



NetConf18 Topics

- **SCTP Offload**
- **UDP Tunnels ICMP Process**

Red Hat, Inc. Beijing, China

Xin Long

Linux Networking Conference in Boston, 2018



Outline

1 **SCTP Offload**

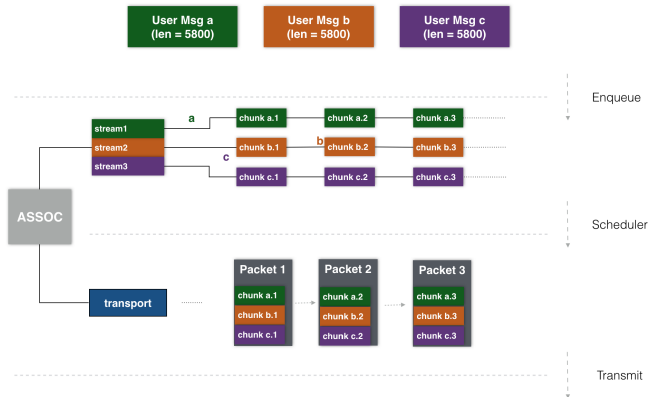
- SCTP GSO Basics
- SCTP GRO Basics
- SCTP GSO Over LOOPBACK and VETH
- SCTP GSO Over TUN and VIRTIO_NET
- SCTP Offload Over METAL DEV

2 **UDP Tunnels ICMP Process**

- ICMP FragNeeded and Redirect
- IP Tunnels ICMP Process
- UDP Tunnels ICMP Problem
- L2TP Tunnels ICMP Problem

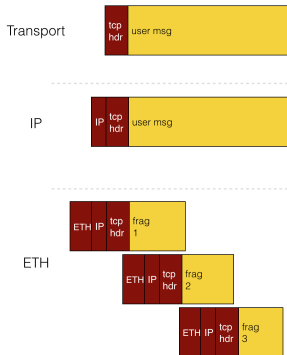
SCTP Chunks General Process

- Fragment Too Early?
- Fraglist, not Nr_frags?

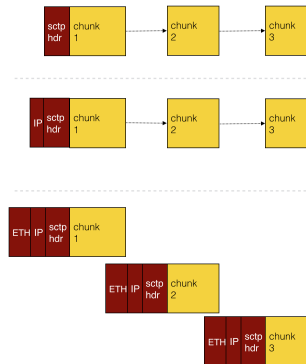


SCTP GSO Basics

TCP GSO



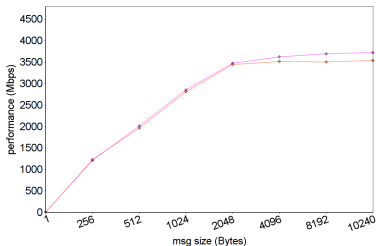
SCTP GSO



Performance Test

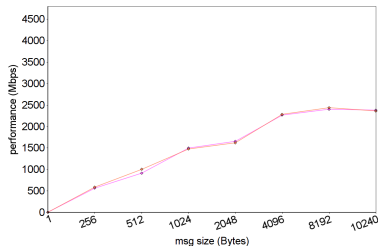
- slow cpu on send side.
- segmentation being done by the stack.

TCP General GSO Performance



non-gso gso

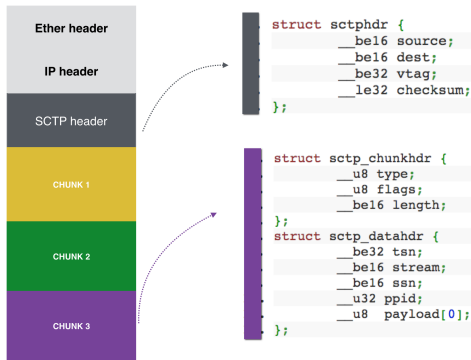
SCTP General GSO Performance



non-gso gso

SCTP Packet Format

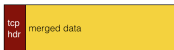
- When to Flush?
- Linearize?



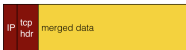
SCTP GRO Basics

TCP GRO

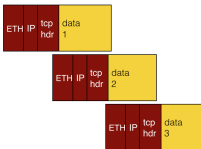
Transport



IP

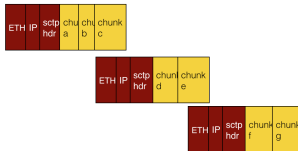


ETH



PUSH

SCTP GRO

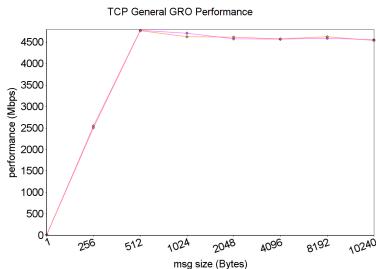


E

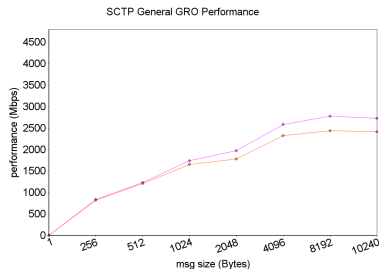


Performance Test

- slow cpu on recv side.
- segmentation being done by the stack.



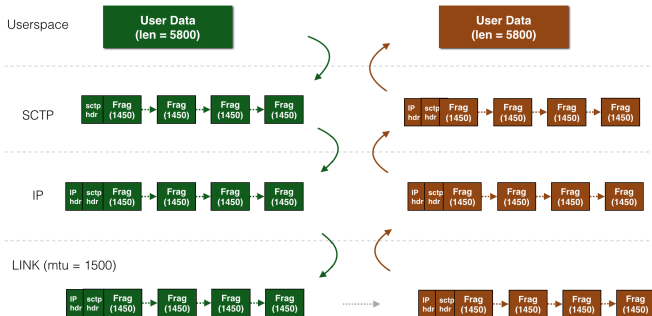
non-gro gro



non-gro gro

SCTP GSO Over LOOPBACK and VETH

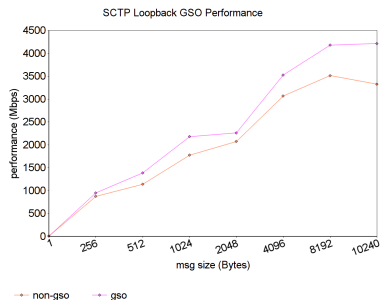
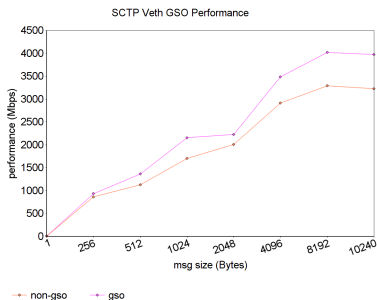
on sctp rx path:
+ support for fraglist process





Performance Test

- ip link set veth/lo mtu 1500.



SCTP GSO Over TUN and VIRTIO_NET

```
in xmit_skb():
- sg_init_table(sq->sg, skb_shinfo(skb)->nr_frags + (can_push ? 1 : 2));
+ sg_init_table(sq->sg, ARRAY_SIZE(sq->sg));
in virtnet_probe():
+ dev->hw_features |= NETIF_F_FRAGLIST;
```

Userspace

User Data
(len = 5800)

User Data
(len = 5800)

SCTP



IP



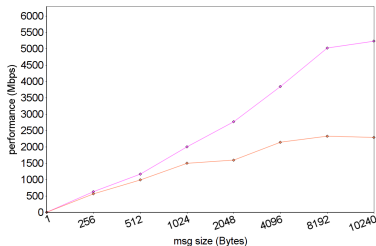
VIRTIO_NET



Performance Test

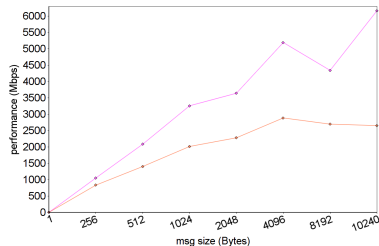
- host $\leftarrow$ guest (virtnet testing)
- host $\rightarrow$ guest (tap testing)

SCTP Virtio_net GSO Performance



non-gso gso

SCTP Tap GSO Performance

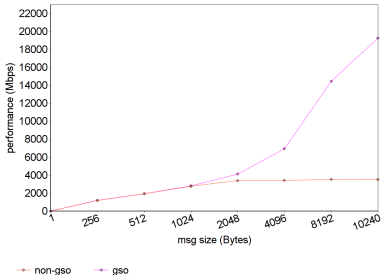


non-gso gso

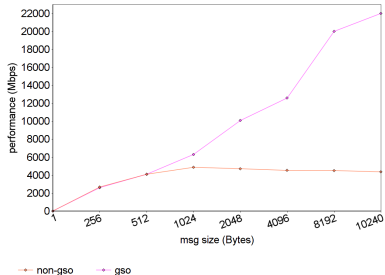
VS TCP GSO Over TUN and VIRTIO_NET

- Does Fragment Delay Matter?

TCP Virtio_net GSO Performance



TCP Tap GSO Performance



SCTP Offload Over METAL DEV

- Depends on Vendors, including CRC checksum offload.
- Will **tx-scatter-gather-fraglist** be supported?



Outline

1 SCTP Offload

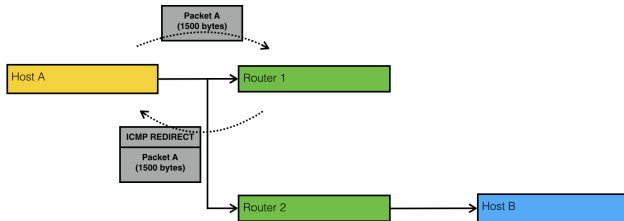
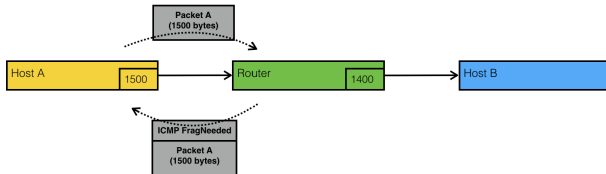
- SCTP GSO Basics
- SCTP GRO Basics
- SCTP GSO Over LOOPBACK and VETH
- SCTP GSO Over TUN and VIRTIO_NET
- SCTP Offload Over METAL DEV

2 UDP Tunnels ICMP Process

- ICMP FragNeeded and Redirect
- IP Tunnels ICMP Process
- UDP Tunnels ICMP Problem
- L2TP Tunnels ICMP Problem

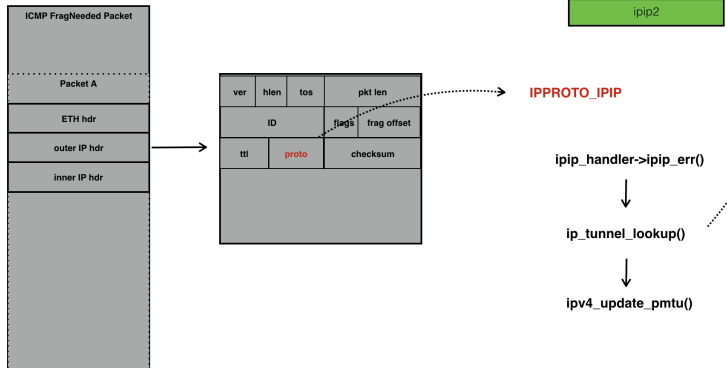
ICMP FragNeeded and Redirect

- Two important ICMP packets.



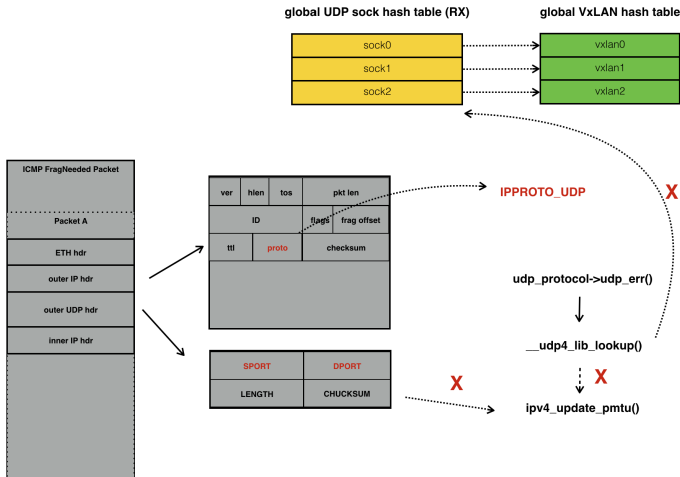
IP Tunnels ICMP Process

- ipv4: ipip, ipgre, vti, sit ...
- ipv6: ip(6)ip6, ip6gre, ip6vti ...
- others



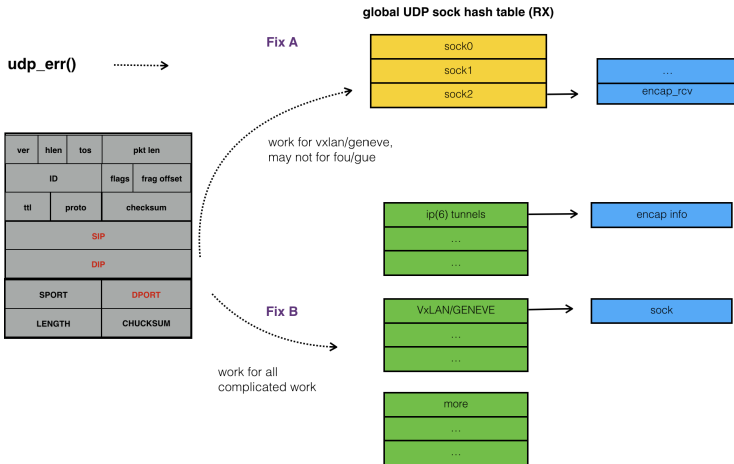
UDP Tunnels ICMP problem

- VxLAN/GENEVE
- FOU4/6, GUE4/6
- others



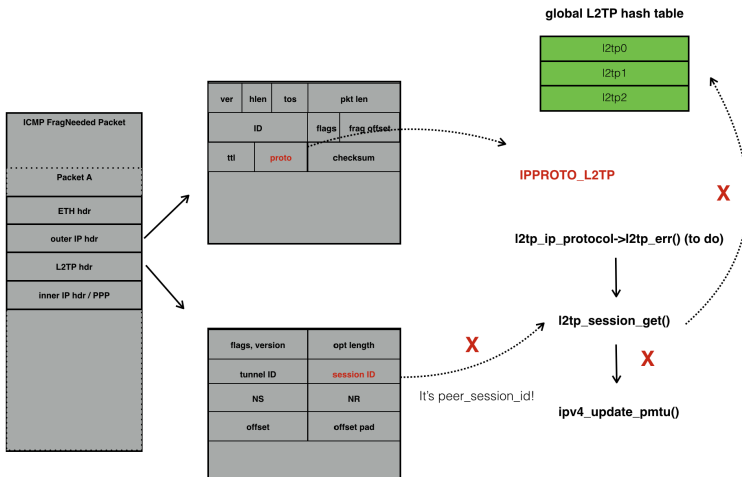
What We Can Do

- A: search UDP sock by SIP/DPORT/encap_rcv
- B: go to possible Tunnels and search by SIP/DIP/encap_info



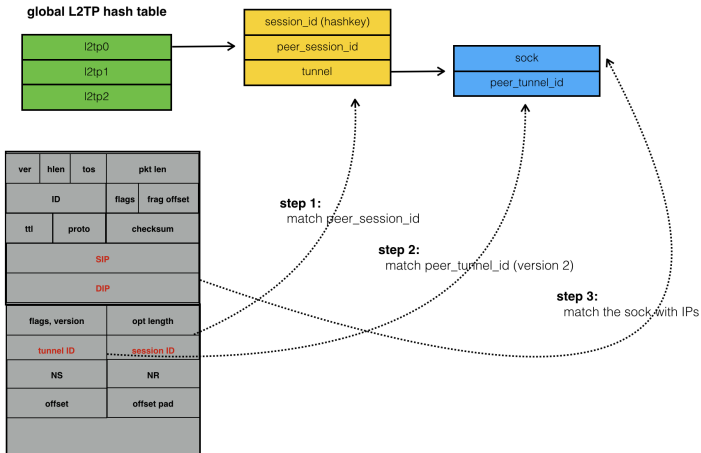
L2TP Tunnels ICMP Problem

- IP encap
- UDP encap



What We Can Do

- A: match L2TP with peer_session_id, peer_tunnel_id and socket by SIP/DIP



Ideas ?