Durham Research Online

Deposited in DRO:
21 May 2019

Version of attached file:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:

Further information on publisher’s website:
https://doi.org/10.1109/IWCMC.2019.8766787

Publisher’s copyright statement:
© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.
**Abstract**—In this paper, we present the experimental measurement and evaluation of Transmission Control Protocol (TCP) performance over Internet Protocol (IP) using a real, heterogeneous network environment, incorporating at least one leg of satellite and land mobile link that, together, make an Integrated Satellite-Terrestrial Network (ISTN) testbed for our investigation and performance analysis. Originally, the TCP algorithm was developed for short latency and low link error network environments and has become a de-facto standard protocol for the reliable delivery of IP traffic over the Internet, which, in reality, is a heterogeneous network environment nowadays. Using the real latency figures measured with our testbed systems, we numerically analyse the performance of a standard TCP scheme and compare it with the newly developed TCP Hybla algorithm that claims to address performance degradation due to long round-trip-time (RTT) and high wireless link error channels such as Geostationary Satellite Links. The overall performance was compared with the achievable throughput of each of the two TCP algorithms and available bandwidth of the real testbed system. TCP Hybla performed better even with changing real values of RTT obtained from a real hybrid ISTN environment with a Geostationary Satellite link as the testbed.

**Index Terms**—Latency, TCP, Performance, SatCom, ISTNs.

I. INTRODUCTION

Satellite Communications (SatComs) will play an integral role in the Fifth-Generation (5G) New Radio (NR) network infrastructure for achieving most, if not all, of the key requirements for enhanced capabilities of 5G as stated in IMT-2020 and beyond (see Fig. 1a and 1b) [1]. This includes using SatCom’s unique features such as ubiquitous connectivity, availability (99.999%), extreme capacity of GEO/non-GEO High Throughput Satellites (HTS), asymmetric bandwidth capabilities, backhauling and resilience [2]. The global data traffic growth has been projected to reach 69% of which 69% is contributed by mobile traffic for 2016 [4], and the volume of video traffic has grown by 50 petabytes per month by 2021 [3], which is about 12x the data traffic of 2018 which of 69% is contributed by mobile video traffic as shown in Fig. 3 [5]. Satellite channels will be an integral part of the overall communication map, with service providers needing to provide seamless connectivity between terrestrial and satellite. Communication traffic will be dynamically steered using a suite of orchestrators to the best transport options available according to bandwidth, latency, network conditions and other application-specific requirements in order to cope with the growth of the Internet and data-driven society. The indispensable integration of satellite and terrestrial networks (ISTN) is well recognised and promoted in the 3rd Generation Partnership Project (3GPP) standards to provide service ubiquity, continuity and scalability [6], and also in the 5G Public-Private Partnership (5GPPP) phase 2 project initiative. The Satellite and Terrestrial networks for 5G (SaTSG) project explicitly attempts to address the hybrid integration of SatCom and 5G systems [7]. The role of SatCom in terrestrial networks, particularly 5G (see Fig. 2), has been studied extensively in [1, 2, 5, 7]–[9]. Satellite full integration within the virtualised architecture will be facilitated by Software Defined Networks and Network Function Virtualisation (SDN/NFV) beginning with the core of the network and expanding to the edge [2, 9]. Management of the NFV infrastructure will be performed through a Management and Orchestration (MANO) framework, which allows easy integration of multiple applications while software orchestration mechanisms, using SDN, provide programmable network infrastructure, dynamic optimisation of the system and some management functions centralisation. The software-based SDN controllers hold the network intelligence, and configuration of network devices can be done externally by vendor-independent management software [2].

The combination of satellite and terrestrial components to form a single hybrid network has been regarded as a promising approach to significantly improve, and extend, the delivery of communications services. Despite the important advances in satellite technologies, SatCom has not evolved at the same pace as terrestrial communications systems due to its lower economies of scale and inherent technological complexities. Recently, the satellite industry has clearly committed to revisit and revamp the role of satellite communications, in the path towards 5G NR, by improving performance in order to obtain the optimum/efficient utilisation of vast satellite capacity combined with reduced system component complexities [9].

However, to achieve optimum performance and effective utilisation of available capacity of SatCom and hybrid ISTN, there is a need to address performance problems which lead to underutilisation, degraded Quality of Service (QoS) and poor Quality of Experience (QoE). These problems are associated with the widely used/de-facto data Transmission Control Protocol over the Internet Protocol (TCP/IP) network environment, which could involve a satellite channel [10].
Originally, the TCP design concepts were first described in [11], to be connection-oriented and end-to-end (E2E) reliable, and fitting into a layered hierarchy architecture of the IP and supporting multi-network applications [12]. Ideally, TCP should be able to operate over hard-wired connections with packet-switched or circuit-switched networks, but not designed for applications to satellite links [12, 13]. Today, TCP has become the most widely used data transport protocol over IP (TCP/IP), which transports 80-90% Internet traffic [14–18]. However, emerging and evolving network technologies characterised by high-speed (Bandwidth/Capacity) and long distance networks with long Round-Trip-Time (RTT) such as satellite, and wireless links with high link errors introduced new challenges to the performance of TCP, which led to underutilisation of the available capacity [10, 19]. The effective performance of TCP depends on the product of the transfer rate and RTT known as Bandwidth-Delay Product (BDP), which is directly linked to the available capacity [20]. The BDP is a buffer space needed at the sender and receiver to achieve maximum TCP throughput, it also gives a measure of the amount of data that would fill the pipe or the amount of unacknowledged data TCP must handle to keep the pipeline (link/buffer) full [20]. Therefore, in a large BDP network environment such SatCom or hybrid ISTN characterised with both high bandwidth and long RTT called Long, Fat-pipe Network (LFN pronounced "elephant (t)"), TCP performance degrades as shown in this paper and other related research [20]. Here, we conduct an experiment with Satellite-Terrestrial user terminals and equipment (UT/UE) to measure actual latency/RTT and use numerical results obtained with a MATLAB program to evaluate and analyse the performance of standard TCP and TCP Hybla—which is designed for better performance in long RTT networks such as hybrid ISTNs. The paper is structured as follows; Section II provides a summary of works to improve TCP performance over networks with satellite link(s). The experimental set-up and procedure are described in Section III. Performance evaluation and analysis is detailed in Section IV based on experimental results obtained using Geostationary Earth Orbit (GEO) satellite networks.

II. ENHANCING TCP OVER SATELLITE-TERRESTRIAL NETWORKS

Reliable Internet data delivery across any E2E network path can be achieved using de-facto TCP; this includes a hybrid ISTN path. However, the key attributes of Satellite links such as long RTT, high error rates, and high BDP, particularly the GEO HTS, present a serious negative impact on the performance and behaviour of TCP-based transport protocols, posing challenges to Internet connectivity [10, 21]. This performance degradation leads to low channel capacity utilisation of the satellite channel’s extreme capacity. To ad-
dress these issues, TCP congestion control algorithms need to be improved for networks incorporating satellite channel and LFN. Many approaches were proposed until around the year 2010 [10, 19, 21]–[23], but research in this area has slowed down in recent years.

However, popular proposals that improve TCP over satellite channels or in hybrid network environments include, Performance Enhancing Proxies (PEPs), which change the E2E semantics of TCP and were developed to enhance the performance of TCP/IP on networks with link characteristics that suffer performance degradation using the original TCP on an ISTN or purely SatCom network [24]. E2E TCP enhancements without infringing the E2E semantics of the original TCP such as TCP Hybla, based on standard TCP NewReno scheme [25], were also proposed to improve performance deterioration in heterogeneous networks involving satellite links characterised by long RTTs leading to capacity underutilisation. Another E2E TCP improved scheme, based on Binary Increase Congestion Control (BIC), is TCP CUBIC [26]. It improved the window growth function such that it became independent of RTT, and included TCP-friendliness, scalability and RTT-fairness, which allows the window size of the competing TCP flows, sharing the same bottleneck to be approximately the same [26, 27]. E2E TCP enhancement attempted to mitigate the performance deterioration of the TCP in a large BDP network environment such as LFN and hybrid ISTN, particularly involving GEO HTS, more details can be found in [10, 19]–[23, 26]–[28].

However, although there has been considerable research effort to develop a new TCP congestion control algorithms to improve performance over satellite channels and in an ISTN environment, little has been done to evaluate the performance of such an algorithm using the actual parameters such as latency measured by the real experimental set-up of ISTN involving GEO satellites with the highest RTT.

III. EXPERIMENTAL SETUP AND PROCEDURE

A. Experiment Procedure

Experimental measurements were conducted using GEO satellite terminals (SUT) and mobile user equipment (UE) from satellite network service and United Kingdom (UK) Public Land Mobile Network (PLMN) providers respectively. Measurements were conducted three times in a day for fifteen and thirty days for two separate months by transmitting and receiving voice/data signals. The setup assumes a Bent-Pipe (BP) satellite hop, which doubles the latency due to intermediary gateway stations (GWS). Scenarios were developed and used to derive a framework for the measurement, details of which are published in [30].

B. Topology

In this work, we have developed a real testbed with hybrid Integrated Satellite Terrestrial Network (ISTN) environment using a scenario called Satellite-Terrestrial Network Link (STNL) that represents a typical ISTN topology as shown in Fig. 4. This topology relies on GWS for processing and routing, since there is no On-board Processing (OBP) and routing in a BP satellite. However, looking at the dual-hop connection and configuration of the topology, the SUT also served as some form of router and access point, but does little or no data processing like GWS as shown in Fig. 4.

IV. PERFORMANCE EVALUATION

A. TCP Performance Evaluation Framework

To understand how TCP performance deteriorates over a heterogeneous network environment such as the ISTN given by the STNL scenario and configuration shown in Fig. 4, then we need to define the key parameters that affect performance and provide analytical expressions for subsequent evaluation and analysis. The main TCP parameter to determine the performance of any network is the congestion window (CW/ND or W in segments) and its evolution or growth rate function W(t) (in segments or bytes unit). This is also the key to evaluate the performance of the TCP algorithms under study. For instance, the instantaneous transmission rate (or achievable throughput) R(t) (in segments/secs or bytes/secs unit) is directly dependent on W(t) from which capacity utilisation or efficiency can be also derived. In this paper, we focus on performance evaluation and analysis of standard TCP and TCP Hybla using the actual latency measured in an ISTN environment.

Hereafter, we will refer to standard TCP version as TCP while TCP Hybla scheme will just be Hybla. The TCP algorithm implements the window growth function (1) and updates the window after receiving acknowledgement from the receiver using the window update rule (2) [11, 12, 25].

The instantaneous transmission rate is then determined from (1) and the RTT value using (3), while the total data transmitted since the transmission started Td(t) (in segments or bytes unit) is determined by integrating R(t) over the time elapsed from the start to the end of the transmission (Td(t) = \int_0^t R(\tau)d\tau), this is simplified analytically in (4) [11, 12, 25].

\[
W_S(t) = \begin{cases} 
\frac{2(CWND)}{t-t_\gamma} & 0 \leq t < t_\gamma \\
\gamma & t \geq t_\gamma 
\end{cases}
\]

\[
W(t) = \begin{cases} 
SS (Exponent) & 0 \leq t < t_\gamma \\
CA (Linear) & t \geq t_\gamma 
\end{cases}
\]

Where \( t_\gamma = RTT \log_2(\gamma) \), is the time that the Slow Start Threshold (ssthresh) value \( \gamma \) is reached with RTT.
The TCP instantaneous transmission rate $R(t) = W(t)/RTT$ can now be derived from (1) and simplified as;

$$W^S_{i+1} = \begin{cases} W^S_i + \frac{1}{\gamma}; & 0 \leq t < t_\gamma \quad SS \\ W^S_i + \frac{1}{\gamma^2}; & t \geq t_\gamma \quad CA \end{cases} \quad (2)$$

Now we can proceed to compute channel capacity utilisation or efficiency $\eta$ from the calculated achievable transmission rate $R$ and total available capacity $C$ given by (10). This shows how the channel is utilised and whether performance is good or poor.

$$\eta[\%] = \frac{R[kbps]}{C[kbps]} \quad (10)$$

### B. Results, Performance Evaluation and Analysis

The actual latency figures measured by experiments with a real GEO satellite and terrestrial network testbed (see Section III-B), summarized in Tables I and II [30], were used to compute numerical results by implementing the TCP and Hybla congestion control algorithms (see (1) to (9)) in a Matlab program. During algorithm implementations, an ideal channel (i.e. error free with PER = 0) was considered and in the absence of congestion for effective evaluation and analysis and better understanding.

The three overall RTT values (i.e $RTT_{max}$, $RTT_{min}$ and $RTT_{max2}$) from Table II were used to implement the TCP and Hybla algorithms to numerically compute the values of $W(t)$, $R(t)$, and $T(t)$ considering $ssthresh$ ($\gamma$) value of 128-segments as a global variable in both algorithms’ implementation. In the case of TCP, the window evolution rate is slow (larger $t_\gamma$ and longer elapsed time $t$) with the higher values of $RTT$ as shown in Fig. 5, especially during the slow start (SS) phase. This slower window growth in TCP, which takes

### TABLE I

<table>
<thead>
<tr>
<th>Daytime</th>
<th>$\Phi_{max}(m/s)$</th>
<th>$\Phi_{min}(m/s)$</th>
<th>$\Phi_{avg}(m/s)$</th>
<th>$\Phi_{std}(m/s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning</td>
<td>1035</td>
<td>866</td>
<td>971</td>
<td>36</td>
</tr>
<tr>
<td>Afternoon</td>
<td>1025</td>
<td>898</td>
<td>966</td>
<td>33</td>
</tr>
<tr>
<td>Evening</td>
<td>1021</td>
<td>906</td>
<td>964</td>
<td>34</td>
</tr>
<tr>
<td>Overall</td>
<td>1027</td>
<td>890</td>
<td>967</td>
<td>34</td>
</tr>
</tbody>
</table>

$$R^H(t) = \begin{cases} \frac{2^{\gamma RTT_{ref}}}{\gamma^2}; & 0 \leq t < t_\gamma,ref \\ \frac{1}{\gamma} + \frac{R_{Tref}}{RTT_{ref}} (t-t_{ref}) + \frac{\gamma(t-t_{ref})}{RTT_{ref}}; & t \geq t_\gamma,ref \end{cases} \quad (8)$$

Therefore, the final objective of Hybla to achieve maximum data transmission rate (i.e achievable Throughput), independent of a long RTT of channels such as satellite is accomplished by looking at (8).

The amount of data (in segments) transmitted by the TCP Hybla source since the transmission started within the elapsed time $t$, from the beginning of the Hybla connection to the time the threshold value $\gamma$ is reached at time $t_\gamma$ and is computed the same way as TCP from (8) analytically and simplified as in (9) [25].

$$T^H_D(t) = \begin{cases} \frac{2^{\gamma RTT_{ref}}}{\gamma^2}; & 0 \leq t < t_\gamma,ref \\ \frac{1}{\gamma} + \frac{R_{Tref}}{RTT_{ref}} (t-t_{ref}) + \frac{\gamma(t-t_{ref})}{RTT_{ref}}; & t \geq t_\gamma,ref \end{cases} \quad (9)$$

Now we can proceed to compute channel capacity utilisation or efficiency $\eta$ from the calculated achievable transmission rate $R$ and total available capacity $C$ given by (10). This shows how the channel is utilised and whether performance is good or poor.
TABLE II
SUMMARY OF DAYTIME $RTT = 2 * OWD$ COMPUTED

<table>
<thead>
<tr>
<th>Daytime</th>
<th>$RTT_{max}$ (ms)</th>
<th>$RTT_{min}$ (ms)</th>
<th>$RTT_{avg}$ (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning</td>
<td>2070</td>
<td>1732</td>
<td>1942</td>
</tr>
<tr>
<td>Afternoon</td>
<td>2050</td>
<td>1796</td>
<td>1932</td>
</tr>
<tr>
<td>Evening</td>
<td>2042</td>
<td>1812</td>
<td>1928</td>
</tr>
<tr>
<td>Overall</td>
<td>2054</td>
<td>1780</td>
<td>1934</td>
</tr>
</tbody>
</table>

Fig. 5. Standard TCP Congestion Window Growth Rate ($W_S(t)$) Function

Fig. 6. TCP Hybla Congestion Window Growth Rate ($W_H(t)$) Function

Fig. 7. Standard TCP Instantaneous Transmission Rate ($R_S(t)$) Function

Fig. 8. TCP Hybla Instantaneous Transmission Rate ($R_H(t)$) Function

Fig. 9. Standard TCP Total Transmitted Data ($T_S(t)$) Function

Fig. 10. TCP Hybla Total Transmitted Data ($T_H(t)$) Function

of $RTT$ dependence (3) and (4). Throughout the duration of transmission (60 secs) using TCP, the window growth rate was less than 160-segments, while the transmission rate and total data transmitted within the same time period are below 90 segs/secs (73728 bits/secs with Maximum Segment Size (MSS) = 1024 bytes) and 4000 segments respectively.

However, Hybla showed a better performance compared to TCP under the same $RTT$ values and link conditions. Hybla does not show much dependence on different $RTT$ values, especially under the $SS$ phase, thereby achieving better performance, even with the highest value of $RTT_{max}$ measured. For instance, the Hybla window growth rate showed
very little dependence on the changing RTT values during the Congestion Avoidance (CA) phase with $RTT_{\max}$ growing faster (see Fig. 6) due to the multiplier effect of a normalised RTT value $\rho$, while maintaining the same time to reach $\gamma$ value ($t_{\gamma,ref}$) the $RTT_{\max}$ achieved higher $W(t)$ of over 10 $K\text{secs}$ in SS phase as compared to other lower RTT values that exist SS phase with lower values ($\leq 10\ K\text{secs}$). Thus, even the lowest $W(t)$ (over 150 $K\text{secs}$ achieved is better than the highest $W^2(t)$ above). This has significant impact on the $H(t)$ and $T^H(t)$ (see Fig. 8 and 10) that improve performance and capacity utilisation. Hybla implementation achieved a higher transmission rate $R_{\text{ave}}$ of up to 100 $K\text{secs/sec}$ ($819200\ \text{Kbits/sec}$) and total data transmitted $T_D(t)$ of over 300 $K\text{secs}$, all independent of the RTT values as in analytical expressions (8) and (9).

Therefore, using (10) with the same channel capacity $C$, Hybla will achieve better and efficient utilization of channel capacity considering the values obtained for $R_{\text{ave}}$ and $R^*$ above, which highlight the implication of using the standard TCP algorithm over heterogeneous network environment such as hybrid ISTN with at least one GEO satellite link.

V. CONCLUSIONS

The performance evaluation and analysis in this paper confirmed the performance deterioration of TCP in a long RTT heterogeneous network environment such as integrated satellite-terrestrial networks (ISTNs) that will play a vital role in future 5G networks and beyond. Although, TCP Hybla was designed to mitigate this problem, and it was found to perform better than standard TCP even with actual RTT obtained from real experiments using an ISTN testbed, there is a need for a more robust solution that considers more realistically higher $RTT_{\max}$ (above 25 ms). The results of our investigation showed that latency ($RTT$) is one of the key parameters that degrades the performance of TCP protocol over the internet.

REFERENCES


