Performance Enhancement for Congestion Control in GEO Satellite Networks

Liang Zong¹, Gaofeng Luo²,³, Min Deng² and Chenglin Zhao¹

¹College of Information Engineering, Shaoyang University, Shaoyang 422000, China
²School of Automatic and Electronic Information, Xiangtan University, Xiangtan 411105, China

Received 3 July 2021; Accepted 22 September 2021

Abstract

The high-speed broadband access technology of satellite networks is becoming increasingly mature, but high-capacity data transmission is prone to congestion, which greatly affects the transmission impact of geostationary orbit (GEO) satellite networks. To alleviate the congestion in satellite networks and improve the transmission rate of links, this study proposed a scheme to increase the data transmission volume. The impact of long delay and high bandwidth in satellite networks was mitigated by increasing the congestion window on the sender side, and increasing the number of data transmissions ensured that the proposed scheme fully utilized the high bandwidth delay product of satellite networks. Different types of data loss were used to improve the stability of network transmission. Results show that the proposed scheme can significantly improve the throughput and reduce the response time of a satellite link when a satellite network is congested. The scheme also significantly improves the network performance when the satellite network is not congested. When the network is not congested, the proposed scheme improves the throughput by a factor of 6 over the standard protocol. Even in the case of congestion, the former also improves the throughput by a factor of 2 over the latter. The proposed method provides a useful reference for improving the transmission performance of satellite networks.

Keywords: Congestion control, Satellite terrestrial network, Transmission control

1. Introduction

With the maturity of broadband satellite Internet access technology, various application services are connected to the Internet through satellite links to provide fast, stable, and reliable network communication services for people under cross-regional network coverage. That is, the full Internet coverage is achieved through satellite networks, and people can acquire Internet experience anytime and anywhere through wired local area network (LAN) or wireless (5G, WiFi, and satellite) connection [1-3]. As a result, mobile terminals can connect to different value-added servers through satellite networks to achieve access to massive videos, news and information, and various services. However, the large number of application accesses will cause a dramatic increase in the number of data transmissions, which will exert a great impact on the transmission performance of satellite networks. Especially in remote areas, the degradation of the transmission performance of satellite networks will affect user experience. Current satellite transmission control technology in the transmission of large-capacity data generally adopts a cross-layer design strategy, and the increased complexity of the protocol algorithm on the sender side, which directly affects the end-to-end semantics, is bound to increase the cost of equipment input. In the case of mobile devices at the terminal, the increase in algorithm complexity is likely to lead to excessive energy consumption of terminal equipment, which is inconducive to the promotion and application of satellite networks.

Conventional transmission control techniques typically set an increased initial window in the satellite Internet when dealing with the high bandwidth delay product of satellite networks. Although this strategy can greatly increase the amount of data sent over a satellite network, the large backlog of data can lead to data loss. In a heterogeneous network of satellites and terrestrial networks, data may be lost randomly due to the high bit error rate (BER) of wireless links in addition to the loss caused by congestion. Such data loss due to wireless links is also an urgent problem in satellite communication. In dealing with the long latency of satellite networks, a transmission control protocol (TCP) of performance-enhancing proxy (PEP) can separate a satellite link from a wired network. It is a highly effective solution, but its partitioning of the network protocol is prone to cause data connection reset, which will further degrade the transmission performance of the network. Scholars have conducted considerable work to improve the performance of satellite networks. The Internet Engineering Task Force suggested that the congestion window be increased from one segment to four maximum segments or 4380 bytes [4-6], but the performance of this approach still needs further improvement. The new cross-layer approaches based on the fast-start mechanism and reducing the time of slow-start and congestion avoidance phases can also reduce the impact of long latency in satellite networks [7-10], but the performance enhancement of satellite networks for long latency is still worth exploring. Therefore, how to balance the impact of increasing data transmission volume and reducing BER in satellite networks is still an urgent problem to be solved.

To this end, a new scheme for data transmission is proposed in this study. To ensure that the new scheme fully utilizes the high bandwidth delay product of a satellite network, the scheme mitigates the effects of long delay and high bandwidth in the satellite network by increasing the...
congestion window value on the sender side. At the same time, the new scheme evaluates the congestion level of the network by using the backlog value in the network to distinguish the type of data loss and improve the stability of network transmission. The scheme can provide a reference for the development and optimization of high-performance transmission protocols in satellite networks.

2. State of the art

Much work has been done by scholars on transmission control in satellite networks, focusing on link symmetry, software-defined networks, network system architecture, and multipath transmission. The asymmetric bandwidth of reverse links often leads to congestion in satellite networks. It can be solved by reducing the congestion window when reverse link congestion occurs, thus simultaneously reducing the forward link throughput of satellite networks. Guan et al. [11] proposed a delay-based forward congestion control algorithm for TCP Vegas to alleviate congestion and improve network bandwidth utilization by classifying congestion into different types (forward and backward). The applicability and effectiveness of the proposed algorithm were verified through simulation on the OPNET software, but the scheme did not make good use of increasing the congestion window to exploit the high bandwidth delay product of the satellite link. Govindarajan et al. [12] proposed a collaborative flow regulation (CFR) protocol that provides fair access to real-time and non-real-time applications over networks with high congestion and error rates. The approach simulation showed that the CFR protocol reduced the video frame loss by 30%-40% compared with conventional schemes and by 50%-60% compared with User Datagram Protocol (UDP) traffic, without compromising video quality. The application of this scheme in TCP should be further investigated. Hamdoun et al. [13] used turbo coding (TC) and network coding (NC) techniques to achieve enhanced video quality over a noisy satellite link by using UDP at the transport layer. The techniques evaluated the performance gain of turbo NC (TNC-UDP) over the traditional TC (TC-UDP) protocol and achieved good results. Further simulations are needed to test the performance of these techniques in other applications.

Existing disruption-tolerant approaches may affect network performance due to resource constraints, especially in heterogeneous networks between satellite and terrestrial networks. Babu et al. [14] proposed a new framework, the medium-term disruption-tolerant software-defined network, which implements network control through an additional STORE operation that uses a node’s memory to buffer link disruptions during packets in the TCP/IP stack and forwards the data when the link becomes active. The study improved the throughput by at least a quarter by using the random movement model. Considering that this scheme changes the integrity of the protocol, it can have a significant impact on practicality. The University of Surrey [15] built a satellite-integrated 5G testbed, which was standards-compliant for mobile core and radio access networks, detailing how to integrate satellite and gateway components into the testbed by using virtualized and software-defined solutions. The authors also provided tests for experimentation in a real-time geostationary orbit satellite system, in which the use of satellites in 5G networks can lead to large performance gains. Effective congestion control is essential for Multipath TCP (MPTCP) operation. Thomas et al. [16] proposed a normalized multipath congestion control scheme that achieved TCP friendliness faster by normalizing the growth of individual subflow throughputs rather than the throughputs themselves. The protocol satisfied TCP friendliness in throughput growth and throughput reduction periods.

To support ships at sea with satellite Internet connections and provide bandwidth optimization over high-latency and low-bandwidth links, Arfeen et al. [17] investigated how optimization can be performed on the client side to improve the quality of the user’s Internet experience at sea without further investment in expensive WAN equipment and changes to the operational network. The proposed scheme performed a statistical analysis of the traffic of various web browsers and suggested ways to reduce the amount of transport layer traffic originating from the client. Bujari et al. [18] revisited the role of PEPs with an alternative to custom virtual networks to optimize the performance of traffic over satellite links. The resulting virtual PEP complied with virtualization standards and leveraged advanced features of the latest technologies. The solution applied network softwareization and virtualization technologies to satellite networks, providing a new way of thinking for future studies in space networking technologies. The widespread deployment of space information networks (SINs) had allowed space communication requirements to change dramatically over the past decade. The dramatic increase in traffic had placed tremendous pressure on the network traffic control of SINs. Pan et al. [19] designed a SIN architecture based on network traffic control to enhance network performance. The scheme used intranetwork load balancing and intranetwork congestion control to demonstrate the feasibility of the proposed architecture. Experimental results showed that the scheme can enhance network performance. Feng et al. [20] introduced the emerging software-defined networking into satellite networks for agile management and automation. The authors focused on resilient recovery of software-defined satellite networks and proposed preliminary solutions to address the fundamental challenges of continuity, fault detection, and recovery of satellite networks. The final study discussed several key outstanding issues that should be addressed urgently. Service requests in nonterrestrial networks were noncentralized in time and spatial domains, which led to service uncertainty and volatility. Lin et al. [21] proposed a model based on state machine tracking trees and a traffic allocation scheme based on cumulative entropy. The new scheme improved the performance in latency and bandwidth utilization by 37.8% and 12.0%, respectively.

With the increasingly large satellite networks, monitoring the network status in a high real-time manner has become increasingly important. Niu et al. [22] proposed a high real-time space constellation network status monitoring technology scheme to support real-time acquisition of various space network status information. The authors carried out the design and implementation of satellite and ground ends on the basis of the idea of star-ground integration. Gao et al. [23] proposed a new game-enhanced compensated switching scheme for MPTCP in 6G software-defined vehicular networks. Experimental results showed that the proposed scheme can effectively solve the data transmission problem. The highly dynamic topology and heterogeneous information in clustered satellite networks presented new challenges for topology control at the information level, and the topology control of traditional networks was extremely inefficient when applied to...
clustered satellite networks. Li et al. [24] proposed a scheme for information topology control in clustered satellite networks, constructed a general information topology model, detailed existing satellite topology control algorithms, and analyzed their problems. However, the experimental scheme lacked experimental simulation. Traditional NC-based multipath routing protocols were inefficient for low Earth orbit satellite networks with long link delays and regular network topologies. Tang et al. [25] first formulated the multipath collaborative routing problem and then proposed an multipath collaborative routing protocol. It designed an efficient no-stop-wait acknowledgment (ACK) mechanism to accelerate data transmission where the source node continuously sent subsequent batches before receiving the ACK messages of the previously sent batches. Simulation results showed that the proposed scheme has some advantages over existing schemes.

The above studies mainly focused on the improvement of satellite networks in terms of the symmetry of links, software-defined networks, network system architecture, and multipath transmission, while less study has been conducted for satellite networks in the scenario of combining long delay and high BER. In this study, we utilize the high bandwidth delay product of satellite networks and propose a scheme to adjust the congestion window only on the sender side by taking full advantage of end-to-end protocols. On the basis of ensuring the end-to-end semantics of satellite networks, the transmission performance of heterogeneous networks is improved, providing a reference for transmission protocols in the congested environment of space networks.

The rest of this study is organized as follows. Section III describes the problems encountered by satellite networks when increasing the congestion window and proposes a new transmission control scheme. Section IV presents the experimental simulations of the proposed scheme. Lastly, Section V provides relevant conclusions.

3. Methodology

Carlo Caini [26] proposed the TCP Hybla scheme, which arises from the analytical estimation of the dynamic performance of TCP (Tahoe, Reno) congestion windows, reducing the dependence of TCP performance on round-trip time (RTT). The dynamics of Hybla congestion windows are as follows:

\[
W_{i+1} = \begin{cases} 
W_i + 2\rho - 1, & \text{SS} \\
W_i + \rho^2 W_i, & \text{CA}
\end{cases}
\]  

(1)

where \(W_{i+1}\) and \(W_i\) are two consecutive congestion windows with initial values of 1 and the CA and SS are slow start and congestion avoidance, respectively.

\[
\rho = \frac{RTT}{RTT_0}
\]

(2)

where \(RTT_0\) is the round-trip delay of the reference connection defined to achieve the compensation effect, and \(RTT\) is the round-trip delay in real time. From the above equation, the algorithm maintains the ACK mechanism used in the standard TCP window update. The update rules for the CA phase are similar to those of the constant rate algorithm. The initial value of the TCP Hybla algorithm is a minimum value of 1. This way, for fast connections \(RTT < RTT_0\), TCP Hybla performs the same control as standard TCP. In addition, the initial slow-start threshold (sthresh) and send cache need to be multiplied by \(\rho\), and the receive window should be large enough; otherwise, the receive window needs to be multiplied by \(\rho\) as well.

3.1 TCP Hybla performance

TCP Hybla allows the long delay connection to obtain the same transmission rate as a relatively fast TCP connection and minimizes the performance reduction result from the long RTT. It may adapt to the long RTT link when no congestion occurs in the satellite network. If network congestion occurs, however, the performance of TCP Hybla will be reduced greatly compared with that of TCP New Reno. TCP Hybla sends a significant amount of data; as data gather in the network, congestion occurs in the gateways and (or) routers. The sender experiences frequent time-out or receives redundant ACK, which enters slow start or fast recovery, and the network performance is dramatically reduced.

Figure 1 illustrates a performance comparison of TCP Hybla and TCP New Reno in a satellite link when network congestion occurs. The throughput of TCP Hybla is lower than that of TCP New Reno in the satellite link when BER is low (1.E-09 to 1.E-07), whereas the throughput is reduced greatly compared with that of TCP New Reno when BER is 1.E-05. Random packet loss probability is larger in the satellite link when BER is high. Nonetheless, large packet loss may also cause congestion; thus, increased packet losses in the network decrease the performance of TCP Hybla.

![Fig. 1. Throughput comparison with congestion loss](image)

3.2 Congestion Window Adjustment

TCP Hybla increases the congestion window (cwnd) via the proportion of the actual \(RTT\) compared with wired network \(RTT_0\) in the slow-start phase. As shown in Equation (1), the value of \(RTT_0\) in the simulations is the wired connection \(RTT\), in which \(RTT_0=25\) ms. However, the \(RTT\) of geostatic satellites in orbit generally can reach 550 ms, such that the size of \(\rho\) is up to 22. Consequently, the congestion window increases up to 4,194,303 (\(2^{22} - 1\)), and this value is extremely large for satellite networks in the slow start. Park et al. suggested that the value of \(RTT_0\) takes 70 ms [27]. Considering the actual existence of congestion, this study sets 75 ms; thereof, the value of \(\rho\) becomes 7, and the congestion window is 127 (\(2^7 - 1\)).

In addition, the slow-start congestion window increase is overly single, always increasing in accordance with \(W_{i+1} = W_i + 2\rho - 1\), which does not reflect the difference.
This condition does not reflect the network state well, such as massive congestion and excessive random packet loss. On this basis, we use a different increase in the congestion window for slow starts, as shown in Figure 2.

![Congestion window adjustment in the slow start](image)

In the first stage, when the congestion window and flightsize are less than \(2\times ssthresh/5\) and \(2\times rwnd/5\), respectively, that is, when they are in the corresponding first interval, the amount of data transmission in the network is small. From the value of the receiver advertised window, we can know that the number of bytes unacknowledged by the receiver is relatively small. Thus, a large amount of data can be injected into the network, setting the congestion window to \(2\times W_{i+1}\). Under this condition, the amount of data in the satellite network can increase rapidly to utilize the network bandwidth.

In the second stage, when the congestion window and flightsize are in the middle of the above boundary points, i.e., both are in the second interval, the data transmission in the satellite network is in a relatively stable state, and the value of the congestion window is set to \(W_{i+1}\). Under this condition, an appropriate amount of data can be injected into the satellite network to maintain the current state of the network.

In the third stage, when the congestion window and flightsize are larger than \(4\times ssthresh/5\) and \(4\times rwnd/5\), respectively, that is, when they are both in the corresponding third interval, a large number of unacknowledged data exist in the satellite network in the transmission state, which can easily lead to congestion. Therefore, the congestion window is set to \(W_{i+1}/2\). Under this condition, the data transmission of the satellite network can be reduced appropriately, which is conducive to the alleviation of congestion.

In the above slow start, not only the magnitude of Hybla increase is retained, but also the three variables cwnd, flightsize, and ssthresh in the network are combined to determine the increase in cwnd jointly. This strategy can combine the network conditions to optimize the increase in cwnd and maintain the network stability. In the detection of random loss, the proposed scheme uses the backlog value in the network to evaluate the congestion level of the network [28], such that the threshold can be reasonably adjusted to mitigate the impact caused by the high BER of the wireless links in the satellite network. The proposed scheme determines whether the data loss is congestion loss or random loss on the basis of the backlog value in the network and then adjusts the threshold value in accordance with the different backlog values to regulate the amount of data sent in the satellite network reasonably.

### 4 Result Analysis and Discussion

The topology of the experimental simulation is shown in Figure 3. The client is connected to the network on the ground through the satellite network, the client is connected to the gateway, and the FTP server is connected to the ground route. To consider the transmission performance of the satellite network fully, the BER range of the satellite link takes the value range of \([10^{-9}, 10^{-5}]\). For the convenience of testing the transmission performance of the satellite network, the client receives a 50 MB file from the server side. The transmission performance under the congestion environment is convenient to monitor during the experiment, and the cache size of the terrestrial gateway is set to simulate the congestion of the satellite network.

![Simulation topology](image)

#### 4.1 Network simulation without congestion

Setting a large cache at the terrestrial gateway can imply that no congestion occurs in the network. Figure 4 shows the response time to complete the download for various schemes at different BERs. The download response time increases with increasing BER for various schemes on 1.E-09 to 1.E-07. The download response time of each scheme is basically the same at 1.E-06 to 1.E-05, but the response time of the proposed scheme is lower than those of the three other schemes. In particular, at a high BER, such as 1.E-05, the proposed scheme has the lowest response time, whereas the New Reno scheme has the longest response time. The proposed scheme is slightly lower than Hybla. At a high BER, Hybla increases the congestion window to reduce the...
response time of downloads effectively if no congestion occurs in the network.

With the same gateway cache settings, Figure 5 shows the throughput of the satellite links, which gradually decreases for all schemes as BER increases. In the range of BER of 1.E-06 to 1.E-08, the throughputs of the four schemes are considerably close. At the BER of 1.E-07 to 1.E-06, the New Reno scheme has the lowest throughput, followed by Veno, and the throughput of the proposed scheme is slightly higher than that of Hybla. When the BER is 1.E-05, the proposed scheme has the highest throughput, whereas the New Reno scheme has the lowest throughput. The throughput of the proposed scheme is slightly higher than that of Hybla. Overall, with only BER and no congestion, the proposed scheme increases the congestion window, which can effectively improve the throughput of the links. When the BER is 1.E-05, if the network condition is good, the link throughput of the proposed scheme is higher than those of the three other schemes in the case of high BER of the satellite network.

Figure 6 depicts that the download response time gradually increases with increasing BER for all schemes. In the range of BER of 1.E-06 to 1.E-05, the download response time pull up sharply, which shows that BER has a great impact on the transmission performance of the satellite network. In the range of BER of 1.E-09 to 1.E-06, the response time fluctuates minimally for different schemes. With a small cache at the gateway and at a high BER (1.E-05), the download response time of Hybla is the longest, followed by that of New Reno, and the proposed scheme has the shortest download response time among the four TCPs. In the case of congestion and high BER, Hybla increases the congestion window and does not effectively reduce the response time to complete the download. The proposed scheme, on the contrary, can distinguish well between random and congestion packet losses, such that it can adapt well to the congestion in the case of high random packet loss and greatly reduce the response time to complete the download.

With a small cache at the gateway, Figure 7 shows the throughput of the satellite link for various schemes with different BERs. As the BER in the satellite link gradually increases, the throughput of the satellite link of different schemes gradually decreases. In the BER range from 1.E-09 to 1.E-07, the throughput of Hybla decreases compared with those of the three other schemes, and the three other TCP versions have roughly similar throughput. In the BER range from 1.E-07 to 1.E-05, the proposed scheme’s satellite link has the highest throughput. Specifically, the proposed scheme has the highest throughput compared with the Hybla scheme. The reason is that the proposed scheme not only has a larger congestion window than New Reno and Veno but also optimizes the size of the congestion window compared with Hybla. The key is that the proposed scheme distinguishes between random and congestion packet losses and reasonably adjusts the threshold setting for fast retransmission. It delays the time to enter the congestion avoidance phase and thus improves the throughput of the satellite link.

Table 1 shows the throughput improvement of the proposed scheme relative to the three other schemes when a small cache is set at the gateway. At different BERs, the proposed scheme has a great advantage. Especially at the
BER of 1.E-05, the proposed scheme improves by 98.44%, 90.21%, and 414.32% relative to the New Reno, Veno, and Hybla schemes, respectively. Increasing the amount of data transmission and distinguishing data loss at a high BER can greatly improve the transmission performance of satellite networks.

From the above simulation, the large window strategy of the proposed scheme achieves high throughput even with a large number of packet losses in the presence of high BER. The threshold adjustment of the proposed scheme also fully utilizes the high bandwidth delay product when data loss occurs. When the network is congested, the satellite link throughput of the proposed scheme is $1.18 \times 10^7$ b/s, and Hybla’s throughput is $0.23 \times 10^7$ b/s (BER is 1.E-05). The proposed scheme sets different thresholds in accordance with the network state to maintain the network stability and avoid continuing to send excessive data to the network in case of congestion.

5. Conclusions

To address the transmission inefficiency of satellite terrestrial networks in congested environments, this study proposed an improved algorithm based on TCP Hybla that adapts well to congested, high-BER satellite networks. It largely maintained TCP Hybla’s congestion window increase rate. Nevertheless, it optimized the size of the slow-start congestion window and effectively distinguished congestion packet loss from random packet loss. The following conclusions could be drawn:

1. Two scenarios for gateways on land were considered: with substantial congestion and without congestion. The improved scheme had an improvement in satellite network performance for various BERs of satellite links.

2. The proposed scheme had a minimal performance improvement over TCP Hybla in the absence of congestion in the network.

3. In the presence of network congestion, the proposed scheme exhibited a significant performance improvement over TCP Hybla, especially at a high BER.

This study proposed an optimized transmission scheme for satellite networks with low transmission efficiency in congested states to improve the transmission inefficiency caused by a single increasing congestion window. It utilized the high bandwidth delay product of satellite networks and proposed a scheme that adjusts the amount of data sent only on the sender side by taking full advantage of the end-to-end protocol. On the basis of ensuring the end-to-end semantics of satellite ground networks, the transmission performance of heterogeneous networks was improved, and it provided a reference for transmission protocols in the congested environment of space networks. Given that the scheme proposed in this study only considered the problem of control of transmission under the characteristics of high BER and long delay, the characteristic environment of other satellite networks lacks in-depth study. Subsequent study can improve the transmission performance of satellite networks over asymmetric links.

Acknowledgements

This work was supported by the Natural Science Foundation of Hunan Province, China (Grant Nos. 2019JJ50620 and 2020JJ4557) and the Scientific Research Fund of Hunan Provincial Education Department (Grant Nos. 20B519, 19B512, and 19A446).

This is an Open Access article distributed under the terms of the Creative Commons Attribution License.

---

Table 1. Percentage of throughput improvement

<table>
<thead>
<tr>
<th>BERs</th>
<th>New Reno</th>
<th>Veno</th>
<th>Hybla</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.E-06</td>
<td>39.12%</td>
<td>16.43%</td>
<td>34.21%</td>
</tr>
<tr>
<td>1.E-05</td>
<td>98.44%</td>
<td>90.21%</td>
<td>414.32%</td>
</tr>
</tbody>
</table>

References


