Applications That Limit Network Performance

Joel Goergen Force10

Abstract: This presentation is a late request based on data the author received late last week covering applications that slow network performance. This data should be used by HSSG to determine link behavioral changes that may need to be done to support next generation networks.

HSSG Goals Include:

- Increase Available Network Bandwidth at both Aggregation Points and the Core through a MAC Rate of 100Gbps.
- Reduce the OpX Cost of both LAG and physical link management by deploying larger pipes.
- Increase Network Efficiency by reducing the LAG environment through larger pipe sizes.

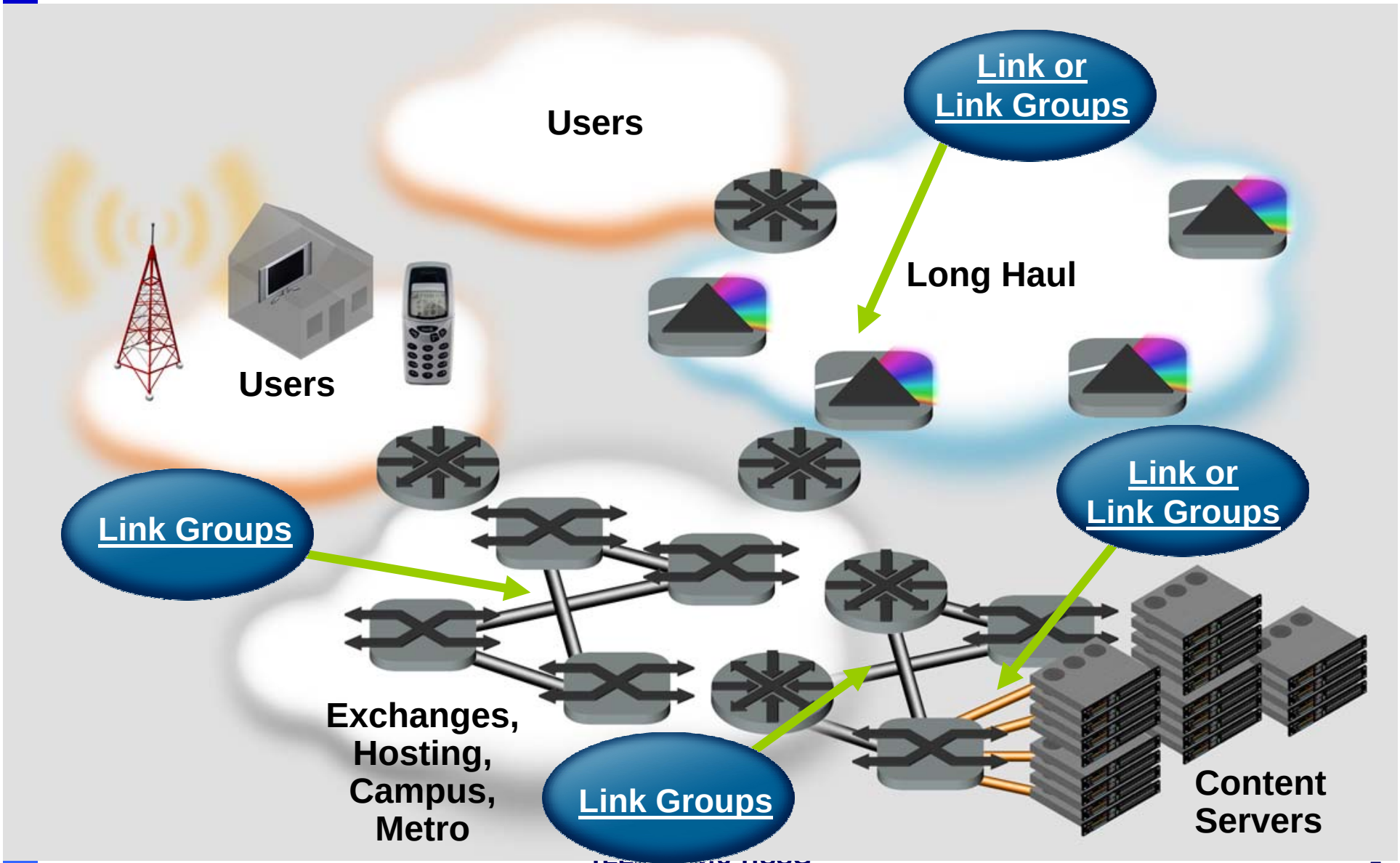
Growing the HSSG Goals to Include:

- Server connections to the aggregation point at a rate of 40Gbps.

40Gbps Server Space:

- Could be addressed by deploying 4x10Gbps LAG today.
- Market for this space is suggesting that LAG won't work here.
- Contributors in this market space are requesting a 40Gbps pipe.
- Server space covers short distance in the data center or long distances interconnecting data centers.
- *Let's look at this closely … because systems companies have to develop the Aggregation points for the 40Gbps connections.*

Defining the Physical Network

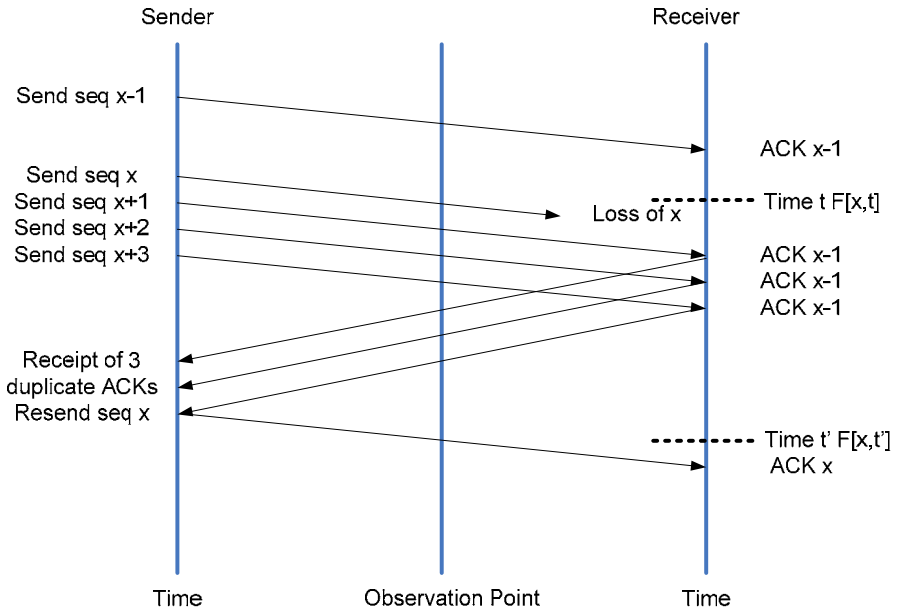


Server Applications Used on the Network

- TCP / IP
  - Transport Control Protocol / Internet Protocol
  - Uses Include:
    - File Transfer
    - File Sharing
    - Internet Relay Chat
    - VOIP
    - … goes on
  - Great concept, but at times is very inefficient use of bandwidth.
    - Is LAG slowing TCP down?
    - Would a bigger pipe work more efficiently?
    - What makes TCP not so efficient?

When TCP Packets are Out Of Order

- TCP receiver gets a segment out of order or receives segments in the buffer with one missing and is waiting for the missing segment to be resent.
- Receiver Generates duplicate ACK
- TCP Fast Retransmit algorithm looks for three ACK as an indication the packet was lost and then resends.
- The Sender waits for three duplicate ACK to distinguish between packet loss and packet reordering.



Retransmission due to packet loss

TCP Packet Delivery Performance

- Application Performance is Effected by Out Of Order Arrival of Packets During the Following Operations [1]:
  - Packet Striping: When packets from the same flow are placed in multiple queues with different priorities or different loading.
  - *Retransmission: When a lost packet is resent.*
  - Load Sharing: When packets from the same flow are processed on multiple processors.
  - *Priority Scheduling: When a flow exceeds the negotiated bandwidth constraints, packets are either dropped or delayed using a lower priority.*
  - Route Fluttering: Packets from the same flow take different routes.

TCP Out of Order Causes Congestion

- ***Study:*** Characterization of Out Of Order behavior in a network with 19 million TCP connections show consistently the following [2]:
  - About 5% of packets are out of order.
  - *Those packets out of order are primarily the result of packet loss.*
- Fast Retransmit during packet loss causes the sender to believe the network is experiencing congestion and therefore erroneously slows flow rate.

Throughput of a TCP Connection

- Mean Steady-State throughput of a TCP connection can be approximated by the following equation [3]:

$$S = \frac{B}{RTT}\sqrt{\frac{3}{2bl}}$$

- S: Mean throughput
- B: Connection packet size
- RTT: Round Trip Time
- b: Number of packets acknowledged be each ACK packet
- l: Loss event rate as seen by the sender

Summary

- TCP Performance between the sender and the receiver, looking at packet loss and reordering, is *primarily effected by packet loss, not out of order packets. This is further compounded by distance between the sender and the receiver.*
- Both 4x10Gbps LAG or a 40Gbps pipe will perform close to the same because out of order packets are not the issue in link degradation in this space.
- If we are going to address the Server Market, providing a 40Gbps pipe is not “THE” answer.

Recommendations

- Recommend HSSG Continue defining 100Gbps.
- Recommend the 40Gbps Server Market use LAG rather then a single 40Gbps pipe.
- Recommend IEEE802.3, IEEE802.1, and IETF begin a joint project to optimize both LAG and TCP performance over IEEE802.3 links to address TCP packet loss deficiency for the Server Space.

References

- [1] “Reorder Buffer-Occupancy Density and its Applications for Measurement and Evaluation of packet Reordering”, Piratla.
- [2] “Measurements and Clissification of Out-of-sequence Packets in a Tier-1 IP Backbone”, Jaiswal.
- [3] “TCP-Load Balancing: The Aequitas Equilibrium”, Elhaddad.

- Credit to both Dr. Thomas Hacker, Purdue University, and Dr. Dennis Guster, St. Cloud State University for their help in identifying root cause of performance degradation in TCP environments crossing single and multiple physical links.