Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

SHRI SHANKARLAL SUNDARBAI SHASUN JAIN COLLEGE FOR WOMEN
ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

DYNAMIC ENABLEMENT OF LSO (LARGESEND OFFLOAD) IN NETWORK VIRTUALIZED ENVIRONMENT FOR BETTER NETWORK THROUGHPUT

S.Annie Christilla
Associate Professor
Department of Computer Science, St. Francis De Sales College, Bangalore, Karnatka, India

ABSTRACT

LSO is a very important feature of a GigE/10G NICs providing a good amount of performance benefits. Using this feature Network layers can send bigger packets instead of smaller packets. On virtualized environmentNetwork Interface Card(NIC) will be attached to VIOS(Virutal IO servers) and logical partitions will be sharing this NIC through virtualization technique. In these kind of environment, LSO feature can be turned on/off on the NIC, the bridge layerSEA (Shared Ethernet Adapter) present in VIOS and in the virtual adapter present in logical partition. LSO need to be enabled in all these components to exploit this feature. If LSO is turned off on the bridge(SEA) present in the VIOS, will cause poor network performance. In this paper, method to achieve better network throughput when LSO feature is turned off/on dynamically is being proposed in this draft.

***keywords*:** LSO(Largesend Offload), VIOS(Virtual IO Server), SEA(Shared Ethernet Adapter), NIC (Network Interface Card), LPAR (Logical PARtition)

1. INTRODUCTION

Maximum Transmission Unit (MTU) of a network is the maximum protocol data unit that can be transferred on the physical medium. MTU is an inherent property of the physical media. For instance MTU in Ethernet is 1500 bytes. In a Network Protocol Stack, Network Layer or Internet Protocol (IP) layer implements datagram fragmentation so that packets with size larger than the network interface's MTU are fragmented to MTU size before being delivered to the data link layer. Transport protocols such as TCP negotiate MSS (Maximum Segment Size) during connection establishment which is the largest amount of data that TCP is willing to send in a single segment. To avoid IP fragmentation, MSS is always set lower than MTU. Large Send or TCP Segmentation Offload (TSO) is a feature supported by Network Adapters in which the job of fragmenting a larger packets into MTU size is done by the Network Interface Card (NIC) in hardware. network protocol stack can send larger size packets (without having to do the job of fragmentation in software) to NICs and the NIC hardware will do the fragmentation in hardware which would help in improving performance.LSO is a very important feature of a GigE/10G NICs providing a good amount of performance benefits. Using this feature Network layers can send bigger packets instead of smaller packets. However if the network applications have been written in a way to send smaller packets, this hardware feature cannot be exploited well. This is because if applications send smaller packets through the network stack, the stack will end up sending smaller packets to the NIC. With this there would be a lot of packets being sent down from the application to the NIC through the network protocol stack in kernel. This can be avoided by making applications send large size packets so that the lesser number of packets are being sent down by the protocol stack to the NIC

2. PROBLEM STATEMENT

On virtualized environment, LSO is turned on NIC, SEA and Virtual Adapter on the logical partition. When TCP connection is established from LPAR to host that resides outside the system, LPAR will be sending bigger packets upto 64k to the VIOS. The NIC on the VIOS will segment this packet based on MTU and send the data out. When the connection is established if LSO is turned off on SEA, LPAR will keep sending large packet with IF_DF flag on. As a result SEA will not be able to fragment this packet and send ICMP back to LPAR. However LPAR will still continue to send bigger packets. When the retransmission timer expires LPAR will send one packet with size of 1500 bytes. This will result poor network throughput as after every retransmission time out one packet will be sent to the remote host.

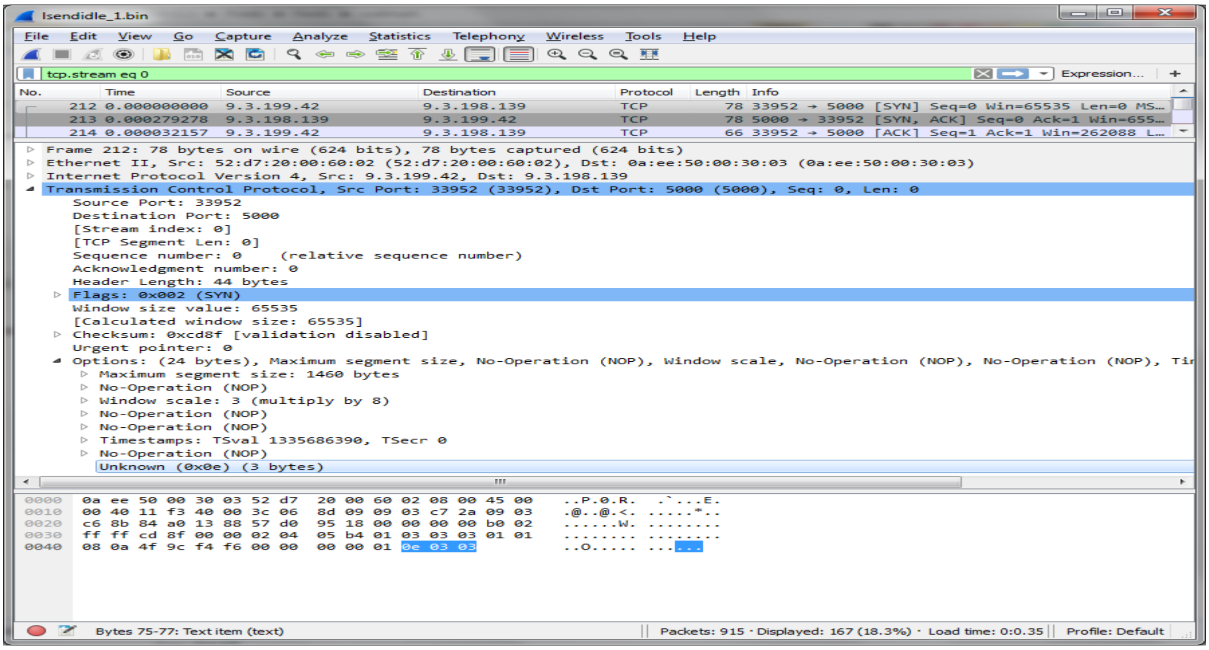
3. EXISTING METHOD

When TCP connection gets establishedLPAR will be sending TCP options0x0E0303 along with SYN packet. If SEA has LSO option turned on while it gets response (SYNACK) back from the destinationit will piggy pack the same TCP option. Upon receiving SYNACK and ifTCP option present, the Transport layer onLPAR will identify LSO is turned on SEA and turns on LSO flag for the connection.Post that transport layer will be sending larger packetsto the VIOS. NIC on VIOS will in turn fragmentthe packet and send it to the destination.TCP connection on the LPAR will keep sending largePacketsupto 64k. The checksum field of the packet will have the MSS value that adapter will use the packet to fragment. Later if LSO is turned off on SEA, and largePacket is received from LPAR(>1500 bytes), packets will be fragmented in the SEA layer and sent down to adapter as small packets. If packet also has IF_DF flag, then those packets will get dropped in SEA layer. In this case SEA willbe sending ICMP “fragmention needed” packetback to the LPAR. On receiving this packet LPAR will turn off LSO feature for the connection. Later if LSO is enabled on SEA, there is no wayTCP connections on the LPAR will be aware of and still will be sending smaller packets. This Results in poor network performance.

4. PROPOSED METHOD

In this paper, an algorithm to achieve better network throughput is proposed. On VIOS at SEA layer information (Source IP, Source Port, Destination IP, Destination Port, Largesend capable connection or not) about each connection will be maintained. When LSO is turned off, ICMP message with type 42 will be generated to Source IP, Source Port with information whether LSO is turned off or not. ICMP message will be generated for all the connections. On LPAR once it receives the ICMP message, it turns of the LSO feature in Transport layer for the specified connection. ICMP message will also have Destination IP and Destination Port and protocol no in the Data section. This will help the LPAR to find the correct connection and turn off the LSO feature. Later lets assume LSO is turned on on SEA then again the ICMP packets will get generated for the LPARS with the information LSO is turned on. Upon receiving this message LPARs will turn on the LSO for the particular TCP connection. At SEA layer, when it receives SYN Packet the entry about connection will be added. This entry will be removed upon receiving FIN for the same connection.

ICMP message used for communication between SEA and LPAR will be type 42 and code will have value either 0 or 1. 0 represents LSO is turned off on SEA and 1 represents LSO is turned on at the SEA layer. LSO at SEA can only be turn on if the LSO is enabled on the underlying network adapter. Data will have SourceIP, Source Port, Destination IP, Destination port and protocol no. Refer Figure 2 for the ICMP format.



```
tcp.stream eq 0

No.    Time         Source         Destination     Protocol  Length  Info
   212 0.000000000  9.3.199.42     9.3.198.139     TCP          78 33952 → 5000 [SYN] Seq=0 Win=65535 Len=0 MS...
   213 0.000279278  9.3.198.139    9.3.199.42      TCP          78 5000 → 33952 [SYN, ACK] Seq=0 Ack=1 Win=655...
   214 0.000032157  9.3.199.42     9.3.198.139     TCP          66 33952 → 5000 [ACK] Seq=1 Ack=1 Win=262088 L...

▷ Frame 212: 78 bytes on wire (624 bits), 78 bytes captured (624 bits)
▷ Ethernet II, Src: 52:d7:20:00:60:02 (52:d7:20:00:60:02), Dst: 0a:ee:50:00:30:03 (0a:ee:50:00:30:03)
▷ Internet Protocol Version 4, Src: 9.3.199.42, Dst: 9.3.198.139
◢ Transmission Control Protocol, Src Port: 33952 (33952), Dst Port: 5000 (5000), Seq: 0, Len: 0
    Source Port: 33952
    Destination Port: 5000
    [Stream index: 0]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 44 bytes
  ▷ Flags: 0x002 (SYN)
    Window size value: 65535
    [Calculated window size: 65535]
  ▷ Checksum: 0xcd8f [validation disabled]
    Urgent pointer: 0
  ◢ Options: (24 bytes), Maximum segment size, No-Operation (NOP), Window scale, No-Operation (NOP), No-Operation (NOP), Ti
    ▷ Maximum segment size: 1460 bytes
    ▷ No-Operation (NOP)
    ▷ Window scale: 3 (multiply by 8)
    ▷ No-Operation (NOP)
    ▷ No-Operation (NOP)
    ▷ Timestamps: TSval 1335686390, TSecr 0
    ▷ No-Operation (NOP)
      Unknown (0x0e) (3 bytes)

0000  0a ee 50 00 30 03 52 d7  20 00 60 02 08 00 45 00   ..P.0.R. .`...E.
0010  00 40 11 f3 40 00 3c 06  8d 09 09 03 c7 2a 09 03   .@..@.<. .....*..
0020  c6 8b 84 a0 13 88 57 d0  95 18 00 00 00 00 b0 02   ......W. ........
0030  ff ff cd 8f 00 00 02 04  05 b4 01 03 03 03 01 01   ........ ........
0040  08 0a 4f 9c f4 f6 00 00  00 00 01 0e 03 03         ..O..... ......

Bytes 75-77: Text item (text)    || Packets: 915 · Displayed: 167 (18.3%) · Load time: 0:0.35 || Profile: Default
```

Figure 1. TCP Option to negotiate LSO between LPAR and SEA

Algorithm:

1) During connection establishment phase both LPAR and SEA will exchange the TCP option 0x0E0303 if they are LSO capable.
2) If they are LSO capable, TCP layer on LPAR will send large amount of data (> 1500 bytes) and upto 64K
3) On VIOS entry will be added in the connection cache table with information SourceIP, source Port, Destination IP, Destination port, LSO capable and Protocol number
4) If LSO is turned on the virtual adapter on LPAR, call back will be called and it will search the TCP control block and will turn off LSO for the tcp connections.
5) If LSO is turned off on the SEA, SEA will go through control connection cache table and send ICMP for each entry. ICMP packet type will be 0x42 and code 0 as LSO is turned off. This packet will also have information Source IP, Source Port, Destination IP, Destination port and protocol number.
6) Upon receiving this ICMP packet, LPAR will disable the LARGESEND flag on TCP control block.
7) While sending data, TCP layer on LPAR will not send bigger packet if LARGESEND flag is not set. (As a result SEA need not fragment the packet).
8) Now lets consider a case when LSO is turned on at SEA, then again SEA layer will go through control connection cache table and send ICMP for each entry. ICMP packet type will be 0x42 and code will be 1 as LSO is enabled. This packet will also have information Source IP, Source Port, Destination IP, Destination port and protocol number.
9) Upon receiving this ICMP packet, LPAR will enable the LARGESEND flag on TCP control block.
10) While sending data, TCP layer on LPAR will send bigger packet if LARGESEND flag is set.

<table>
<tr><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td></tr>
<tr><td colspan="8">Type</td><td colspan="8">Code</td><td colspan="16">ICMP header checksum</td></tr>
<tr><td colspan="32">Data :::</td></tr>
</table>

Figure 2. ICMP Protocol format

5. CONCLUSION

The objective of the proposed method is to improve the network throughput by transferring the correct sized data to NIC on the VIOS. The proposed method updates the LSO capability of the SEA/underlying NIC to LPAR dynamically as and when it happens, thereby making sure LPAR sends the correct sized data to the NIC on VIOS. This eliminates the fragmentation effort on

VIOS and also when LSO is enabled uses the feature to full extent. Hence the network throughput is improved using this method.

REFERENCES

1. Retreived from https://tools.ietf.org/html/rfc792.
2. Hai Lin, Lucio Correia, & et. al., IBM PowerVM Virtualization
3. Gary R. Wright & W. Richard Stevens,” TCP/IP Illustrated-Implementation”, Vol. 2
4. Kumar Reddy, “Network Virtualization”.