TCP EXTENSIONS FOR A SATELLITE CHANNEL

Xuenan Hu, Tingxian Zhou

The Department of Electronics and Communication Engineering,
Harbin Institute of Technology, P.R. China

ABSTRACT

The usage of Internet is explosively growing. Satellite has become a choice solution breaking through the bandwidth bottleneck and the terrain limit. TCP, which is well suited to terrestrial networks, performs poorly on a satellite channel. The reduced efficiency and QoS(Quality of Service) mainly result from three characteristics of a satellite link: higher bit error rate, the high latency, asymmetry. For this issue, the paper presents connection-subsection network architecture, and brings forward S-TCP based on the architecture.

KEYWORDS

TCP, Satellite, QoS.

INRODUCTION

The demand for Internet has grown explosively in the past several years and the need for enterprises to employ IP standards to transport the information has become increasingly important. Terrestrial networks like PSTN and ISDN, however, which are used to access the Internet, are not fast enough to transfer multimedia information. Moreover, the desire to transport common content to numerous sites has strained the use of traditional terrestrial methods for transmitting such multicast data. Satellite networks, on the other hand, have enough bandwidth to easily transmit large amounts of data to multiple points simultaneously, and satellite receiver equipment has become cheaper since the start of digital satellite broadcasting. But, the traditional TCP/IP, which is designed and optimized to operate in terrestrial environment, performs inefficiently over a satellite link. The reduced efficiency and quality of service can be attributed to three characteristics of a satellite link: higher latency, higher bit error rate and asymmetry. Of course these issues mainly revolve around the fact that while the TCP/IP flow control algorithms are optimized to operate on terrestrial networks it leads to serious bandwidth waste operating on satellite links.

ISSUES AND CHANLLENGES

TCP throughput is the most significant QoS in the system because it is a performance factor that users can perceive directly. The decrease of TCP throughput is mainly result from the three inherent characteristics of satellite and it is important to analyze the characteristics and their effect on TCP throughput.

Higher Bit Error Rate: Measurements demonstrate that satellite channels have higher BER (Bit Error Rate) values around 10^{-6} than common terrestrial channels. Unfortunately, the TCP flow control algorithms are a loss sensitive and always mistake link corruption loss falsely for congestion loss. So satellite channels corruption loss falsely triggers congestion avoidance mechanism and reduce window size, even though the network is uncongested [1]. Additionally, the loss of ACK packets reduces throughputs further. As a result, TCP's congestion avoidance mechanism can severely limit growth of congestion window, so that satellite channel bandwidth can be wasted.

Higher Latency: Satellite links have a higher latency than terrestrial links. For example, one-way propagation delay is approximately 280ms for GEO (geostationary earth orbits). TCP uses a closed-loop feedback mechanism to transmit data. To avoid congestion in the network, every connection sends data at the lowest rate at first and then increases its transmission rate as corresponding acknowledgements are received for the data sent [2]. The bandwidth detection process consists of two phases. In the first phase (i.e. slow start phase), congestion window is initialized to one segment and is increased by one segment for every new acknowledgement received. Once congestion window threshold is reached, the congestion avoidance phase is started and the increase in congestion window is increased by one segment each round-trip time. So the amount of time required for a TCP connection to achieve full bandwidth is substantially dependent on round-trip time [3]. And the average channel utilization efficiency is very low in satellite links of long latency and large amount of bandwidth is wasted. However, the algorithms are necessary to prevent congestive collapse in a shared network.

Asymmetry: There are large bandwidth asymmetries between up and down data channels in many satellite systems. Bandwidth asymmetry can limit TCP's throughput. The downlink bandwidth (from the satellite to the ground) is substantially larger than the uplink bandwidth, with rations of 1000:1 common. This is due to cost limitations associated with increased transmitter power and antenna size requirement for higher bandwidth. Slower reverse channels seriously limit ACK speed so that forward channels data throughput can be greatly reduced.

CONNECTION-SUBSECTION NETWORK ARCHITECTURE

Since TCP inefficiency is result from the difference between satellite link and terrestrial link, if satellite link can be made transparent to end client, the issues above would be solved. Based on this thought, connection-subsection network architecture is brought forward. In this architecture, the connection between client and server is broken into three separate connections (see Figure 1): a TCP connection between the client and the gateway G1, a satellite protocol connection between the two gateways(G1 and G2)and a TCP connection between another gateway G2 and server. The gateway G1 translates TCP to a specialized protocol for satellite channel, and then gateway G2 translates back to TCP. The primary advantage of this architecture is that the segment over the satellite can use a protocol optimized for satellite channel, while the terrestrial segments of the connection can continue to use TCP. This provides improved performance while remaining entirely transparent to the end user. More importantly, a protocol optimized for satellite channel on segment between two gateways should be considered and designed. In fact, this optimized protocol is an updated TCP for satellite channel, so we name it S-TCP.

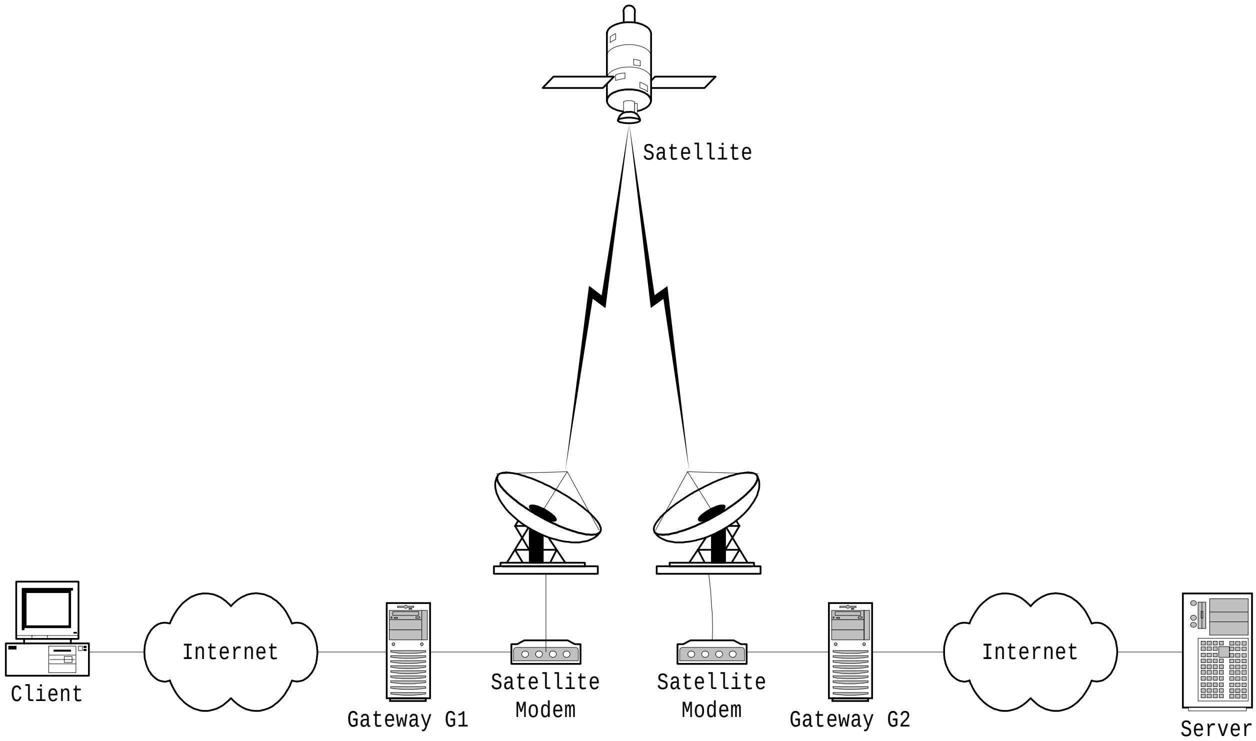

Figure 1 connection-subsection network architecture

S-TCP

Since S-TCP is updated TCP, it is necessary to discuss how to improve on TCP for a satellite channel.

Large Window: TCP utilizes a sliding window mechanism to limit number of packets in flight. The original TCP limits the window size to 64KB. The throughput of a TCP connection is bounded by the TCP’s window size and the round-trip time (RTT) [4](see Formula (1)):

$$throughput = \frac{window}{RTT} \quad (1)$$

TCP historically has been limited to a maximum window size of 64KB(65536bytes) and many TCP implementations use a default window size of only 8KB, so only throughput 114kbps per connection can be achieved. It is also obvious that the throughput of a TCP connection is independent on the channel bandwidth. Large mount of satellite channel bandwidth is wasted. It is imperative under the situation to adjust window size to a large value for a satellite channel.

$$throughput = \frac{8kbytes}{560ms} \approx 114kbps \quad (2)$$

A New Acknowledgement Mechanism: From the section 2, we can conclude that the traffic of acknowledgement must be decreased to improve throughput of a TCP connection since back channel bandwidth is a bottleneck to the throughput of forward data. Traditionally, receiving terminal send an acknowledgement for every received data packet. To decrease the mount of reverse acknowledgement data, a new acknowledgement mechanism should be thought of, in which receiving terminal sends an acknowledgement for several received packets. The question that how many pockets should be received before sending an acknowledgement need be studied. It is not only dependent on the bandwidth rate of back channel to forward channel.

New Flow Control algorithms: TCP’s flow control algorithms are sensitive to delay so that the long delay of satellite links results in a decreased utilization efficiency of channel bandwidth. What we can do is to change flow control algorithm because the delay of link cannot be adjusted. It is not difficult to discover that increasing the initial window size (i.e. allowing more packets to be sent at slow start, thereby triggering more ACKs and hence quickening growth) will increase bandwidth utilization efficiency. But we have to consider the fact that too large initial window can be aggressive and increase the packet loss probability. So the exact choice of the initial window size is an open issue. It is cheering that the exact initial window size has been found out. An initial window of four segments save start time significantly [5]. If the rate of window growth is increased, then average ramp up time before reaching full window is decreased and average channel utilization efficiency is increased. To achieve this goal, the step length of window growth should be increased or the law of window growth should be changed. In the new slow start algorithm, the value of window should be increased by multiple segments for each received ACK, while the value of window is increased by 1 segment for each received ACK in original slow start algorithm. In the new congestion avoidance algorithm, the value of window should be increased by several segments every RTT, while the value of window is increased by 1 segments every RTT in original congestion avoidance algorithm.

Loss Distinguishing Mechanism: TCP’s congestion avoidance mechanism cannot distinguish between the corruption loss and the congestion loss, and then both the corruption loss and the congestion loss can trigger congestion avoidance mechanism so that the value of window and throughput is decreased. When only the corruption loss occurs, the congestion avoidance shouldn’t be triggered because there is no congestion in the network path. If a mechanism that can distinguish packet loss due to congestion from a loss due to corruption is found, a large amount of bandwidth wasted will be saved on satellite links of higher BER. However, confusion between the corruption loss and the congestion loss is not allowed. A failure to identify congestion loss can exacerbate congestion.

CONCLUSIONS

TCP flow control algorithms affect its throughput on satellite channel substantially. The issue mainly arises from three characteristics of satellite links: higher BER, high latency and asymmetry. Fortunately, many methods are found to improve TCP performance on satellite channel. This paper proposes the connection-subsection network architecture and S-TCP (i.e. an updated version of TCP for satellite channel) based on this network architecture. This design provides a transparent solution to terminal users and improves end-to-end performance. Nevertheless, there is a lot of work to do for the optimization of the performance.

REFERENCES

1. M. Allman. et al., “Ongoing TCP Research Related to Satellites,” Internet draft, draft-ietf-tcpsat-res-issues-08.txt, June 1999.
2. V. Jacobsen, RFC1072. May 1992.
3. W. Stevens, “TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms,” RFC2001, January 1997.
4. J. Postel, “Transmission Control Protocol”, RFC793, September 1981.
5. Sally Floyd, February 1997. Note to end2end-interest mailing list.