffffffff8108aa1b>] ? process_one_work+0x14b/0x420
[18366.100374]  [<ffffffff8108b543>] ? worker_thread+0x53/0x550
[18366.101208]  [<ffffffff8108b4f0>] ? rescuer_thread+0x380/0x380
[18366.101928]  [<ffffffff81090981>] ? kthread+0xc1/0xe0
[18366.102552]  [<ffffffff810908c0>] ? kthread_create_on_node+0x180/0x180
[18366.103252]  [<ffffffff81576572>] ? ret_from_fork+0x42/0x70
[18366.103931]  [<ffffffff810908c0>] ? kthread_create_on_node+0x180/0x180
[18366.104685] Code: 48 89 03 48 b8 00 02 20 00 00 00 ad de 48 89 43 08 e8 e0 23 28 e1 48 89 df 48 89 eb e8 d5
 b9 fc e0 48 8b 2b 48 8b 43 08 48 85 ed <48> 8$
[18366.106799] RIP  [<ffffffffa01e4115>] batadv_frag_clear_chain+0x55/0xa0 [batman_adv]
[18366.107351]  RSP <ffff88003b07fd28>

this one is from today.

---------------- strange? ------------------
[ 5945.937671] kernel tried to execute NX-protected page - exploit attempt? (uid: 0)
[ 5945.937711] BUG: unable to handle kernel paging request at ffff88003d0175c8
[ 5945.937751] IP: [<ffff88003d0175c8>] 0xffff88003d0175c8
[ 5945.937778] PGD 1b24067 PUD 1b25067 PMD 800000003d0001e3
[ 5945.937805] Oops: 0011 [#1] SMP
-------------------------------------------------------
```

#20 - 08/27/2015 05:07 PM - Matthias Fehl

Great tip Björn!
I´ve switchet to single core and so far no crash on all six nodes today.

#21 - 08/27/2015 10:39 PM - Bjoern Franke

- File networksetup 2015.8.png added

Marek Lindner wrote:

> Odd, didn't know kthread_create() could crash that way. Did the crash followed the usual behavior ? Meaning, did it take longer to crash or was it faster ?

I think it take longer and the crashes are fewer (before: 20-30 a day, after: 4-5 a day).

> Is your smallest MTU also in the routers and not on the gateway itself ? Or did you have to re-configure something to make the test with turned off fragmentation work ?

I only tested if the setup crashes with turned off fragmentation, not if data could be transmitted. But the VPN-links (Router-Supernode & Supernode-Supernode) have a MTU of 1312, so I think this could fit.

#22 - 08/28/2015 10:24 AM - Marek Lindner

Matthias Fehl wrote:

> I´ve switchet to single core and so far no crash on all six nodes today.

Thanks for testing the single core setup. It would have been my next suggestion. So, it seems you are seeing the same issue which only manifests itself on SMP systems.

#23 - 08/28/2015 11:24 AM - Marek Lindner

Sven Eckelmann wrote:

> Now I am a little bit confused because I've said to Simon that the fragmentation list is properly locked everywhere. Maybe someone else could check it to make sure that I haven't overlooked something.

Sven, could you comment on the following theory:

batadv_frag_check_entry() and batadv_frag_init_chain() rely on hlist_empty() to check whether or not fragments are present. However, the fragmentation code does not use hlist_del_init() which sets the pointers to NULL. Thus, the check could lead to random kfree()'s ?

#24 - 08/28/2015 11:57 AM - Marek Lindner

Marek Lindner wrote:

> Sven, could you comment on the following theory:
>
> batadv_frag_check_entry() and batadv_frag_init_chain() rely on hlist_empty() to check whether or not fragments are present. However, the fragmentation code does not use hlist_del_init() which sets the pointers to NULL. Thus, the check could lead to random kfree()'s ?

Scratch that. Couldn't find a codepath leading to an issue because of that.

#25 - 08/29/2015 12:42 AM - Matthias Fehl

Marek if i switch just "one" gateway back to more than one cpu its like
that there is now a shitload of stuff in a queue unloading all at once.

```
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 18:07 - 00:38  (06:31)
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 18:06 - 00:38  (06:31)
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 18:05 - 00:38  (06:33)
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 18:04 - 00:38  (06:33)
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 18:03 - 00:38  (06:34)
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 18:03 - 00:38  (06:34)
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 17:59 - 00:38  (06:38)
reboot   system boot  4.1.0-1-amd64    Fri Aug 28 17:59 - 00:38  (06:39)
```

before it was every 5-10min on this gateway.
btw. if its helping, i could give access to this node.

#26 - 08/29/2015 10:33 AM - Bjoern Franke

We tested now again with 2015.1, kernel 4.1 and fragmentation enabled. 2 reboots in 10 hours, this is much better than with 3.16.

```
[38409.734812] BUG: unable to handle kernel NULL pointer dereference at 0000000000000074
[38409.739111] IP: [<ffffffff811b04ee>] ksize+0x4e/0x70
[38409.743306] PGD 0
[38409.747416] Oops: 0000 [#1] SMP
[38409.751453] Modules linked in: nls_utf8 fuse btrfs xor raid6_pq ufs qnx4 hfsplus hfs minix ntfs vfat msdos
fat jfs xfs dm_mod ipt_MASQUERADE nf_nat_masquerade_ipv4 iptable_nat nf_conntrack_ipv4 nf_defrag_ipv4 nf_nat_i
pv4 nf_nat nf_conntrack ip_tables x_tables tun bridge stp llc batman_adv(O) crc32c_generic libcrc32c ip_gre ip
_tunnel gre kvm_amd radeon kvm evdev sp5100_tco k10temp i2c_piix4 ttm pcspkr drm_kms_helper drm amd64_edac_mod
 edac_mce_amd edac_core i2c_algo_bit tpm_infineon tpm_tis tpm shpchp button acpi_cpufreq processor thermal_sys
 autofs4 ext4 crc16 mbcache jbd2 sg sd_mod ata_generic ohci_pci ahci libahci pata_atiixp ohci_hcd ehci_pci ehc
i_hcd tg3 ptp pps_core libphy libata scsi_mod usbcore usb_common
[38409.778987] CPU: 0 PID: 802 Comm: openvpn Tainted: G           O    4.1.0-0.bpo.1-amd64 #1 Debian 4.1.3-1~b
po8+1
[38409.783980] Hardware name: HP ProLiant MicroServer, BIOS O41     07/29/2011
[38409.789038] task: ffff8800d0af5470 ti: ffff88011a06c000 task.ti: ffff88011a06c000
[38409.794189] RIP: 0010:[<ffffffff811b04ee>]  [<ffffffff811b04ee>] ksize+0x4e/0x70
[38409.799393] RSP: 0018:ffff88011fc03740  EFLAGS: 00010246
[38409.804594] RAX: 0000000000000000 RBX: ffff8800d1d34500 RCX: 0000000000000004
[38409.809867] RDX: 01ffff800002006c RSI: ffff88011fc1c900 RDI: 000077ff80000000
[38409.815130] RBP: 0000000000000080 R08: ffffea0003498ca0 R09: ffff88011fc1c900
[38409.820394] R10: 00000000d2bc109a R11: 0000000000000000 R12: 0000000000000080
[38409.825676] R13: ffff8800d19d0000 R14: ffff8800d78c22ac R15: 0000000000000020
[38409.830938] FS:  00007efdf8729700(0000) GS:ffff88011fc00000(0000) knlGS:0000000000000000
[38409.836240] CS:  0010 DS: 0000 ES: 0000 CR0: 0000000080050033
[38409.841552] CR2: 0000000000000074 CR3: 000000011a089000 CR4: 00000000000007f0
[38409.846919] Stack:
[38409.852278]  ffffffff8146a831 000000000000052e ffff8800d78c22ac 0000000000000001
[38409.857844]  ffff8800d1d34500 ffff88011a604000 ffff8800d1d34500 000000000000009e
[38409.863399]  ffff8800d78c22ac 0000000000000001 ffffffffa0371915 ffff8800d1d34500
[38409.868880] Call Trace:
[38409.874226]  <IRQ>
[38409.874271]  [<ffffffff8146a831>] ? pskb_expand_head+0x81/0x250
[38409.884886]  [<ffffffffa0371915>] ? gre_tap_xmit+0xc5/0x110 [ip_gre]
[38409.890155]  [<ffffffff8147d7dc>] ? dev_hard_start_xmit+0x24c/0x3b0
[38409.895435]  [<ffffffff8149f85f>] ? sch_direct_xmit+0xcf/0x1f0
[38409.900650]  [<ffffffff8147db76>] ? __dev_queue_xmit+0x236/0x570
[38409.905836]  [<ffffffffa03836cb>] ? batadv_frag_send_packet+0x2eb/0x360 [batman_adv]
[38409.910954]  [<ffffffff81474c0a>] ? dev_get_by_index+0xa/0x20
[38409.916031]  [<ffffffffa038c726>] ? batadv_send_skb_to_orig+0x86/0xb0 [batman_adv]
[38409.921127]  [<ffffffffa038c90e>] ? batadv_send_skb_unicast+0x8e/0x100 [batman_adv]
[38409.926175]  [<ffffffffa038dd21>] ? batadv_interface_tx+0x3e1/0x430 [batman_adv]
[38409.931219]  [<ffffffff8149f7ee>] ? sch_direct_xmit+0x5e/0x1f0
[38409.936235]  [<ffffffff8147d7dc>] ? dev_hard_start_xmit+0x24c/0x3b0
[38409.941238]  [<ffffffff8147de45>] ? __dev_queue_xmit+0x505/0x570
[38409.946172]  [<ffffffffa03b6c5e>] ? br_dev_queue_push_xmit+0x4e/0x70 [bridge]
[38409.951057]  [<ffffffffa03b6ca5>] ? br_forward_finish+0x25/0x80 [bridge]
[38409.955883]  [<ffffffffa03f19a9>] ? __nf_ct_refresh_acct+0xa9/0xc0 [nf_conntrack]
[38409.960682]  [<ffffffffa03b6d4f>] ? __br_deliver+0x4f/0x120 [bridge]
[38409.965473]  [<ffffffffa03b728b>] ? br_deliver+0x3b/0x70 [bridge]
[38409.970236]  [<ffffffffa03b48b5>] ? br_dev_xmit+0x135/0x240 [bridge]
[38409.974930]  [<ffffffff8147d7dc>] ? dev_hard_start_xmit+0x24c/0x3b0
[38409.979577]  [<ffffffff8147de45>] ? __dev_queue_xmit+0x505/0x570
[38409.984214]  [<ffffffff814ba72e>] ? ip_finish_output+0x4be/0x860
[38409.988862]  [<ffffffff814bc43b>] ? ip_output+0x6b/0xc0
[38409.993380]  [<ffffffff814ba270>] ? ip_fragment+0xa60/0xa60
[38409.997766]  [<ffffffff814b7a23>] ? ip_forward+0x3a3/0x4a0
[38410.002020]  [<ffffffff814b6034>] ? ip_rcv+0x294/0x3c0
[38410.006138]  [<ffffffff814b56a0>] ? inet_del_offload+0x40/0x40
[38410.010143]  [<ffffffff8147b7d5>] ? __netif_receive_skb_core+0x7e5/0x990
[38410.014051]  [<ffffffff8147c712>] ? process_backlog+0xa2/0x140
[38410.017852]  [<ffffffff8147bf22>] ? net_rx_action+0x212/0x340
[38410.021560]  [<ffffffff810772da>] ? __do_softirq+0x11a/0x290
[38410.025212]  [<ffffffff8157aefc>] ? do_softirq_own_stack+0x1c/0x30
[38410.028824]  <EOI>
[38410.028865]  [<ffffffff81076b95>] ? do_softirq.part.20+0x35/0x40
[38410.035948]  [<ffffffff8147a71f>] ? netif_rx_ni+0x2f/0x70
[38410.039529]  [<ffffffffa03cf570>] ? tun_get_user+0x520/0x9e0 [tun]
[38410.043114]  [<ffffffff81460201>] ? move_addr_to_user+0xb1/0xd0
[38410.046688]  [<ffffffffa03cfaf1>] ? tun_chr_write_iter+0x51/0x80 [tun]
[38410.050281]  [<ffffffff811cad48>] ? new_sync_write+0x88/0xb0
[38410.053821]  [<ffffffff811cb3e4>] ? vfs_write+0xa4/0x1b0
```

```
[38410.057298]  [<ffffffff811cb31c>] ? vfs_read+0x10c/0x130
[38410.060718]  [<ffffffff811cc152>] ? SyS_write+0x42/0xb0
[38410.064071]  [<ffffffff815793b2>] ? system_call_fast_compare_end+0xc/0x6b
[38410.067409] Code: bf 00 00 00 80 ff 77 00 00 48 0f 42 3d 3c 3b 66 00 48 01 f8 48 c1 e8 0c 48 c1 e0 06 48 01
 d0 48 8b 10 80 e6 80 75 1e 48 8b 40 30 <48> 63 40 74 c3 0f 1f 44 00 00 0f 0b 66 0f 1f 44 00 00 31 c0 c3
[38410.074809] RIP  [<ffffffff811b04ee>] ksize+0x4e/0x70
[38410.078326]  RSP <ffff88011fc03740>
[38410.081851] CR2: 0000000000000074
```

#27 - 08/29/2015 01:55 PM - Bjoern Franke

Sorry, I forgot to mention that the test was with both cores.

#28 - 08/29/2015 02:36 PM - Linus Lüssing

Hi Bjoern,

The last call trace looks different. Note that the module actually crashing is ip_gre ("gre_tap_xmit+0xc5/0x110 [ip_gre]").

Hm, for GRE it seems there is no explicit mailinglist to contact, would the netdev mailinglist be the right place to forward that this particular calltrace to?

#29 - 08/29/2015 07:18 PM - Antonio Quartulli

Linus Lüssing wrote:

> Hm, for GRE it seems there is no explicit mailinglist to contact, would the netdev mailinglist be the right place to forward that this particular calltrace to?

```
# ./scripts/get_maintainer.pl -f net/ipv4/ip_gre.c
"David S. Miller" <davem@davemloft.net> (maintainer:NETWORKING [IPv4/...)
Alexey Kuznetsov <kuznet@ms2.inr.ac.ru> (maintainer:NETWORKING [IPv4/...)
James Morris <jmorris@namei.org> (maintainer:NETWORKING [IPv4/...)
Hideaki YOSHIFUJI <yoshfuji@linux-ipv6.org> (maintainer:NETWORKING [IPv4/...)
Patrick McHardy <kaber@trash.net> (maintainer:NETWORKING [IPv4/...)
netdev@vger.kernel.org (open list:NETWORKING [IPv4/...)
linux-kernel@vger.kernel.org (open list)
```

definitely netdev mailing list.

#30 - 08/30/2015 02:47 AM - Marek Lindner

Matthias Fehl wrote:

> https://air-raid.de/crash/dump.tar

Just checked the dump file: How many interfaces to you actually add to batman-adv ? Looks like a ton .. :)

What is the larger dump file for ?

Do you crashes on 4.1 also go away by limiting the systems to a single core setup ?

Do all 4.1 crashes look batman-adv unrelated ?

#31 - 08/30/2015 02:55 AM - Marek Lindner

Linus Lüssing wrote:

> The last call trace looks different. Note that the module actually crashing is ip_gre ("gre_tap_xmit+0xc5/0x110 [ip_gre]").

I agree, this crash looks quite different. @Bjoern: Do you get these batman-adv fragmentation backtraces with 4.1 at all or do they all look like what you posted above ? If so, it could mean we are having a bug in our older Linux compatibility code.

You also mention this test using dual core - does it also crash with a single core ? If the latter, it would further indicate that the Linux 4.1 crash is different from the 3.16 crash.

#32 - 09/02/2015 09:15 AM - Bjoern Franke

Marek Lindner wrote:

> Linus Lüssing wrote:
>
> > The last call trace looks different. Note that the module actually crashing is ip_gre ("gre_tap_xmit+0xc5/0x110 [ip_gre]").
>
> I agree, this crash looks quite different. @Bjoern: Do you get these batman-adv fragmentation backtraces with 4.1 at all or do they all look like what you posted above ? If so, it could mean we are having a bug in our older Linux compatibility code.

They seem to be no fragmentation backtraces, but all also not ip_gre related:

```
[ 6126.681689] BUG: unable to handle kernel paging request at 00000000ffffffff
[ 6126.681839] IP: [<ffffffff811afceb>] free_block+0x11b/0x1a0
[ 6126.681948] PGD d2d94067 PUD 0
[ 6126.682013] Oops: 0002 [#1] SMP
[ 6126.682083] Modules linked in: ipt_MASQUERADE nf_nat_masquerade_ipv4 iptable_nat nf_conntrack_ipv4 nf_defra
g_ipv4 nf_nat_ipv4 nf_nat nf_conntrack ip_tables x_tables tun bridge stp llc batman_adv(O) crc32c_generic libc
rc32c ip_gre ip_tunnel gre radeon amd64_edac_mod evdev tpm_infineon tpm_tis kvm_amd ttm drm_kms_helper drm eda
c_mce_amd k10temp tpm kvm sp5100_tco edac_core pcspkr i2c_algo_bit i2c_piix4 shpchp button acpi_cpufreq proces
sor thermal_sys autofs4 ext4 crc16 mbcache jbd2 sg sd_mod ata_generic ohci_pci tg3 ptp pps_core libphy pata_at
iixp ahci libahci libata ohci_hcd ehci_pci ehci_hcd scsi_mod usbcore usb_common
[ 6126.683368] CPU: 1 PID: 3540 Comm: kworker/1:1 Tainted: G           O    4.1.0-0.bpo.1-amd64 #1 Debian 4.1.
3-1~bpo8+1
[ 6126.683546] Hardware name: HP ProLiant MicroServer, BIOS O41     07/29/2011
[ 6126.683669] Workqueue: events cache_reap
[ 6126.683743] task: ffff88011a18ac20 ti: ffff88011a740000 task.ti: ffff88011a740000
[ 6126.683869] RIP: 0010:[<ffffffff811afceb>]  [<ffffffff811afceb>] free_block+0x11b/0x1a0
[ 6126.684012] RSP: 0018:ffff88011a743cf8  EFLAGS: 00010016
[ 6126.684104] RAX: 0000000001a40d81 RBX: ffff8800d7abbd00 RCX: 00000000fffffffe
[ 6126.684224] RDX: ffffea0003481ae0 RSI: 0000000000000001 RDI: 00000000ffffffff
[ 6126.684344] RBP: ffffe8ffffc89fa8 R08: ffffea0003481b20 R09: 0000000000000010
[ 6126.684464] R10: ffff8800d2021040 R11: ffff8800d2021080 R12: ffffea0003481b20
[ 6126.684584] R13: ffff8800d206c080 R14: ffffea0003481b00 R15: ffff8800d7ab9840
[ 6126.684705] FS:  00007fcc8bdb2700(0000) GS:ffff88011fc80000(0000) knlGS:0000000000000000
[ 6126.684842] CS:  0010 DS: 0000 ES: 0000 CR0: 000000008005003b
[ 6126.684940] CR2: 00000000ffffffff CR3: 00000000d2d88000 CR4: 00000000000007e0
[ 6126.685059] Stack:
[ 6126.685097]  ffffe8ffffc89fe0 ffff8800d7ab9868 ffff88011a743d68 ffff8800d7ab9848
[ 6126.685242]  ffff88011fc93710 ffffe8ffffc89f90 ffff8800d7ab9840 0000000000000008
[ 6126.685386]  0000000000000000 ffff8800d7abbd00 ffffe8ffffc89fa0 ffffffff811b0c47
[ 6126.685529] Call Trace:
[ 6126.685579]  [<ffffffff811b0c47>] ? drain_array+0xa7/0x140
[ 6126.685677]  [<ffffffff811b1dec>] ? cache_reap+0x6c/0x260
[ 6126.685773]  [<ffffffff8108b46b>] ? process_one_work+0x14b/0x430
[ 6126.685878]  [<ffffffff8108bfcb>] ? worker_thread+0x6b/0x560
[ 6126.685978]  [<ffffffff8108bf60>] ? rescuer_thread+0x390/0x390
[ 6126.686081]  [<ffffffff81091423>] ? kthread+0xd3/0xf0
[ 6126.686171]  [<ffffffff81091350>] ? kthread_create_on_node+0x180/0x180
[ 6126.686287]  [<ffffffff815797f2>] ? ret_from_fork+0x42/0x70
[ 6126.686384]  [<ffffffff81091350>] ? kthread_create_on_node+0x180/0x180
[ 6126.686495] Code: 4b 18 41 8b 7e 18 89 c6 48 0f af f1 0f b6 4b 1c 48 c1 ee 20 29 f0 d3 e8 0f b6 4b 1d 01 f0
 49 8b 76 10 d3 e8 8d 4f ff 41 89 4e 18 <88> 04 0e 49 8b 47 38 48 83 c0 01 49 89 47 38 41 8b 56 18 85 d2
[ 6126.687162] RIP  [<ffffffff811afceb>] free_block+0x11b/0x1a0
[ 6126.687269]  RSP <ffff88011a743cf8>
[ 6126.687329] CR2: 00000000ffffffff
```

> You also mention this test using dual core - does it also crash with a single core ? If the latter, it would further indicate that the Linux 4.1 crash is different from the 3.16 crash.

No, turned off the second core on August 30th and no crashes so far.

#33 - 10/19/2015 08:17 PM - Lars B

Hi,

we are seeing an similar issue on our gateways after activating a second cpu.

Setup:
Kernel 4.1.0-0.bpo.2-amd64
VMWare Machine with 2 vCPU and 2GB Memory
6 batman instances with 3 interfaces each, 1 batman instance with one interface

```
[593442.752609] Skipping free of non-empty nc_path (c6:72:1f:c7:11:a8 -> 32:b9:c2:b0:63:2c)!
[593498.917130] Skipping free of non-empty nc_path (04:af:fe:42:02:06 -> 32:b9:c2:d9:24:84)!
[593502.778118] Skipping free of non-empty nc_path (fa:1e:67:2f:02:76 -> 04:af:fe:02:02:06)!
[593502.778337] Skipping free of non-empty nc_path (c6:72:1f:fe:8b:bc -> 04:af:fe:22:02:06)!
[593555.599770] ------------[ cut here ]------------
[593555.599782] WARNING: CPU: 1 PID: 2240 at /build/linux-PoJsUp/linux-4.1.6/lib/list_debug.c:53 list_del+0x9/
0x30()
[593555.599786] list_del corruption, ffff88007a479550->next is LIST_POISON1 (dead000000100100)
[593555.599868] Modules linked in: tun bridge stp llc ip_gre ip_tunnel gre xt_nat iptable_nat nf_conntrack_ipv
4 nf_defrag_ipv4 nf_nat_ipv4 nf_nat nf_conntrack xt_TCPMSS xt_tcpudp iptable_mangle iptable_filter ip_tables x
_tables crct10dif_pclmul crc32_pclmul ghash_clmulni_intel aesni_intel aes_x86_64 lrw gf128mul ppdev evdev psmo
use glue_helper vmwgfx vmw_balloon serio_raw ablk_helper cryptd ttm battery pcspkr drm_kms_helper drm acpi_cpu
freq processor parport_pc ac parport thermal_sys i2c_piix4 shpchp vmw_vmci 8250_fintek button batman_adv libcr
c32c autofs4 ext4 crc16 mbcache jbd2 sr_mod cdrom sg ata_generic sd_mod crc32c_intel vmxnet3 floppy ata_piix m
ptspi scsi_transport_spi mptscsih libata mptbase scsi_mod
[593555.600251] CPU: 1 PID: 2240 Comm: fastd Tainted: G        W       4.1.0-0.bpo.2-amd64 #1 Debian 4.1.6-1~b
po8+1
[593555.600254] Hardware name: VMware, Inc. VMware Virtual Platform/440BX Desktop Reference Platform, BIOS 6.0
0 04/14/2014
[593555.600257]  0000000000000000 ffffffff81752c08 ffffffff815732fe ffff88007fd03b68
[593555.600292]  ffffffff81072ed1 ffff88007a479550 ffff88007a479540 ffff88006fab02b2
[593555.600297]  ffff88007b73f840 ffff88007ade1340 ffffffff81072f4a ffffffff81752cc0
[593555.600302] Call Trace:
[593555.600305]  <IRQ>  [<ffffffff815732fe>] ? dump_stack+0x40/0x50
[593555.600315]  [<ffffffff81072ed1>] ? warn_slowpath_common+0x81/0xb0
[593555.600324]  [<ffffffff81072f4a>] ? warn_slowpath_fmt+0x4a/0x50
[593555.600332]  [<ffffffff812fa519>] ? list_del+0x9/0x30
[593555.600346]  [<ffffffffa01c69ca>] ? batadv_tt_tvlv_ogm_handler_v1+0x29a/0x450 [batman_adv]
[593555.600356]  [<ffffffffa01ba302>] ? batadv_tvlv_containers_process+0xf2/0x190 [batman_adv]
[593555.600367]  [<ffffffffa01ba3dd>] ? batadv_tvlv_ogm_receive+0x3d/0x50 [batman_adv]
[593555.600388]  [<ffffffffa01b11ce>] ? batadv_iv_ogm_process_per_outif+0xb6e/0xe60 [batman_adv]
[593555.600393]  [<ffffffff814654c3>] ? sock_wfree+0x63/0x70
[593555.600401]  [<ffffffffa01b17a8>] ? batadv_iv_ogm_receive+0x2e8/0x3c0 [batman_adv]
[593555.600408]  [<ffffffffa01b17e0>] ? batadv_iv_ogm_receive+0x320/0x3c0 [batman_adv]
[593555.600412]  [<ffffffff8101d285>] ? read_tsc+0x5/0x10
[593555.600422]  [<ffffffffa01b9cd3>] ? batadv_batman_skb_recv+0xd3/0x120 [batman_adv]
[593555.600426]  [<ffffffff8147b90f>] ? netif_receive_skb_internal+0x1f/0x90
[593555.600431]  [<ffffffff8147b6d5>] ? __netif_receive_skb_core+0x7e5/0x990
[593555.600435]  [<ffffffff8147c612>] ? process_backlog+0xa2/0x140
[593555.600439]  [<ffffffff8147be22>] ? net_rx_action+0x212/0x340
[593555.600443]  [<ffffffff810772fa>] ? __do_softirq+0x11a/0x290
[593555.600447]  [<ffffffff8157adfc>] ? do_softirq_own_stack+0x1c/0x30
[593555.600449]  <EOI>  [<ffffffff81076bb5>] ? do_softirq.part.20+0x35/0x40
[593555.600455]  [<ffffffff8147a61f>] ? netif_rx_ni+0x2f/0x70
[593555.600466]  [<ffffffffa040e570>] ? tun_get_user+0x520/0x9e0 [tun]
[593555.600471]  [<ffffffffa040eaf1>] ? tun_chr_write_iter+0x51/0x80 [tun]
[593555.600475]  [<ffffffff811cadc8>] ? new_sync_write+0x88/0xb0
[593555.600479]  [<ffffffff811cb464>] ? vfs_write+0xa4/0x1b0
[593555.600483]  [<ffffffff8146209e>] ? __sys_recvmsg+0x3e/0x80
[593555.600504]  [<ffffffff811cc1d2>] ? SyS_write+0x42/0xb0
[593555.600509]  [<ffffffff815792b2>] ? system_call_fast_compare_end+0xc/0x6b
[593555.600512] ---[ end trace caff6b6f0c3d917c ]---
[593618.429309] Skipping free of non-empty nc_path (04:af:fe:52:02:06 -> c6:72:1f:c7:11:a8)!
[593618.429321] Skipping free of non-empty nc_path (04:af:fe:52:02:06 -> a2:f7:c1:5e:ca:9e)!
[593621.966514] BUG: unable to handle kernel paging request at ffff88017b436f80
[593621.966593] IP: [<ffffffff81469f04>] __alloc_skb+0x134/0x1f0
[593621.966636] PGD 1b24067 PUD 0
[593621.966661] Oops: 0002 [#1] SMP
[593621.966694] Modules linked in: tun bridge stp llc ip_gre ip_tunnel gre xt_nat iptable_nat nf_conntrack_ipv
4 nf_defrag_ipv4 nf_nat_ipv4 nf_nat nf_conntrack xt_TCPMSS xt_tcpudp iptable_mangle iptable_filter ip_tables x
_tables crct10dif_pclmul crc32_pclmul ghash_clmulni_intel aesni_intel aes_x86_64 lrw gf128mul ppdev evdev psmo
use glue_helper vmwgfx vmw_balloon serio_raw ablk_helper cryptd ttm battery pcspkr drm_kms_helper drm acpi_cpu
freq processor parport_pc ac parport thermal_sys i2c_piix4 shpchp vmw_vmci 8250_fintek button batman_adv libcr
c32c autofs4 ext4 crc16 mbcache jbd2 sr_mod cdrom sg ata_generic sd_mod crc32c_intel vmxnet3 floppy ata_piix m
```

```
ptspi scsi_transport_spi mptscsih libata mptbase scsi_mod
[593621.972475] CPU: 1 PID: 2187 Comm: fastd Tainted: G        W       4.1.0-0.bpo.2-amd64 #1 Debian 4.1.6-1~b
po8+1
[593621.972634] Hardware name: VMware, Inc. VMware Virtual Platform/440BX Desktop Reference Platform, BIOS 6.0
0 04/14/2014
[593621.972799] task: ffff8800369c55f0 ti: ffff880079fb8000 task.ti: ffff880079fb8000
[593621.972898] RIP: 0010:[<ffffffff81469f04>]  [<ffffffff81469f04>] __alloc_skb+0x134/0x1f0
[593621.973014] RSP: 0018:ffff880079fbb968  EFLAGS: 00010246
[593621.973091] RAX: 00000000ffffffff RBX: ffff88007b6e5a00 RCX: 00000000ffffffff
[593621.973168] RDX: ffff88017b436f80 RSI: 00000000ffffff80 RDI: ffff88007b6e5ac8
[593621.973206] RBP: ffff88007b437000 R08: 00000000000001c0 R09: 000000000001ca20
[593621.973244] R10: 0000000000000014 R11: 000000000762f388 R12: 0000000000000000
[593621.973291] R13: ffff88007f800400 R14: 00000000000004d0 R15: 00000000ffffffff
[593621.973331] FS:  00007f133d6c6700(0000) GS:ffff88007fd00000(0000) knlGS:0000000000000000
[593621.981781] CS:  0010 DS: 0000 ES: 0000 CR0: 0000000080050033
[593621.981813] CR2: ffff88017b436f80 CR3: 000000007cbf9000 CR4: 00000000000406e0
[593621.981979] Stack:
[593621.981994]  0000031e00000000 00ff880036e2c000 ffff88006e44b700 0000000000000000
[593621.982034]  0000000000000000 ffff8800369c55f0 ffff8800369c55f0 ffff88007c92c7c0
[593621.982073]  0000000000000000 ffffffff8146ae86 ffff880036e2c000 ffff88006e44b700
[593621.982178] Call Trace:
[593621.982200]  [<ffffffff8146ae86>] ? alloc_skb_with_frags+0x56/0x1f0
[593621.982231]  [<ffffffff814656b1>] ? sock_alloc_send_pskb+0x1e1/0x260
[593621.984035]  [<ffffffff814bb515>] ? __ip_append_data.isra.43+0x775/0xb00
[593621.987846]  [<ffffffff814b93d0>] ? ip_reply_glue_bits+0x60/0x60
[593621.989439]  [<ffffffff814bcf5a>] ? ip_make_skb+0xda/0x130
[593621.990350]  [<ffffffff814b93d0>] ? ip_reply_glue_bits+0x60/0x60
[593621.991591]  [<ffffffff814e464e>] ? udp_sendmsg+0x29e/0x950
[593621.992949]  [<ffffffff811cc3ec>] ? rw_copy_check_uvector+0x6c/0x110
[593622.004491]  [<ffffffff81460b5c>] ? sock_sendmsg+0x3c/0x50
[593622.005452]  [<ffffffff814616ab>] ? ___sys_sendmsg+0x27b/0x290
[593622.006361]  [<ffffffffa040db57>] ? tun_do_read+0x267/0x440 [tun]
[593622.008010]  [<ffffffff8120fccb>] ? ep_scan_ready_list+0x1cb/0x1e0
[593622.010634]  [<ffffffffa040de50>] ? tun_chr_read_iter+0x50/0x90 [tun]
[593622.011778]  [<ffffffff81461e2e>] ? __sys_sendmsg+0x3e/0x80
[593622.012886]  [<ffffffff815792b2>] ? system_call_fast_compare_end+0xc/0x6b
[593622.013809] Code: e1 01 83 e0 f7 c1 e1 03 09 c1 b8 ff ff ff ff 45 85 e4 88 8b 90 00 00 00 b9 ff ff ff ff 6
6 89 83 c6 00 00 00 66 89 8b c2 00 00 00 <48> c7 02 00 00 00 00 48 c7 42 08 00 00 00 00 48 c7 42 10 00 00
[593622.019904] RIP  [<ffffffff81469f04>] __alloc_skb+0x134/0x1f0
[593622.021879]  RSP <ffff880079fbb968>
[593622.025800] CR2: ffff88017b436f80
```

A kernel dump is also available.
The same setup runs fine on a single core machine, so I don't believe in a multiple interfaces issue (28> days uptime).

Is there any way, we can assist to find the issue?

Kind Regards,
Lars

#34 - 10/27/2015 12:04 PM - Simon Wunderlich

Hi Lars,

sorry for the late reply - could you please try to turn off network coding? There is a directive CONFIG_BATMAN_ADV_NC=n in the makefile. We consider network coding experimental, and it seems it causes some problem for you ....

Thank you very much!

#35 - 10/30/2015 10:45 PM - Lars B

Hi Simon,

no problem :-)
I checked our Makefile and CONFIG_BATMAN_ADV_NC=n is set?

Kind Regards,
Lars

#36 - 11/02/2015 05:51 PM - Simon Wunderlich

I believe there must be a mistake or your config doesn't match whatever has been compiled. Error messages like "Skipping free of non-empty nc_path (c6:72:1f:c7:11:a8 -> 32:b9:c2:b0:63:2c)!" can only appear when network coding is enabled.

You could inspect your binary e.g. with Strings and see if the "Skipping free of non-empty nc_path" is found. If it is, then you compiled with network coding enabled.

#37 - 11/23/2015 12:58 PM - Ben Oswald

We now have the same problem with crashes of this type. Maybe it is connected to my other issue #225but the stacktrace looks different there. Both problems started with adding a second interface to our bat device.

#38 - 11/23/2015 01:31 PM - Simon Wunderlich

Ben,

please post your stack traces, batman-adv version, kernel version, and anything else you think could be useful for debugging.

Thanks!

#39 - 11/23/2015 01:53 PM - Martin Weinelt

We also experience problems when introducing more than one hard link into a batman-adv interface.

We've tried this with Debian Jessies (amd64) and experienced the following stack traces: https://gist.github.com/andir/f9b834462e0dad9a99e1

```
Kernel 3.16.0-4-amd64 #1 Debian 3.16.7-ckt11-1+deb8u5
batman-adv 2015.1-10-gb307e72-dirty (maint branch)
```

Then upgraded to Kernel 4.2, experienced the same behaviour with the intree module.

```
# modinfo batman-adv
filename:       /lib/modules/4.2.0-0.bpo.1-amd64/kernel/net/batman-adv/batman-adv.ko
version:        2015.1
description:    B.A.T.M.A.N. advanced
author:         Marek Lindner <mareklindner@neomailbox.ch>, Simon Wunderlich <sw@simonwunderlich.de>
license:        GPL
srcversion:     C81BBE85714B26E3EC3F9A5
depends:        libcrc32c,crc16
intree:         Y
vermagic:       4.2.0-0.bpo.1-amd64 SMP mod_unload modversions
```

#40 - 12/05/2015 04:50 PM - Ben Oswald

- File Screenshot2015-11-13_10-11-51.png added

- File Screenshot2015-11-22_19-58-47.png added

Sorry that I've forgotten these information. The Kernel Version is 3.16.0-4-amd64 form the current debian stable with the following batman-adv module:

```
filename:       /lib/modules/3.16.0-4-amd64/kernel/net/batman-adv/batman-adv.ko
version:        2014.3.0
description:    B.A.T.M.A.N. advanced
author:         Marek Lindner <mareklindner@neomailbox.ch>, Simon Wunderlich <sw@simonwunderlich.de>
license:        GPL
srcversion:     39F613C8D996EE89ACE5144
depends:        libcrc32c,crc16
intree:         Y
vermagic:       3.16.0-4-amd64 SMP mod_unload modversions
```

Some screenshots of the stacktrace are attached.

My personal observations show that the probability of a crash increases with number of direct peers of the server. With 35 peers it seems stable for a few days now, with around 100 peers it crashes every few hours or minutes.

#41 - 12/06/2015 11:21 AM - Ben Oswald

Ok, I'll have to correct my post from yesterday it crashes also with 34 peers.

#42 - 12/10/2015 09:37 PM - Lars B

Hi,

i recompiled with CONFIG_BATMAN_ADV_NC=n and the stability increased.
Now uptimes from 4-7 days are possible.

Setup:
Kernel 4.1.0-0.bpo.2-amd64 (4.2.5-1~bpo8+1)
VMWare Machine with 2 vCPU and 1GB Memory
batman 2015.1

Latest batman related trace:

```
[478289.937119] general protection fault: 0000 [#1] SMP
[478289.937176] Modules linked in: tun ip_gre ip_tunnel gre bridge stp llc coretemp crct10dif_pclmul crc32_pcl
mul sha256_ssse3 sha256_generic hmac drbg ansi_cprng aesni_intel aes_x86_64 lrw gf128mul glue_helper ppdev par
port_pc evdev ablk_helper vmw_balloon cryptd psmouse parport serio_raw pcspkr vmwgfx ttm drm_kms_helper i2c_pi
ix4 battery 8250_fintek drm vmw_vmci shpchp acpi_cpufreq processor thermal_sys ac button batman_adv(O) libcrc3
2c autofs4 ext4 crc16 mbcache jbd2 sr_mod cdrom sg ata_generic sd_mod crc32c_intel vmxnet3 floppy ata_piix vmw
_pvscsi libata scsi_mod
[478289.937417] CPU: 0 PID: 2253 Comm: batadv-vis Tainted: G           O    4.2.0-0.bpo.1-amd64 #1 Debian 4.2.
5-1~bpo8+1
[478289.937484] Hardware name: VMware, Inc. VMware Virtual Platform/440BX Desktop Reference Platform, BIOS 6.0
0 04/14/2014
[478289.937545] task: ffff88007c33efc0 ti: ffff88007ba44000 task.ti: ffff88007ba44000
[478289.937568] RIP: 0010:[<ffffffffa01ea74e>]  [<ffffffffa01ea74e>] batadv_orig_router_get+0xe/0x60 [batman_a
dv]
[478289.937617] RSP: 0018:ffff88007ba47db0  EFLAGS: 00010206
[478289.937634] RAX: 3074616228203630 RBX: ffff88007bf6b000 RCX: 0000000000004001
[478289.937658] RDX: 0000000000004000 RSI: 0000000000000000 RDI: ffff88007bf6b000
[478289.937679] RBP: ffff88007b9428c0 R08: 0000000000004000 R09: 00000000fffffffd
[478289.937701] R10: 0000000000000000 R11: 0000000000000036 R12: 0000000000000000
[478289.937722] R13: ffff88007c16ef40 R14: ffff880000034dc0 R15: 0000000000000000
[478289.937744] FS:  00007f1707b57700(0000) GS:ffff88007fc00000(0000) knlGS:0000000000000000
[478289.937768] CS:  0010 DS: 0000 ES: 0000 CR0: 0000000080050033
[478289.937786] CR2: 00007f5349509000 CR3: 000000007b87e000 CR4: 00000000000006f0
[478289.937848] Stack:
[478289.937858]  ffffffffa01de413 ffff88007b9428d8 ffff88007a65c000 ffffffffa01f69ff
[478289.937884]  ffffffff811e2203 ffff880000000001 ffff8800798a1100 ffff88007c97c300
[478289.937909]  0000007100000322 ffffffff811e18b3 ffff880036cbd000 ffff880079405a00
[478289.937934] Call Trace:
[478289.937951]  [<ffffffffa01de413>] ? batadv_iv_ogm_orig_print+0xe3/0x240 [batman_adv]
[478289.937978]  [<ffffffff811e2203>] ? seq_printf+0x43/0x50
[478289.937995]  [<ffffffff811e18b3>] ? seq_buf_alloc+0x13/0x30
[478289.938018]  [<ffffffffa01eb46d>] ? batadv_orig_seq_print_text+0x7d/0xb0 [batman_adv]
[478289.938043]  [<ffffffff811e1c4b>] ? seq_read+0xcb/0x380
[478289.938062]  [<ffffffff811bfac1>] ? vfs_read+0x81/0x120
[478289.938682]  [<ffffffff811c0822>] ? SyS_read+0x42/0xa0
[478289.939309]  [<ffffffff815586f2>] ? system_call_fast_compare_end+0xc/0x6b
[478289.939921] Code: 50 01 44 89 c0 f0 0f b1 11 44 39 c0 74 c2 eb e6 45 31 ed eb d3 0f 1f 84 00 00 00 00 00 6
6 66 66 66 90 48 8b 47 08 48 85 c0 74 0e <48> 39 70 10 74 0e 48 8b 00 48 85 c0 75 f2 31 d2 48 89 d0 c3 48
[478289.941829] RIP  [<ffffffffa01ea74e>] batadv_orig_router_get+0xe/0x60 [batman_adv]
[478289.942441]  RSP <ffff88007ba47db0>
```

#43 - 12/16/2015 10:12 AM - Sven Eckelmann

Lars B. everyone is in the dark and everyone is just guessing around. But your crash looks a little bit like something which (hopefully) is addressed by the patchset I've posted yesterday:

- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-1-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-2-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-3-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-4-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-5-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-6-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-7-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-8-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-9-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-10-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-11-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-12-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-13-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-14-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-15-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-16-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-17-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-18-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-19-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-20-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-21-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-22-git-send-email-sven@narfation.org/
- https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1450222316-1764-23-git-send-email-sven@narfation.org/

#44 - 12/20/2015 02:24 PM - Lars B

Hi Sven, thanks for your work!
I upgrade to latest git master, applied the patches and will provide feedback in the next days.

#45 - 12/22/2015 05:13 AM - Martin Weinelt

v2015.2-64-g5fe476a (without patchset)
https://gist.github.com/mweinelt/feb83cf938ed9cc96d19

v2015.2-94-g29a2684 (including patchset)
https://gist.github.com/mweinelt/e3e4946d178813235260

#46 - 01/29/2016 10:49 AM - Marek Lindner

- Subject changed from Kernel Crash when using more than on interface in bat0 to Kernel Crash when using more than one interface in bat0

#47 - 05/02/2016 09:31 PM - Sven Eckelmann

Did anyone triy v2016.1?

#48 - 05/29/2016 10:44 PM - Sven Eckelmann

Maybe waiting for v2016.2 is also not a bad idea. At least I hope to get following patch (or a variant of it) merged for this release:
https://patchwork.open-mesh.org/project/b.a.t.m.a.n./patch/1464588694-19855-1-git-send-email-sven@narfation.org/

It tackles a weird memory corruption problem. A memory corruption problem like the ones you may have here.

#49 - 05/31/2016 09:40 AM - Sven Eckelmann

- Assignee set to Marek Lindner

- Status changed from New to In Progress

Martin is testing it for Freifunk Darmstadt. It looks good at the moment and he already sent his "Tested-by".

Marek Lindner: Maybe you can add the patch to maint.

#50 - 05/31/2016 09:41 AM - Sven Eckelmann

- Related to Bug #217: Oops: "Unable to handle kernel paging request" in batadv_tt_local_remove added

#51 - 05/31/2016 09:41 AM - Sven Eckelmann

- Related to Bug #228: Workqueue: bat_events batadv_send_outstanding_bat_ogm_packet added

#52 - 06/18/2016 11:42 AM - Sven Eckelmann

- Assignee deleted (Marek Lindner)

- Status changed from In Progress to Feedback

batman-adv 2016.2 was released last week. I suspect that this release fixes this problem. At least I have reports from Freifunk Darmstadt and Freifunk Chemnitz that an included patch solved a similar problem for them.

This ticket doesn't seem to show a lot activity anymore and thus I would like to close it soon to avoid a dead but still open ticket without a chance to mark it as fixed. I will wait until mid of July for feedback but will close this ticket if nothing happens.

#53 - 06/25/2016 05:43 PM - Lars B

Confirmed, issue is solved with 2016.2 for me :-)

Thanks!

#54 - 06/25/2016 06:04 PM - Sven Eckelmann

- % Done changed from 0 to 100

- Status changed from Feedback to Resolved

Thanks a lot for the feedback :)

#55 - 07/16/2016 11:14 PM - Sven Eckelmann

- Status changed from Resolved to Closed

#56 - 02/11/2017 08:44 AM - Sven Eckelmann

- Target version set to 2016.2

#57 - 05/27/2020 10:27 PM - Sven Eckelmann

- Description updated

Files

0001-batman-adv-TEST-if-batadv_frag_clear_chain-is-valid.patch	964 Bytes	08/20/2015	Sven Eckelmann
networksetup 2015.8.png	23.4 KB	08/27/2015	Bjoern Franke
Screenshot2015-11-22_19-58-47.png	18.6 KB	12/05/2015	Ben Oswald
Screenshot2015-11-13_10-11-51.png	70 KB	12/05/2015	Ben Oswald