Bug 207225 - Malformed headroom in umem request of XDP socket could lead to out of bound write

Status: RESOLVED PATCH_ALREADY_AVAILABLE
Alias: None
Product: Networking
Component: Other (show other bugs)
Hardware: All Linux
Importance: P1 low
Assignee: Stephen Hemminger
URL:
Keywords:
Depends on:
Blocks:

Reported: 2020-04-13 14:27 UTC by Bùi Quang Minh
Modified: 2020-04-17 03:53 UTC (History)
CC List: 1 user (show)
See Also:
Kernel Version: 5.5.11, 5.5.17, 5.7-rc1
Tree: Mainline
Regression: No

Attachments	
POC registers malformed headroom in umem registration (2.59 KB, text/plain) 2020-04-13 14:27 UTC, Bùi Quang Minh	Details
Add an attachment (proposed patch, testcase, etc.)	

Bùi Quang Minh 2020-04-13 14:27:36 UTC Description

```
Created attachment 288417 [details]
POC registers malformed headroom in umem registration

- When user calls setsockopt to register umem ring on XDP socket, the headroom can
be a big unsigned 32 bit number, which leads to
   + This check in xdp_umem_reg function (net/xdp/xdp_umem.c) is bypassed
   size_chk = chunk_size - headroom - XDP_PACKET_HEADROOM;
   if (size_chk < 0)
        return -EINVAL;
   + This initialization in the same function, the chunk_size_nohr becomes larger
than actual size
   umem->chunk_size_nohr = chunk_size - headroom;

- Consequence: I see that the chunk_size_nohr is used to check that the xdp_buff
can fit into the chunk in xsk receive functions; with this malformed
chunk_size_nohr, we can put a larger than chunk size xdp_buff to chunk, leads to an
out of bound write. However, I research some more and find that to trigger to
receive functions, we must redirect the packets from XDP program using xskmap which
requires CAP_NET_ADMIN capability, which makes this very low impact.

- Unfortunately, I cannot trigger xsk receive functions (I am new to Linux kernel)
due to some error when binding XDP program to an interface. I can only prove the
register side, the initialization of chunk_size_nohr via debugging. I attached the
POC of malformed headroom umem register below, which I tested on kernel 5.5.11. The
POC needs to be run with root privilege (or a user with CAP_NET_RAW, this could be
achieve with new user namespace on kernel with CONFIG_USER_NS=y, however, as far as
I know, next phases when allocate xskmap, CAP_NET_ADMIN is required and user
namespace is not permitted).

Thank you very much for reviewing this report
```

Note
You need to log in before you can comment on or make changes to this bug.