International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 3, Issue 3 (May-June 2015), PP. 360-364

OVERVIEW OF TCP PERFORMANCE IN SATELLITE COMMUNICATION NETWORKS

Omorogiuwa, .O and Osahenvemwen, O. A.
Department of Electrical and Electronic
Faculty of Engineering and Technology.
Ambrose Alli, University.
Ekpoma, Edo State, Nigeria.
osahenvemwnaustin@ymail.com

Abstract- This study presents the "overview of TCP performance on satellite communication networks", aimed at satellite characteristics, their effects on throughput selected link control protocols and various method proposed for enhancing TCP throughput on satellite networks. Literature reviews on satellite link characteristics and their effects on TCP operation in satellite communication networks. Different improve strategies that have been proposed to enhance TCP data throughput on satellite links. The choice of frame size (n in bits) and window size (W in number of frames) used to improve data throughput on satellite links were considered in this study. Also, the role of sliding window flow control protocol was considered. However, the sliding window method ensures that traffic congestion on the networks is minimized and also, increases the TCP throughput in satellite communication networks.

KEYWORDS: Frame size, Satellite, TCP, Throughput, and Window flow.

I. INTRODUCTION

Recent time, as the quest and demand for multimedia began to witnessed an increase in both the number of subscribers and infrastructure development such satellite. Satellite communication is a microwave transmission system utilizing non terrestrial relay stations positioned in space, where is no atmosphere. Different protocol and application such as Transmission Control Protocol (TCP) and world-wide communications were developed to enhance satellite internet services. Also, users in remote areas, users in regions without well-developed terrestrial networks and mobile users are all potential beneficiaries of satellite communication services.

The Transmission Control Protocol (TCP) is one of the core protocols of the Internet protocol suite (IP), and is so common that the entire suite is often called TCP/IP. TCP provides reliable, ordered, error-checked delivery of a stream of octets between programs running on computers connected to a Local Area Network (LAN), internet or the public Internet. The TCP/IP resides at the transport layer. Web browsers use the TCP whenever it is connected to servers on the World Wide Web, and it is used to deliver emails and transfer files from one location to another (Mario, 2001). TCP is a reliable stream delivery service that guarantees all bytes received will be identical with bytes sent and in the correct order. To guarantee packets transfer reliability over the communication networks, a technique known as positive acknowledgment with retransmission is deploy. This fundamental technique requires the receiver to respond with an acknowledgment message as it receives the data and also, the sender keeps a record of each packet it sends. The sender also maintains a timer from when the packet was sent, and retransmits a packet if the timer expires before the message has been acknowledged. The timer is needed in case a packet gets lost or corrupted (Thomas, 1999). While IP handles actual delivery of the data, TCP keeps track of the individual units of data transmission, called segments; a message is divided into segments for efficient routing through the network. For example, when an HTML file is sent from a web server, the TCP software layer of that server divides the sequence of octets of the file into segments and forwards them individually to the IP software layer (Internet Layer). The Internet Layer encapsulates each TCP segment into an IP packet by adding a header that includes (among other data) the destination IP address. When the client program on the destination computer receives them, the TCP layer (Transport Layer) reassembles the individual segments and ensures they are correctly ordered and error free as it streams them to an application (Luglio et al; 2009). Transmission Control Protocol accepts data from a data stream, divides it into chunks, and adds a TCP header creating a TCP segment. The TCP segment is then encapsulated into an Internet Protocol (IP) datagram, and exchanged with peers (Luglio et al; 2009). The TCP packages the data from these buffers into segments and calls on the internet module [e.g. IP] to transmit each segment to the destination TCP (Kohei, 2011; Alain, 2013; www.cse.wusti.edu/-jain//cv/raj-jain-book1-high). A TCP segment consists of a segment header and a data section. The TCP header contains 10 mandatory fields, and an optional extension field. The data section follows the header. It contents are the payload data carried for the application. The length of the data section is not specified in the TCP segment header. It can be calculated by subtracting the combined length of the TCP header and the encapsulating IP header from the total IP datagram length (specified in the IP header) TCP is the layer 4 protocol which ensures reliable end-to-end communication implementing the concept of the acknowledgement of the received data. When the end-to-end delay is high, as when a satellite link is part of the path, TCP performance rapidly decreases because the window takes a very long time to increase as well as the pipe to be filled. In order to improve TCP mechanism efficiency over the satellite links, many solutions has be adopted. Some of them are specifically proposed for satellite while others more general (Abdelrahman et al, 2002; Alain, 2013). The long propagation delay of a geosynchronous satellite path can result to fundamental problem for some satellite applications. Interactive applications, such as a telnet session or a game of Quake, are going to be frustrating for many satellite users. These inherent delays in the delivery of a message over a satellite link due to the finite speed of light and the altitude of communications satellites. Many communications satellites are located at Geostationary Orbit (GSO) with an altitude of approximately 36,000 km. At this altitude the orbit period is the same as the Earth's rotation period. Therefore, each ground station is always able to "see" the orbiting satellite at the same position in the sky. The propagation time for a radio signal to travel twice that distance (corresponding to a ground station directly below the satellite) is 239.6 milliseconds (ms). The ground stations at the edge of the view area of the satellite, the distance traveled is 2 x 41,756 km for a total propagation delay of 279.0 ms. These delays are for one ground station-to-satellite-to-ground station route (or "hop"). Therefore, the propagation delay for a message and the corresponding reply on one Round-Trip Time (RTT) could be at least 558 ms. The RTT is not based solely on satellite propagation time. The RTT will be increased by other factors in the network, such as the transmission time and propagation time of other links in the network path and queuing delay in gateways. Furthermore, the satellite propagation delay will be longer if the link includes multiple hops or if intersatellite links are used. As satellites become more complex and include on-board processing of signals, additional delay may be added (Kohei, 2011, wood, et al; 2000). Other orbits are still used by communications satellites including Low Earth Orbit (LEO) and Medium Earth Orbit (MEO). The lower orbits require the use of constellations of satellites for constant coverage. In other words, as one satellite leaves the ground station's sight, another satellite appears on the horizon and the channel is switched to it. The propagation delay to a LEO orbit ranges from several milliseconds when communicating with a satellite directly overhead, to as much as 80 ms when the satellite is on the horizon.

II. TCP PERFORMANCE IN SATELLITE

Satellite channel characteristics may have effects on the way transport protocols, such as the Transmission Control Protocol (TCP), behave. When protocols, such as TCP, perform poorly, channel utilization is low. While the performance of a transport protocol is important, it is not the only consideration when constructing a network containing satellite links. For example, data link protocol, application protocol, router buffer size, queuing discipline and proxy location are some of the considerations that must be taken into account when designing satellite network. Also, the higher latency with respect to terrestrial networks implies that it takes longer time to reach the optimum window size while a higher packet loss can be experienced as a consequence of the greater BER in particular channel conditions. Furthermore, when the satellite provides wide band access the bandwidth multiple by delay, it resulted to a very large impact on the ramping time. In mitigate impairments several counter measures has be implemented both at physical level (mainly to reduce losses) and at network (including layer 4) level (Alain et al, 2013; http// www isosat.net/user file/file/routers/iso tropic%20; Thomas, 1999).

III. DRAWBACK OF TCP

It is possible to get around some of the inefficiencies of TCP by going around the protocol. This must be considering carefully, because the congestion control in TCP is there for a reason. However, in some situations, such as at the edge of the network, it may be possible to translate to another protocol for a satellite hop or to spoof TCP by acknowledging packets before the satellite hop. It is possible to use multiple TCP connections to improve overall throughput. This can be done by splitting a large file up and sending the pieces over separate connections. It is also done by some browsers, but using multiple connections has been frowned on as an unfair practice by some computer scientists. While bypassing TCP’s slow start and congestion control algorithms may seem attractive in some cases, care must be taken to make sure that the network as a whole is not abused. This is most easily done at the edge of the network where the last leg of the connection is controlled. It might make sense in some cases to disable congestion control completely (for a deep space probe, for example, if it were using TCP for some reason). On the Internet, if one person disabled their congestion control, they might notice an improvement in performance, but if everyone disabled their congestion control, there would be no more Internets (Alain et al, 2013, Tarik et al, 2004).

IV. SATELLITE LINK CHARACTERISTICS

Satellite links have various characteristics that can degrade the performance of TCP. These include:

1) Long delay paths (long feedback path)
2) Large delay based on bandwidth availability
3) Transmission errors (as opposed to congestion loss)
4) Limited bandwidth
5) Asymmetric use
6) Variable Round Trip Times (for some constellations)
7) Intermittent connectivity (handoffs and outages)

By far, the most common type of communications satellite today uses the geostationary orbit. Such satellites have an altitude of 22,300 nautical miles and orbit the Earth once a day, thus appearing to be stationary in the sky. These satellites do not normally suffer from characteristics number 6 and 7 above. Characteristic number 5 is a result of the way satellite systems are typically configured for end users (Wood et al; 2000).

V. METHODOLOGY

Review of various satellite link characteristics and their effects on TCP operation. Different improve strategies that have been proposed to enhance TCP data throughput on satellite links. A case study discuss how the choice of frame size (n in bits) and window size (W in number of frames) might be used to improve data throughput on satellite links that employ the sliding window flow control protocol.

The main characteristics of the end-to-end path that affect transport protocol performance are latency, bandwidth, packet loss due to congestion, and losses due to transmission errors. If part of the path includes a satellite channel, these parameters can vary substantially from those found on wired networks. The following assumptions about the performance characteristics are as follows:

Latency: The three main components of latency are propagation delay, transmission delay, and queuing delay. In the broad band satellite case, the dominant portion is expected to be the propagation delay. In connections of traversing GEO links, the one-way propagation delay is typically on the order of 270 ms, and may be more depending on the presence of interleaves for forward error correction. Variations in propagation delay for GEO links are usually removed by using Doppler buffers. Therefore, for connections using GEO links, the dominant addition to the end-to-end latency will be roughly 300 ms (one way) of fixed propagation delay. In the LEO case, this can be an order of magnitude less. For example, satellites at an altitude of 1000 km will contribute roughly an additional 20 ms to the one way delay for a single hop; additional satellite hops will add to the latency depending upon how far apart are the satellites. However, the delay will be more variable for LEO connections since, due to the relative motion of the LEO satellites, propagation delays will vary over time, and the connection path may change. Therefore, for LEO-based transport connections, the fixed propagation delay will generally be smaller (such as from 40-400 ms), but there may be substantial delay variation added due to satellite motion or routing changes, and the queueing delays may be more significant (http://www effect- of-latency -and- packet- loss on TCP-throughput; Kohei, 2011, Juanjos et al 2012)

Asymmetry: With respect to transport protocols, a network exhibits asymmetry when the forward throughput achievable depends not only on the link characteristics and traffic levels in the forward path but also on those of the reverse path (Thomas et al; 1999). Satellite networks can be asymmetric when; a host connected to a satellite network will send all outgoing traffic over a slow terrestrial link (such as a dialup modem channel) and receive incoming traffic via the satellite channel. Another common situation arises when both the incoming and outgoing traffic are sent using a satellite link, but the uplink has less available capacity than the downlink due to the expense of the transmitter required to provide a high bandwidth back channel.This asymmetry may have an impact on TCP performance (Tarik et al, 2004; Geoff, 2000). Some satellite networks are inherently bandwidth asymmetric, such as those based on a direct broadcast satellite (DBS) downlink and a return via a dial-up modem line. Depending on the routing, this may also be the case in future hybrid GEO/LEO systems; for example, a DBS downlink with a return link via the LEO system causes both bandwidth and latency asymmetry. For purely GEO or LEO systems, bandwidth asymmetries may exist for many users due to economic factors. For example, many proposed systems will offer users with small terminals the capability to download at tens of Mb/s but, due to uplink carrier sizing and the cost of power amplifiers, will not allow uplinks at rates faster than several hundred Kb/s or a few Mb/s unless a larger terminal is purchased (Thomas et al,1999;Juanjos et al;2012).

Transmission errors: Satellite channels exhibit a higher Bit-Error Rate (BER) than typical terrestrial networks. TCP uses all packet drops as signals of network congestion and reduces its window size in an attempt to alleviate the congestion. In the absence of knowledge about why a packet was dropped (congestion or corruption), TCP must assume the drop was due to network congestion to avoid congestion collapse (wood, et al; 2000). Therefore, packets dropped due to corruption cause TCP to reduce the size of its sliding window, even though these packet drops do not signal congestion in the network.

Congestion: With the use of very high frequency, high bandwidth radio or optical intersatellite communications links, the bottleneck links in the satellite system will likely be the links between the earth and satellites. These links will be fundamentally limited by the uplink/downlink spectrum; so as a result, the internal satellite network should generally be free of heavy congestion. However, the gateways between the satellite subnet work and the internet could become congested more easily, particularly if admission controls were loose. In summary, we assume future satellite networks characterized by low BERs, potentially high degrees of bandwidth and path asymmetry, high propagation delays (especially for GEO based links), and low internal network congestion.

Long feedback loop

Due to the propagation delay of some satellite channels (e.g., approximately 250 ms over a geosynchronous satellite) it may take a long time for a TCP sender to determine whether or not a packet has been successfully received at the final destination. This delay hurts interactive applications such as telnet, as well as some of the TCP congestion control algorithms (Matthew 2013).

Large delay and bandwidth product

The Delay and Bandwidth Product (DBP) defines the amount of data a protocol should have "in flight" (data that has been transmitted, but not yet acknowledged) at any one time to fully utilize the available channel capacity. The delay used in this equation is the RTT and the bandwidth is the capacity of the bottleneck link in the network path. Because the delay in some satellite environments is large, TCP will need to keep a large number of packets "in flight" (that is, sent but not yet acknowledged).

VI. IMPROVE TCP TECHNIQUES

The improve performance techniques on TCP protocol on satellite.

Window scale: TCP's protocol syntax originally only allowed for windows of 64 KB. The window scale option significantly increases the amount of data which can be outstanding on a connection by introducing a scaling factor to be applied to the window field. This is particularly important in the case of satellite links, which require large windows to realize their high data rates. The standard maximum TCP window size (65,535 bytes) is not adequate to allow a single TCP connection to utilize the entire bandwidth available on some satellite channels. TCP throughput is limited by the following formula (Geoff, 2000)

$$\text{Throughput} = \text{window size} / \text{RTT} \qquad 1$$

Therefore, using the maximum window size of 65,535 bytes and a geosynchronous satellite channel RTT of 560 ms the maximum throughput is limited to:

$$\text{Throughput} = 65{,}535 \text{ bytes} / 560 \text{ ms} = 117{,}027 \text{ bytes/second}$$

Therefore, a single standard TCP connection cannot fully utilize, for example, T1 rate (approximately 192,000 bytes/second) GSO satellite channels. However, TCP has been extended to support larger windows (Tarik et al; 2004; http//www .doc Stoc.com/ docs/2371885/An-Analysis-of-tcp-startup).

Sliding window protocols are used where reliable in-order delivery of packets is required, such as in the Data Link Layer (OSI model) as well as in the Transmission Control Protocol (TCP).Conceptually, each portion of the transmission (packets in most data link layers, but bytes in TCP) is assigned a unique consecutive sequence number, and the receiver uses the numbers to place received packets in the correct order, discarding duplicate packets and identifying missing ones. The problem with this is that there is no limit on the size of the sequence numbers that can be required. The term "window" on transmitter side represents the logical boundary of the total number of packets yet to be acknowledged by the receiver. The receiver informs the transmitter in each acknowledgment packet the current maximum receiver buffer size (window boundary). The TCP header uses a 16 bit field to report the receive window size to the sender. Therefore, the largest window that can be used is 2^{16} = 64 kilobytes. In slow-start mode, the transmitter starts with low packet count and increases the number of packets in each transmission after receiving acknowledgment packets from receiver. For every ack packet received, the window slides by one packet (logically) to transmit one new packet. When the window threshold is reached, the transmitter sends one packet for one ack packet received. If the window limit is 10 packets then in slow start mode the transmitter may start transmitting one packet followed by two packets (before transmitting two packets, one packet ack has to be received), followed by three packets and so on until 10 packets. But after reaching 10 packets, further transmissions are restricted to one packet transmitted for one ack packet received. In a simulation this appears as if the window is moving by one packet distance for every ack packet received. On the receiver side also the window moves one packet for every packet received. The sliding window method ensures that traffic congestion on the network is avoided. The application layer will still be offering data for transmission to TCP without worrying about the network traffic congestion issues as the TCP on sender and receiver side implement sliding windows of packet buffer. The window size may vary dynamically depending on network traffic. For the highest possible throughput, it is important that the transmitter is not forced to stop sending by the sliding window protocol earlier than one round-trip delay time (RTT). The limit on the amount of data that it can send before stopping to wait for an acknowledgment should be larger than the bandwidth-delay product of the communications link. If it is not, the protocol will limit the effective bandwidth of the link.

Selective Acknowledgments (SACK): Selective acknowledgments allow for multiple losses in a transmission window to be recovered in one RTT, significantly lessening the time to recover when the RTT is large.TCP uses a cumulative acknowledgement scheme in which received segments that are out of sequence are not acknowledged and the TCP sender can only learn about a single lost packet per round trip time. This forces the sender to either wait a RTT to realize if packets are lost, or to avoid retransmitting segments which have been correctly received. SACK is a strategy which allows TCP receivers to inform TCP senders exactly which packets arrived, and then to recover more quickly from lost packets avoiding needless retransmissions.

TCP for Transactions (T/TCP): TCP for Transactions, among other refinements, attempts to reduce the connection handshaking latency for most connections, reducing the user-perceived latency from two RTTs to one RTT for small transactions. This reduction can be significant for short transfers over satellite channels.

Path MTU discovery: This option allows the TCP sender to probe the network for the largest allowable Message Transfer Unit (MTU). Using large MTUs is more efficient and helps the congestion window to open faster.

VII. DISCUSSION

Over the years, satellite communication has provided important technique of sending information to remote areas without infrastructure development in that locality. Also, as the demand of internet services increases there is need to meet the increasing population of the world with good quality of service. However, this study presents “Overview of TCP performance on satellite communication networks”, due to some drawback generated from TCP layer 4. The impairment generate from the satellite link characteristics are latency, bandwidth, packet loss due to congestion, and losses due to transmission errors. The sliding window method ensures that traffic congestion on the network is avoided based on each portion of the transmission (packets in most data link layers, but bytes in TCP) is assigned a unique consecutive sequence number, and the receiver uses the numbers to place received packets in the correct order, discarding duplicate packets and identifying missing ones.

VIII. CONCLUSION

However, the important of satellite communication cannot be over emphasized due to it enormous economical value and technology averment. However, due to increase in demand and services higher frequency band are being deployed to meet this challenges. Also, it is observed that there is a drawback on the TCP throughput technology deployed by satellite communication networks. In addition, possible means to resolve these inherent problems associate with satellite network technology. The impairment generated from the satellite link characteristics are as followed; latency, bandwidth, packet loss due to congestion, and losses due to transmission errors. Investigations were focused on how to improve different strategies that have been proposed to enhance TCP data throughput on satellite links. A case study discuss how the choice of frame size (n in bits) and window size (W in number of frames) might be used to improve data throughput on satellite links that employ the sliding window flow control protocol were highlighted. The sliding window method ensures that traffic congestion on the network is avoided.

REFERENCE

[1] M. Luglio, C. Roseti and M. Gerla, (2009). TCP Performance over Satellite in case of Multiple Sessions per Links using Efficient Flow Control and Real OS, nrlweb.cs ucla.edu/publication/download/122/kaband 04_ tcp pdf [Accessed on 21st April 2013]

[2] Optimized TCP/IP for satellite Communication.[Accessed 5 August 2013].Retrieved from http// www isosat.net/user file/file/routers/iso tropic%20[accessed on 6 June, 2013]

[3] Matthew Lehman, (2013, August, 3). Measuring throughput in a TCP/IP network Retrieved from http// www.techrepublic.com/article/ measuring-throughput [Accessed on15 May, 2013]

[4] Effect of latency and packet loss on TCP throughput. [accessed on 8 August, 2013]. Retrieved from http://www effect- of-latency -and- packet- loss on TCP-throughput [accessed on 3rd April 2013]

[5] Mario Marchese, (2001). TCP modification over satellite channel: study and performance evaluation, International Journal of satellite communications,19:93-110.

[6] Alain Pirovano and Fabien Garcia, (2013). A New Survey on Improving TCP Performances over Geostationary Satellite Link Network and Communication Technologies; Vol. 2, No.1; ISSN 1927-064X E-ISSN 1927-0658, Page 71-76.

[7] Thomas R. Henderson and Randy H. Katz, (2004).TCP Performance over Satellite Channels. www.cse.wusti.edu/-jain//cv/raj-jain-book1-high performancetcp-ip-networking.pdf [accessed on 11 April 2013], Page 1-6.

[8] Tarik Taleb, Nei Kato and Yoshiaki Nemoto (2004).An Explicit and Fair Window Adjustment Method to Enhance TCP Efficiency and Fairness Over Multihops Satellite Networks IEEE Journal on Selected areas in communications, Vol. 22, No. 2, Pp 371 -384.

[9] Thomas R. Henderson, and Randy H. Katz,(1999). Networks Transport Protocols for Internet-Compatible Satellite, to appear, IEEE Journal on selected Areas of Communications,Vol. 1, Pp 1-8

[10] N J Kothari, Bhavika Gambhava and K S Dasgupta. (2010) Performance Analysis of TCP-AFC for Satellite-based Networks. International Journal of Computer Applications 1(10):5–10.

[11] Mario marchese (2001), TCP Modification Over Satellite Channels: Study and Performance Evalution, International Journal of Satellite Communications, Volume 19, Issue 1, Pp 93-110.

[12] Ian F. Akyildiz, Giacomo Morabito and Sergio Palazzo. (2001). TCP-Peach for satellite networks: analytical model and performance evaluation, International Journal of Satellite communications Int J .Satell. Commun ; 19:429}442 (DOI: 10.1002/sat.713)

[13] A.M. Abdelrahman, M.S. Abdalla, B.M. Ali, V. Prakash and R.K.Z. Sabudin,(2002). Improving the Performance of TCP in LEO Satellite Environment. Information Technology Journal, 1: 250-254.

[14] Juanjo Alins, Jorge Mata-Diaz, Jose L.Muñoz,Elizabeth Rendón-Morales, Oscar Esparza (2012). XPLIT: A cross-layer architecture for TCP services over DVB-S2/ETSI QoS BSM, Elservier computer networks, Volume 56, issues 1, Page 412-434.

[15] Geoff Huston, Telstra (2000).Future for TCP, Internet protocol Journal- Vol. 3, No-3, Page 1-5

[16] L. wood, A. clerget, I. Andrikopoulos, G. Pavlou and W. Dabbous, (2000), IP Routing Issues in satellite constellation networks, International Journal of Satellite Communications Special Issue on Internet Protocols over Satellite, Vol.18, No.6, Page 45-51.

[17] Kohei Arai (2011), Backup Communication Routing Through Internet Satellite,(IJACSA) International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No. 9-21.

[18] An analysis of TCP startup over an experimental DVB-RCS platform [online] Available at http//www.docStoc.com/docs/2371885/An-Analysis-of-tcp-startup.[Accessed on 17th July,2013].