A Framework for End-to-End Latency Measurements in a Satellite Network Environment

Anas A. Bisu*, Alan Purvis†, Katharine Brigham‡, and Hongjian Sun§
Department of Engineering
Durham University, Durham, England, UK
Emails: *anas.a.bisu@dur.ac.uk , †alan.purvis@dur.ac.uk , ‡katharine.brigham@dur.ac.uk and §hongjian.sun@dur.ac.uk

***Abstract*—In this work, a precise method for end-to-end (E2E) latency measurement in satellite Internet protocol (SIP) network environment is proposed. Latency is considered a key parameter affecting the quality of service (QoS) and performance of communications. This is more pronounced in IP over Satellite. Metrics such as throughput and bandwidth performance of communications systems are dependent on latency, which also has a direct impact on other QoS metrics such as Internet packet transfer delay and delay variation or jitter. The upper limits of QoS objective performance metrics are defined by E2E latency for different QoS traffic classes in this environment. Therefore, there is a need to develop efficient methods for the accurate measurement of E2E latency in a SIP environment. Two case study scenarios were developed for satellite and hybrid networks to measure the latency in a SIP environment. Two Geostationary Satellite Network Services were used to compare the performance of the different scenarios and networks. The results demonstrate that at least 50% of the E2E latency is due to processing and transmitting IP packets over the satellite in both scenarios. Inconsistent latency behaviour was also observed from daily results at different times of the day, which may degrade performance of jitter sensitive applications.**

***Index Terms*—Latency, QoS, Performance, Satellite, Communications.**

I. INTRODUCTION

Satellite communications has interesting and unique attributes such as global coverage, resilience, scalability, multi/broadcast capability, bandwidth-on-demand flexibility, reliability, high data rate and high capacity [1]. These particular features may uniquely position satellite communications as a key technology to provide broadband Internet access to remote isolated communities areas for bridging the digital gap and bring human and economic developments in these remote areas. For example, about 62% of 1.2 billion population of Africa live in remote rural villages while 340 million of them travel about 50km to access the Internet [2, 3]. However, end-to-end (E2E) optimum performance and quality of service (QoS) at all network layers are required to meet current and future media-intense applications using satellite internet protocol (satellite IP). Internet applications mainly use the transmission control protocol (TCP/IP) stack originally developed for terrestrial networks, but the performance of TCP/IP applications is highly sensitive to long latency environments like in satellite networks [1, 4]. Latency (i.e., delay) is the length of time for the signal or information transmitted from source (Tx) to be received by destination (Rx) [5]. High latency has direct significant impact on performance and quality of services (QoS) of communications systems, especially when using TCP/IP applications [1, 6]. Optimum system performance can be achieved by improved latency, with next generation 5G networks aiming at maximum of 1ms latency with 1000-fold capacity [7]–[9]. Performance metrics like Throughput, IP packet transmission delay (IPTD), and jitter or packet delay variation (IPDV) are affected directly by latency [1, 7, 10]. This is more pronounced in satellite IP networks, especially with satellites in geostationary earth orbit (GEO) due to longer propagation paths and additional transmission and processing delays from end-to-end [11, 12]. Satellite networks exhibit long propagation latency and most assumptions of round-trip-time (RTT) of satellite are based on its propagation latency [11, 12]. Considering the E2E satellite network path, propagation latency contributes to only about 50% of the total one-trip-time (OTT) or RTT.

The increased use of satellite IP in remote areas, backhaul links, and plan integration of satellite and terrestrial networks (ISTN) in future 5G networks [8, 9, 13] is a wakeup call for developing a more precise way to measure E2E latency in this environment. Today, applications that can convert smartphones and tablets into satellite phones and access terminals by different satellite network provider (SNP) are on the rise. [14] developed an automated method to measure and calibrate latency in real-time for global navigation satellite system (GNSS) transceivers. However, this work is limited to GNSS and latency within the transceiver system and does not extend to the measurements to ground and space segments as a complete path. Furthermore, it is also limited to the use of a bidirectional cable link to couple and loop the signal from transmitter to receiver without leaving the transceiver system [14]. In contrast, the work proposed herein is a precise measurement technique for E2E latency in a heterogeneous network environment involving several heterogeneous links: terrestrial, Wi-Fi, and satellite for the transmission and reception of voice signal over satellite IP.

II. THEORETICAL FRAMEWORK

The proposed scenario models were developed from general latency theoretical framework models consisting of different component contributions along the E2E network path. The components of the network path contributing to the total latency are described below.

A. *Propagation Latency*

This is the time taken by the signal to travel from one node (Tx) to another (Rx) via the communication link [5, 15]. Expressed as a ratio of distance, *d (km)*, and speed, *v (km/s)*, of the medium (speed of light c = 3x105km/s for wireless links), given in (1).

$$T_{Prop} = \frac{d}{v} \tag{1}$$

B. *Transmission Latency*

The time taken to transmit all of the bits in each packet via a communication link of capacity C (in bps) [5, 15] is the transmission latency. This latency is referred to as store-and-forward (serialization) delay, and is dependent on the link rate and packet (or message) size, M (in bits) [16], given in (2).

$$T_{Tran} = \frac{M}{C} \tag{2}$$

C. *Processing Latency*

Time taken to examine packets by hardware and software such as interfaces, applications and network protocols [5, 15]. Layer interactions with packets generate packet-switching (processing) delay. This may be negligible compared to other delays due to the current high speed of devices and software execution. Mathematically expressed in (3) as ratio of buffer size, Bs (in bits) and device rate of processing, Sp (in bps).

$$T_{Proc} = \frac{B_S}{S_P} \tag{3}$$

D. *Queuing Latency*

Time spent by packet in the output buffer, waiting (queuing) before transmission on to the link [5, 15, 16]. This latency is dependent on the number of packets waiting for transmission. Thus, it is proportional to the traffic (heavy or light) and average packet arrival rate on the link. Considering M/M/1 queuing model [16] the latency can be expressed in (4).

$$T_{Queue} = \sum_{i=1}^{n} \rho_i T_i \,; \quad \rho_i = \frac{\lambda_i}{C_i} \tag{4}$$

where ρ_i is i^{th} link Loading, T_i is average i^{th} system (link/node) delay, λ_i is average (i^{th} packet arrival rate) external traffic on link in packet per second (pps), and C_i is i^{th} link capacity (service rate).

E. *Correlation Function*

Cross-correlation is used for computing the time delay between a pair of signals (Tx and Rx) through convolution to find the maximum similarity of the signals [18]. The cross-correlation for discrete time is defined in (5). The maximum correlation peak can be used to compute the time shift between two signals, and the peak achieves its maximum possible value when the two signals are exactly the same [17].

$$R_{xy}[n] = \sum_{m=-\infty}^{\infty} x[m]y[m+n] \tag{5}$$

Equation (1)-(5) form a general model of latency and is the basis used for actual measurement of total E2E latency using the proposed scenarios in our framework [5, 16].

III. EXPERIMENTAL SETUP AND STUDY SCENARIOS

Satellite networks on geostationary earth orbit (GEO) were used as hypothetical reference paths (HRP) to develop and test our method using two case study scenarios. This involved complete satellite and heterogeneous links to enable performance evaluation and analysis of different link types that are envisioned to be integral part of next generation 5G networks. These study scenarios are code-named satellite-satellite link (SSL), and satellite-terrestrial link (STL). The different experimental setups used are explained in the following subsections and are illustrated in Fig. 1.

A. *Satellite-Satellite Link*

The satellite-satellite link (SSL) established E2E connectivity between two semi-fixed satellite user terminals (SUTs) at about 1m apart and pointed to the satellite through a Line-of-Sight link. Adjustments were made until satisfactory signal strength for communication was obtained within 70-90%. End-users and SUTs were connected using Wi-Fi link 1-20m apart. This scenario was developed to measure the actual E2E latency of Satellite Link Testbed (SLT) as shown in the top region of Fig. 1.

B. *Satellite-Terrestrial Link*

The satellite-terrestrial link (STL) E2E connection was achieved through a heterogeneous network consisting of satellite, Wi-Fi and Public land mobile network (PLMN) links. The user device on the satellite link end connected to the satellite-pointed SUT via Wi-Fi, while the user device on the terrestrial link end connected to PLMN via the nearest base transceiver station (BTS) in the locality. This scenario was developed to measure E2E latency of heterogeneous network testbed as seen in the lower region of Fig. 1.

Communications were routed and processed by a gateway station (GWS) on the ground after being reflected by a transparent satellite, forming a star topology [19].

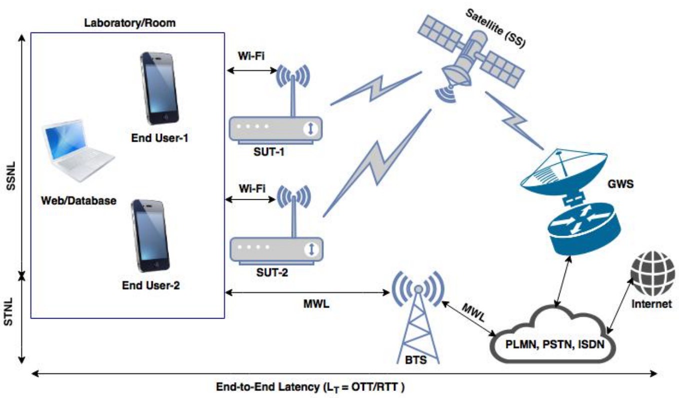

Fig. 1. Experimental Setup of Scenarios

C. Data Acquisition Process

Fig. 2 shows the experimental data acquisition process using the scenarios depicted in Fig. 1, by transmitting (Tx) audio signal from End-user 1 and receiving (Rx) by End-user 2. The correlation function in (5) was used to compute the time delay (latency) between the transmitted and received signals produced under different scenarios through satellite network providers.

D. End-to-End Latency Model For Scenarios

Three mathematical models were established from the two scenarios and network topology. The first is a general model described as linear summation of one-trip-time (OTT) delay (propagation, processing, queuing and transmission) along the path as given by (6). The other two models as described by (7) and (8) were developed from (4) and the two scenarios (SSL and STL) considering the E2E signal path topology.

$$T_L = T_{Prop} + \sum_i T_i \tag{6}$$

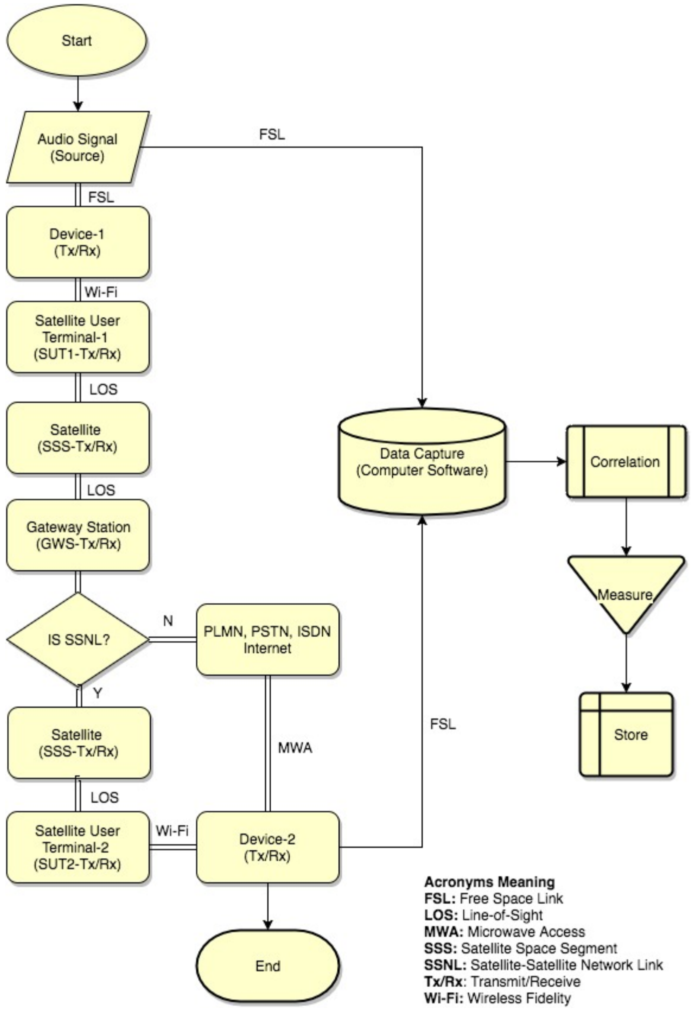

Fig. 2. Data Acquisition Flow Diagram

where T_L is the *total time delay (latency)* in *milliseconds (ms)*, and sum of T_i includes *queuing, processing and transmission* delays.

$$T_{SSL} = 4T_{PropS} + \sum_i T_i \tag{7}$$

$$T_{STL} = 2T_{PropS} + \sum_i T_i + T_{PropT} + \sum_j T_j \tag{8}$$

IV. RESULTS AND ANALYSIS

Results were obtained using two different satellite network providers (SNPs) and the two case study scenarios described in Section III. The first set of experimental results were obtained daily using SNP1 terminals for sixteen (16) days as shown in Fig. 3, while the second set was obtained to increase the data resolution by collecting measurements in the morning (M), afternoon (A) and night (N) for fifteen days using SNP2 and is shown in Fig. 4. These data allow us to study and analyze the latency performance of the two networks using the two developed scenarios as well as the time of day.

A. Daily Performance Analysis of Scenarios

The daily experimental measurements with SNP1 are shown in Fig. 3 and exhibit better performance of latency for STL with average of 956ms and standard deviation of 25ms. SSL has higher average of 1417ms and standard deviation of 31ms using the same network provider (SNP1). The latency was observed to be varying by the day having different values on different days as in Fig. 3 due to different traffic patterns each day and the stochastic nature of latency [16]. These variations may affect IP packet delay variation (jitter) thereby degrading performance of jitter sensitive applications. To improve the resolution of the data, measurements were conducted 3 times a day using SNP2 as shown in Fig. 4.

For SNP2, SSL was observed to have better performance with average latency of 957ms and standard deviation of 64ms, STL performed worst in this case with high average of 1239ms and 72ms standard deviation. The statistical results of the measurements are summarized in Table I.

The resulting analysis of the latency measurements shows that SNP1 performed better under the STL scenario, while SPN2 performed better under the SSL scenario. SNP1 provides broadband Internet service while SNP2 only provides limited Internet services, thus SNP1 may provide an additional advantage of fast Internet access in remote isolated areas and faster voice service to terrestrial destinations in the cities where they exist. SNP2 may serve better as a voice service to connect two remotely isolated users via satellite links in the absence of terrestrial infrastructure on both sites.

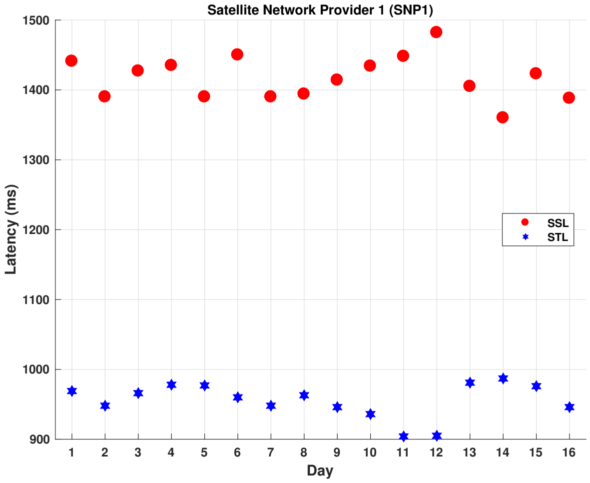

Fig. 3. Performance Comparison of Scenarios with Network provider 1

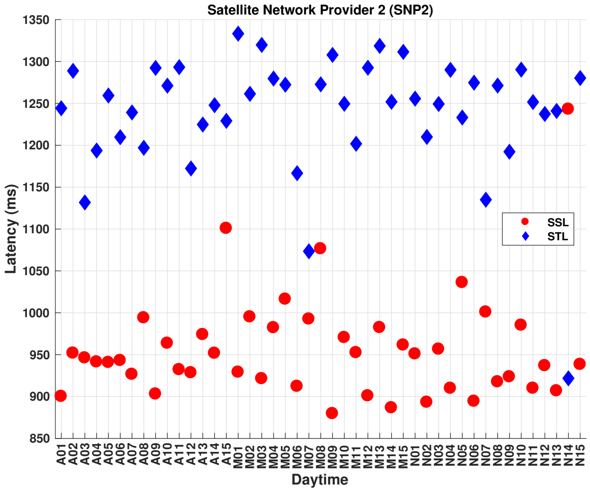

Fig. 4. Performance Comparison of Scenarios with Network provider 2

TABLE I
PERFORMANCE SUMMARY AND COMPARATIVE ANALYSIS

Network		**SNP1**		**SNP2**	
Scenario		**SSL**	**STL**	**SSL**	**STL**
Latency (ms)	**Max**	1482	987	1243	1333
	Min	1360	904	880	992
	Avg	1417	956	957	1239
	Stddev	31	25	64	72

B. Daytime Performance Analysis of Scenarios

A performance analysis of different times of the day was considered by taking measurements in the morning (M), afternoon (A), and night (N) using the SSL and STL scenarios. Fig. 5 shows the daily performance using SSL, while Fig. 6 shows that of STL. The time of day with the lowest average latency (i.e., best performance) using SSL occurred in the afternoon with 953ms and 48ms standard deviation. The highest average latency (i.e., worst performing) occurred at night with 960ms and 88ms standard deviation. This time of the day also has the highest value of 1243ms with SSL scenario as seen in Fig. 5. The STL performance shown in Fig. 6 depicts better performance at night with average latency of 1222ms and 93ms standard deviation while the highest average of 1261ms occurred in the morning, although its standard deviation of 69ms is less compared to that of the night.

Overall, SSL has lowest latency of minimum 880ms in the morning and maximum of 1243ms at night, while STL has lowest standard deviation of 47ms in the afternoon. The statistical results of the measurements recorded throughout the day are summarized in Table II.

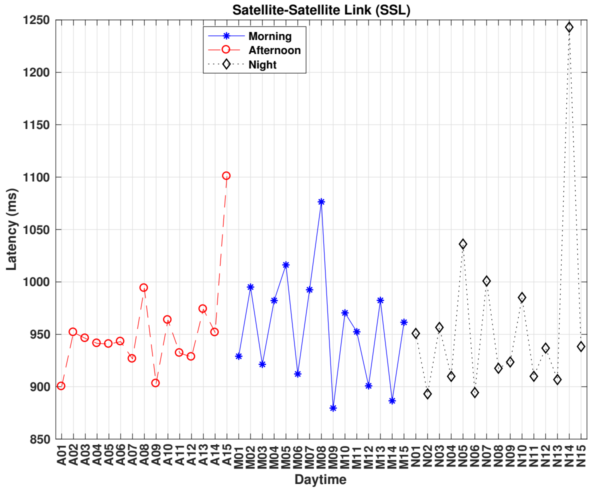

Fig. 5. Daytime Performance Comparison with Satellite-Satellite Link

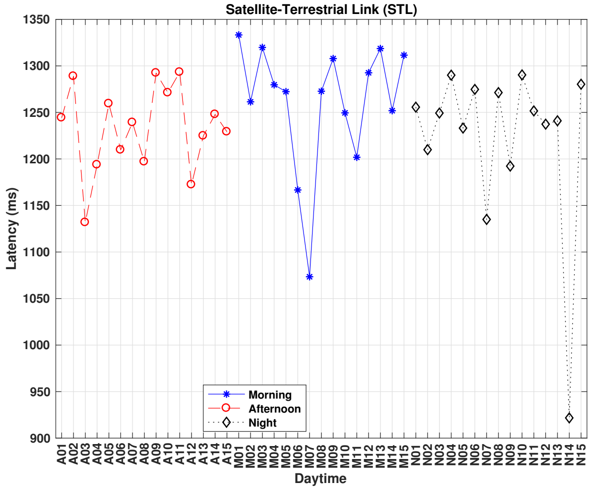

Fig. 6. Daytime Performance Comparison with Satellite-Terrestrial Link

TABLE II
DAYTIME PERFORMANCE COMPARISON OF SCENARIOS

Scenario	**SSL**			**STL**		
Daytime Latency (ms)	L_M	L_A	L_N	L_M	L_A	L_N
Max	1076	1101	1243	1333	1293	1290
Min	880	900	893	1073	1132	922
Avg	957	953	960	1261	1233	1222
Stddev	54	48	88	69	47	93

Statistical information such as the mean and standard deviation are important in latency characterization [16] in order to analyze the effects of E2E latency on data transfer across the network and to identify the level of QoS and performance achievable [1, 5, 6]. Uncertainties normally exist within a communications network; those due to latency can be modeled stochastically in (6) if (6) is viewed as a linear sum of random variables defined by (1)-(4) and (7)-(8) in the case of our proposed scenarios. The main source of uncertainty likely lies in (4) [16] and in future work this will be investigated further.

The performance differences observed across providers may have been due to the respective performance objectives of each network provider. For example, SNP1 may be more focused on (as determined by the customer offerings) the provision of Broadband Internet from remote areas using Broadband Global Area Network (BGAN) terminals that allow up to a 464kbps data rate. Therefore SNP1 may be less optimized for a SSL connection and more focused on heterogeneous links similar to those set up in the STL scenario. SNP2 may be more focused on voice communications than Internet provision over satellite and is only designed for light mobile Internet data usage such as email, instant messaging, and mobile web browsing on smartphones at rates of only 60-144kbps. Therefore, the performance for SNP2 may be more optimized for an SSL-type connection (compared to STL, in which it exhibited lower performance in terms of latency).

V. CONCLUSIONS

This study proposes two scenarios for a precise E2E latency measurement in a satellite IP environment, and the latency performance of the proposed scenarios are also investigated using two different satellite network providers. Our results showed that latency performance in a satellite IP network depends on type of scenario and network provider. Particular scenarios were found to perform better using specific network provider (SNP) while being worse for another. Better performance of latency varies with scenario, network provider and time of day. The daily analysis showed that SSL performance is better with SNP2 while STL is better for SNP1. However, SSL exhibits the lowest latency of 880ms in the morning and maximum of 1243ms at night, while STL has lowest standard deviation of 47ms in the afternoon. Future work will investigate the performance of applications with the recorded maximum, minimum and average latencies. We also intend to study and apply techniques such as Performance Enhancement Proxies (PEPs), Delay-Tolerant Network (DTN) Architecture and User Datagram Protocol to minimize latency and optimize the performance of applications over satellite IP networks. All of the factors investigated in this work (the provider, scenario, and time of day) will be considerations in future work intended to address and minimize latency.

REFERENCES

[1] S. Kota and M. Marchese, "Quality of Service for Satellite IP Networks: A Survey", *International Journal of Satellite Communications and Networking*, vol. 21, pp. 303-349, 2003.

[2] W. B. Group. (2015) Rural Population (% of Total Population). [Online]. accessed: 31/03/2017. Available: http://data.worldbank.org/indicator

[3] G. Faria *et al.*, SatMagazine: Worldworld Satellite Magazine. Nov. 2016 edition, 8-79. [Online]. Available: http://www.satnews.com/magazines.php

[4] P. Woodhead, "Advances in Delay-Tolerant Networks (DTNs): Architecture and Enhanced Performance," in *Electronic and Optical Materials*, vol. 67, pp. 1-269, 2015.

[5] P. Hinton, E. Baker, and C. Hill, "Latency Time for lawyers to get up to Speed?", *Computer Law & Security Review*, vol. 28, no. 3, pp. 340-346, 2012.

[6] T. Braun, M. Diaz, E. J. Gabeiras, and T. Staub, *End-to-End Quality of Service Over Heterogeneous networks*. Berlin, Germany: Springer, 2010.

[7] A. E. Gharsellaoui, S. A. M. Ghanem, D. Tarchi, and A. Vanelli-Coralli, "Adaptive network coding schemes for satellite communications," *14th IEEE Advanced Satellite Multimedia Systems and Signal Processing for Space Communications Conference (ASMS/SPSC)*, Palma de Mallocra, Spain, 5-7 September, 2016.

[8] R. Ferrús *et al.*, "SDN/NFV-enabled satellite communications networks: Opportunities, scenarios and challenges," *IEEE Physical Communication*, vol. 18, no. 2, pp. 95-112, 2016.

[9] X. Artiga, J. Núñez-Martnez, A. Pérez- Neira, G. J. Lendrino Vela, J. M. Faré García, and G. Ziaragkas, "Terrestrial-Satellite Integration in Dynamic 5G Backhaul Networks," IEEE, 2016.

[10] M. Marchese, *QoS Over Heterogeneous Networks*. West Sussex, UK: John Wiley, 2007.

[11] S. Utsumi and S. M. S. P. Zabir, Sumet, 11A new explicit congestion notification scheme for satellite IP networks," *Journal of Network and Computer Applications*, vol. 75, pp. 169-180, November 2016.

[12] C. Caini, P. Cornice, R. Firrincieli, and D. Lacamera, "A DTN approach to satellite communications," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 820-827, 2008.

[13] C. Niephaus, M. Kretschmer, and G. Ghinea, "QoS Provisioning in Converged Satellite and Terrestrial Networks: A Survey of the State-of- the-Art," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2415-2441, 2016.

[14] Y. Xiao, Y. Xu, H. Sun, X. Xu, and A. Zhou, "A Precise Real-Time Delay Calibration Method for Navigation Satellite Transceiver," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 11, pp. 2578-2586, 2016.

[15] T. Eylen and C. F. Bazlamacci, "One-Way Active Delay Measurement With Error Bounds," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 12, pp. 3476-3489, 2015.

[16] J. W. Stahlhut, T. J. Browne, G. T. Heydt, and V. Vittal, "Latency Viewed as a Stochastic Process and its Impact on Wide Area Power System Control Signals," *IEEE Transactions on Power Systems*, vol. 23, no. 1, pp. 84-91, 21 Jan. 2008.

[17] K. Dudacek Jr., K. Dudacek, and V. Vavficka, "Comparison of Short Delay Measurement Methods," *International Conference on Applied Electronics (AE)*, Pilsen, Czech Republic, 8-9 Sept., 2015.

[18] X. Jin, Y. Hua, and Y. Cao, "Research on A New Method of Time Delay Measurement in Telephone Time Service," *IEEE International Frequency Control Symposium (FCS)*, 19-22 May, 2014.

[19] T. Gayraud and P. Berthou, "A QoS Architecture for DVB-RCS Next Generation Satellite Networks," EURASIP Journal on Wireless Communications and Networking, Satellite Communications, pp. 79-87, 2007.