Load Balancing and QoS Provisioning Based on Congestion Prediction for GEO/LEO Hybrid Satellite Networks

---

Citation:

Hiroki Nishiyama, Daigo Kudoh, Nei Kato, and Naoto Kadowaki, "Load Balancing and QoS Provisioning Based on Congestion Prediction for GEO/LEO Hybrid Satellite Networks," Proceedings of the IEEE, vol. 99, no. 11, pp. 1998-2007, Nov. 2011.

URL:

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5934692

Load Balancing and QoS Provisioning Based on Congestion Prediction for GEO/LEO Hybrid Satellite Networks

Hiroki Nishiyama, Daigo Kudoh, Nei Kato, and Naoto Kadowaki

***Abstract*—While GEostationary Orbit (GEO) satellite systems provide us with a wide coverage area, their long delay serves as a significant constraint for real-time applications. On the other hand, Low Earth Orbit (LEO) satellite systems are best suited to delay sensitive applications. However, the coverage and mobility issues of LEO satellites lead to relatively high management costs. In this paper, we devise a new load balancing and Quality of Service (QoS) provisioning scheme to accommodate both real-time and non-real-time traffic based on a new congestion prediction scheme. The effect of this new scheme is expected to improve the efficiency of the GEO/LEO hybrid satellite networks and the QoS satisfaction of end users.**

***Index Terms*—Congestion prediction, Load balancing, Quality of Service, and Satellite communication.**

Paper descriptions

To fully utilize the advantages of hybrid satellite networks integrating different satellite constellations, a QoS-oriented load balancing technology based on congestion prediction is essential.

I. Introduction

Satellite networks providing a global coverage area are an effective solution to provide ubiquitous network access services all over the world. In the past decade, due to the rapid globalization and growing bandwidth of the Internet based on the phenomenal development of the terrestrial wired and wireless network technologies, the role of satellite networks as a system serving network accesses gradually declined. However, it is certainly true that they are still unique network systems in rural, sea, and disaster-stricken areas where no network infrastructure can be easily deployed. In addition, satellite networks have received renewed interest in the recent past. One reason the significance of satellite networks has reappeared is the emergence of high-speed satellite communication networks serving Gigabit links [1]. Also, the promising vision of further broadband satellite networks has been presented [2], [3]. The utilization of Ka-band allows us to achieve several dozen gigabit communications as in optical networks. The increased interest in the observation of the earth by using satellites, the decreased cost of launching satellites, and the development of smaller sized satellites can be also considered as reasons for our interest to study the satellite networks. Indeed, it can be expected that the recent innovations will promote the development of the current satellite networks and lead to the emergence of next generation satellite networks.

As a probable form of the next generation satellite networks, integrating different satellite constellations have been considered with much attention [4], [5]. However, most of the existing satellite network systems have been operating independent of one another. In other words, there is no connection between different satellite constellations. This provides an opportunity to improve the satellites system performance and create the potential for their new usages. In fact, multi-layered satellite networks consisting of multiple satellite constellations having different orbital altitudes have been studied as a space core network, which delivers traffic to any point on the earth [6]–[8]. Fig. 1 shows an example of two-layered satellite networks, i.e., the lower and upper layers represent the constellations of Low Earth Orbit (LEO) and Geostationary Orbit (GEO) satellites, respectively.

In each satellite constellation integrated into the multi-layered satellite network, the number of satellites required to cover almost all of the earth surface is totally different according to their orbital altitudes. For example, in Iridium [9] network system, which consists of LEO satellites orbiting 780km above earth, 66 satellites are needed to cover all of the surface including the poles. On the other hand, just three satellites are enough to provide global coverage except the poles when using GEO satellites orbiting 36000km above the equator. In each constellation, satellites connected by Inter-Satellite Links (ISLs) composes a ring or mesh network, but there is no connection between the constellations. In contrast, the advanced aspect of the multi-layered satellite networks is the existence of Inter-Layer Links (ILLs) connecting the different layers with each other.

In this paper, we focus on the GEO/LEO hybrid satellite network, which is a typical example of two-layered satellite networks, and deal with the load balancing issue with the provision of Quality of Service (QoS) by using the advantage of the interconnection between layers. Furthermore, we demonstrate the effectiveness of our traffic distribution scheme based on the prediction of network congestion through computer simulations. The paper is organized as follows. In Section II, we review the advanced technologies to cope with load balancing and/or QoS provisioning over multi-layered satellite

H. Nishiyama and N. Kato are members of the Graduate School of Information Sciences at Tohoku University, Japan. They may be contacted at bigtree@it.ecei.tohoku.ac.jp, kato@it.ecei.tohoku.ac.jp, respectively.

D. Kudoh works for the NTT East Corporation in Tokyo, Japan. He may be contacted at dkudoh@it.ecei.tohoku.ac.jp.

N. Kadowaki works for the National Institute of Information and Communications Technology (NICT) in Tokyo, Japan. He may be contacted at naoto@nict.go.jp.

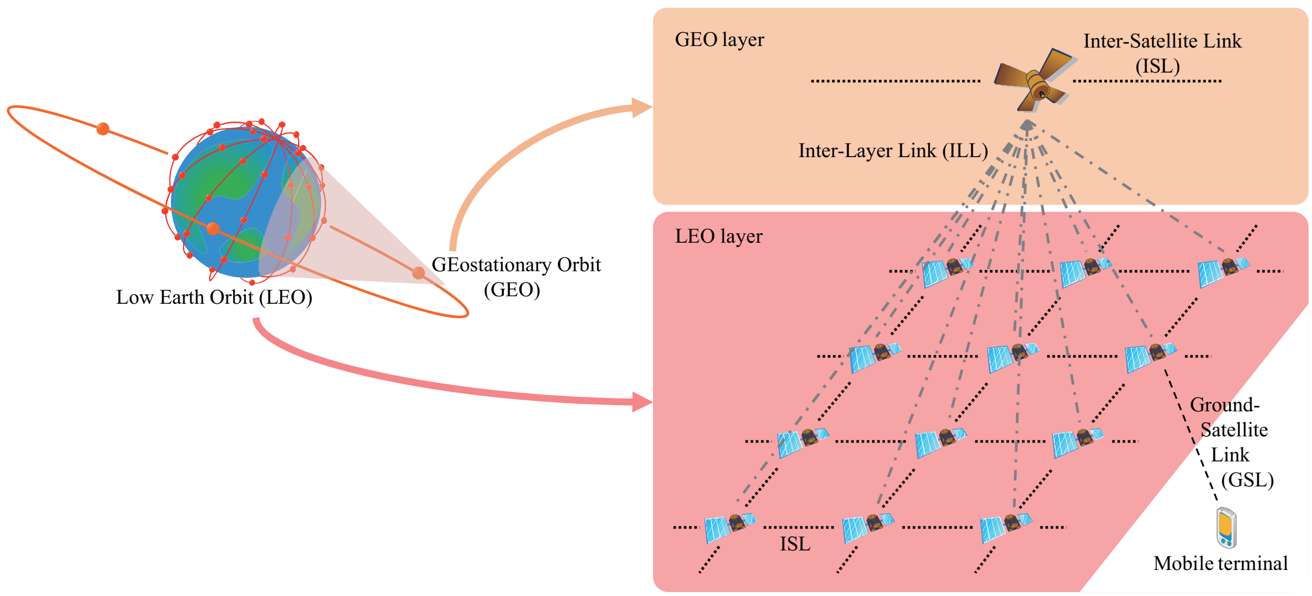


Fig. 1: Considered GEO/LEO hybrid satellite network.

networks. Our scheme aiming at achieving both load balancing and QoS provisioning by using the congestion prediction is described in Section III, and its performance is verified in Section IV. Conclusion is provided in Section V.

II. Load balancing and QoS provisioning in multi-layered satellite networks

Satellite networks can be classified into three categories, namely GEO, Medium Earth Orbit (MEO), and LEO satellites, according to their orbital altitudes. GEO satellites have 36000km orbits, MEO satellites have the orbits from 10000km to 20000km, and LEO satellites have the orbits between 500km and 1500km. Higher altitudes contribute to wide coverage areas, and lower altitudes lead to small communication delays. The most significant advantage of the multi-layered satellite networks is that they have the strengths of combined networks, i.e., the relatively wider coverage area served by upper layers and the comparatively shorter propagation delay provided by the lower layers. From the point of fair and efficient utilization of network resources, wide coverage areas of satellites having higher altitudes are preferred for averaging the load of each satellite. In general, due to the imbalanced population depending on landform, climate, and so on, traffic convergence to certain satellites usually occurs in mesh satellite constellations having lower orbits. For example, satellites above metropolitan areas receive a huge amount of traffic from terrestrial bases while the traffic volume coming from the earth is almost zero on the ocean. On the other hand, short propagation delay due to low altitudes of satellites is attractive for real-time interactive applications such as Voice over IP (VoIP). In fact, there is an enormous difference in round-trip propagation delay between LEO or GEO satellites. Therefore, in multi-layered satellite networks, intelligent routing control technology to effectively balance traffic load according to the QoS requirements is necessary. In the followings, we review the routing schemes developed for multi-layered satellite networks.

As a routing protocol for three-layered satellite networks constructed by integrating GEO, MEO, and LEO satellite constellations, Multi-Layered Satellite Routing (MLSR) [10] has been proposed. In MLSR, GEO satellites gather the information on link delays in each satellite located within their coverage area; the information of LEO satellites is transmitted to the GEO satellite via MEO satellites. GEO satellites calculate and update the routing tables of each satellite, which are notified to MEO and LEO satellites from GEO satellites. By adopting the centralized control mechanism, not only delay-minimized routing but also the reduction of signaling and computation costs can be achieved. On the other hand, MLSR is modified for multicast routing in [11] where the bandwidth utilization of each link is taken into account to calculate the multicast tree instead of the link delay.

The two-layered satellite networks consisting of MEO and LEO satellite constellations have been studied for a long time [12], [13] and several different approaches in route control have been developed [14]–[16]. In the Satellite Grouping and Routing Protocol (SGRP) [14], the MEO and LEO satellite constellations are exactly separated in their roles, i.e., traffic is delivered only by using LEO satellites, and the MEO satellites are just used for network management tasks such as route controlling. LEO satellites positioned within the coverage area of the same MEO satellite are grouped, and they periodically inform the MEO of the information on the delay of their links, which are exchanged among MEO satellites and used for updating the routing tables of LEO satellites. LEO satellites perform route control according to their routing tables calculated by and notified from the MEO. At the same time, to avoid occurrence of network congestion in the LEO satellite network, traffic are delivered without utilizing congested links which are detected by monitoring queue occupancy, i.e., queue length exceeding a certain pre-defined threshold implies the congested link. However, the effect of diverting traffic only in LEO satellite constellation is rather limited.

The efficient utilization of MEO satellite constellation is necessary to effectively distribute traffic to avoid network congestion occurrence. As an example of such solutions, Hierarchical and Distributed QoS Routing Protocol (HDRP) [17] has been proposed. Although HDRP is similar to SGRP in terms of being based on the grouping of LEO satellites, it has two distinct advantages. The first one is the traffic delivery via MEO layer. When the source and destination LEO satellites belong to different groups, the traffic is transferred through the MEO layer, which results in reducing the traffic volume in LEO layer. While HDRP can reduce the risk of occurrence of network congestion compared with SGRP, it is possible that the delay unnecessarily becomes too long even if there is an ISL between the source and destination LEO satellites when they belong to different groups. The second advanced feature of HDRP is the QoS provisioning, i.e., bandwidth guarantee. In the HDRP, upon receiving a new communication establishment request, the optimal route is determined to guarantee the requested bandwidth and minimize transmission delay.

To guarantee the QoS for delay-sensitive applications such as VoIP, Adaptive Routing Protocol for QoS (ARPQ) [18] has been developed. ARPQ employs the routing mechanism similar to that used in SGRP and HDRP. In LEO satellites, when the occupied queue length exceeds a threshold, traffic except the delay sensitive ones are diverted, at a constant ratio, to the neighboring least congested LEO satellite to prevent further increase of end-to-end delay due to the growing queuing delay. On the other hand, delay sensitive traffic is detoured to MEO layer to prevent it from suffering from large queuing delays. Actually, the satisfaction of QoS requirement for specific applications can be improved by using ARPQ. However, the traffic diverted in the LEO layer may cause an additional network congestion at the neighboring LEO satellite because the volume and the direction of traffic being diverted is not appropriately dynamically adjusted according to network congestion situations.

Although there are some route control schemes developed for multi-layered satellite networks as overviewed above, almost all of them have not been designed to fundamentally solve the load balancing problem derived from imbalanced population distribution as described at the beginning of this section. Indeed, the integration of satellite constellations having different orbital altitudes has enough potential to solve this issue. The use of not only a single layer but also multiple layers for traffic distribution by following the appropriate routing algorithms seems to dramatically and efficiently mitigate the network congestion. Therefore, the most straightforward way to innovate a routing technology suitable for multi-layered satellite networks is to develop an enhancement of the routing algorithms designed for a single satellite constellation [19]–[24]. In this regard, the totally different link delays among different layers also need to be taken into account, i.e., QoS-aware control mechanisms developed for multi-layered satellite networks are necessary [25].

A novel approach to cope with load imbalance issue derived from imbalanced geographical distribution of population in LEO or MEO satellite constellations is to predict the occurrence of network congestion based on the positional relation between the satellite orbits and the congested areas as we have presented in [19]. The congested areas imply regions such as metropolitan areas including a lot of traffic sources and destinations. When satellites roam over congested areas, they experience a high probability of network congestion. On the other hand, since the orbits of satellites are fixed, and multiple satellites go around the same orbit, satellites are able to predict the occurrence of network congestion before they actually experience it by exchanging network congestion information among the neighboring satellites. In the next section, we propose an enhancement of the congestion prediction-based load balancing method for multi-layered satellite networks.

III. Congestion prediction-based QoS-aware routing

While some different combinations of satellite constellations can be considered as candidates of the next generation multi-layered satellite network models, we assume the integration of GEO and LEO satellites as depicted in Fig. 1 because their combination has both noticeable aspects, namely, the wide coverage of GEO satellites and the short delay of LEO satellites. The considered GEO/LEO hybrid satellite network has a ring constellation of GEO satellites as the upper layer and a mesh constellation of LEO satellites as the lower layer. While the numbers of GEO satellites and LEO satellites are not regulated to generalize the model, we assume that each LEO satellite maintains an ILL to the above GEO satellite and four ISLs to the backward, forward, right, and left neighboring LEO satellites. As sources and destinations of traffic, mobile users are assumed. Due to the practical limitations in battery, antenna, and devices of the mobile terminals, we assume that each user terminal can simultaneously establish a connection to a LEO satellite by a Ground-Satellite Link (GSL); simultaneous multiple connections and the direct connections to GEO satellites are not assumed.

In the GEO/LEO hybrid satellite network, the load imbalance issue due to the imbalanced distribution of population occurs in the LEO layer. To overcome this issue, we attempt to apply the congestion prediction-based traffic distribution mechanism [19] for a single constellation of LEO satellites to the GEO/LEO hybrid satellite network, and modify its traffic detouring policy to control QoS. In the proposed scheme, only LEO satellites equip the load balancing function based on congestion prediction. Each LEO satellite is allowed to divert traffic not only to its neighboring LEO satellites, but depending on the traffic type, it may also decide to send data up to its overhead GEO satellite. However, except for this load balancing traffic detouring function, the other algorithms in the proposed scheme, such as traffic measurement and path diversification, are similar to those mentioned in [19].

A. Congestion Prediction

The most general way to detect network congestion is to monitor the changes in the incoming traffic rate, and regard the measurements of higher rates exceeding a pre-defined or dynamically adjusted threshold as an occurrence of network congestion. However, the commencement of load distribution upon detecting network congestion is too late in some situations, which can cause significant network congestion. In other words, employing a congestion prediction mechanism for load balancing is necessary to effectively avoid the occurrence of actual network congestion. This serves as the motivation behind our work. Since it is possible in mesh constellations of LEO satellites to know which LEO satellite is going toward the congested area, in the proposed scheme, the satellite above the congested area preliminarily informs the neighboring satellite following itself with the coordinates of the congested area. The congested area can be defined as a circle with its center at the informed coordinate. The radius of the circular field needs to be determined according to the configuration of the satellite constellation and the LEO satellites coverage area. By exchanging such information, the satellite approaching the congested area can predict network congestion and immediately begin traffic detouring upon entering the area without awaiting detection of actual network congestion events. The detailed procedures of congestion prediction are overviewed below.

In the proposed scheme, two different states, i.e., normal and warning states, are defined to present the status of each LEO satellite. Normal and warning states imply that the corresponding LEO satellite has low and high probability of the occurrence of congestion, respectively. In the normal state where no congestion event is detected, a LEO satellite attempts to detect a congestion precursor by just monitoring the incoming traffic rate, and does not exhibit any load balancing behavior. On the other hand, in the warning state where there is a threat of the cause of congestion, not only the monitoring of the traffic but also traffic distribution and the information exchange for congestion prediction are performed according to whether it exists in congested or non-congested areas. The status transition from the normal to warning states occurs when observing a higher incoming traffic rate than the threshold, denoted by λ, or entering into the congested area. On the other hand, each LEO satellite returns its status from the warning to normal states when the incoming traffic rate becomes less than the pre-defined threshold, ν.

When a LEO satellite in the non-congested area moves to the warning state by detecting a growing incoming traffic rate, it begins detouring traffic, and declares the area as a congested area and notifies the neighboring LEO satellites. In contrast, after a LEO satellite transitions to the warning state by entering the congested area, it begins the traffic detouring when the incoming traffic rate exceeds the threshold μ, which is less than λ. Since the probability of network congestion is relatively higher than that in the non-congested area, we use the small threshold to invoke traffic detouring compared with the case of the non-congested area. If the LEO satellite decides to detour traffic due to a high incoming traffic rate, it informs its neighbors that the area is congested. In contrast, if the traffic detouring was never carried out while the LEO satellite flies over the congested area, the area is regarded as being non-congested, and the handover of the information on the congested area will be terminated.

TABLE I: Traffic detouring ratios of each traffic class.

	Traffic class		
	A	B	C
$r_d < I^c_{\text{relay}}$	0	0	r_d/I^c_{relay}
$I^c_{\text{relay}} \leq r_d < I^c_{\text{relay}} + I^b_{\text{relay}}$	0	$(r_d - I^c_{\text{relay}})/I^b_{\text{relay}}$	1
$I^c_{\text{relay}} + I^b_{\text{relay}} \leq r_d$	0	1	1

B. QoS aware load distribution

In the GEO/LEO hybrid satellite network where there is a non-negligible delay difference between ISL and ILL, we need to consider the influence of large ISLs in traffic diversion, i.e., dramatic increase of end-to-end delivery delays. The motivation behind this lies on the fact that it is not preferred to divert traffic originating from delay-sensitive applications such as VoIP to GEO satellites. Actually, in the proposed scheme, traffic is categorized into three classes, namely A, B, and C, and the direction of detouring is determined according to the traffic classes. Class A has the highest priority and the delay-sensitive interactive applications such as VoIP are involved. The traffic class A is never detoured in any situation because the increased detouring delay significantly degrades the QoS. On the other hand, class B's traffic consisting of relatively delay-robust applications such as real-time video streaming applications are delivered only through LEO satellites, i.e., traffic detouring is performed within the LEO layer. Class C represents best-effort traffic, which is allowed to be diverted to GEO satellites because of its robustness to long delays and delay changes. The identification of traffic classes can be easily implemented by adopting a labeling mechanism, e.g., Diffserv.

When a LEO satellite performs traffic detouring in the warning state, the traffic detouring ratio, η, is calculated from the following equation:

$$\eta = \frac{I_{\text{relay}} + I_{\text{self}} - \mu \cdot C_{\text{ISL}}}{I_{\text{relay}}}, \tag{1}$$

where I_{relay} and I_{self} indicate the rates of traffic coming from neighboring LEO satellites and mobile terminals within the coverage area, respectively, that are passing through the congested link having the capacity equal to C_{ISL}. Here, the traffic rate which needs to be detoured, r_d, can be expressed as the product of η and I_{relay}. When r_d is smaller than the traffic rate of class C, I^c_{relay}, a part of class C's traffic is detoured to the above GEO satellite in order to mitigate congestion in the LEO layer. On the other hand, if r_d takes a traffic volume between I^c_{relay} and the summation of I^c_{relay} and the traffic rate of class B, I^b_{relay}, not only the class C's traffic but also a part of the traffic belonging to class B are detoured; class B's traffic is diverted to neighboring LEO satellites to mitigate the delay growing due to the detouring. When r_d exceeds the summation of I^c_{relay} and I^b_{relay}, all of the traffic classified into class C or B are detoured while class A's traffic is never detoured. Fig. 2 shows an example where the detouring of all of the classes B's and C's traffic is required. To actually perform traffic detouring at the neighboring LEO satellites of the congested LEO

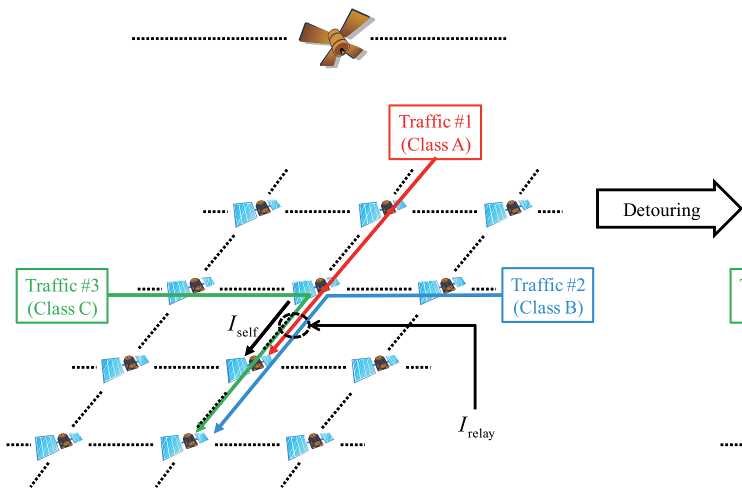


Fig. 2: A traffic detouring scenario.

TABLE II: Configuration of links.

Link	Parameter	Value
ISL in GEO layer	Bandwidth	25Mbps
	Queue size	200kB
ISL in LEO layer	Bandwidth	2.5Mbps
	Queue size	20kB
ILL	Bandwidth	25Mbps
	Queue size	200kB
GSL	Bandwidth	5Mbps
	Queue size	20kB

TABLE III: Distribution of traffic flows.

Source	Destination					
	NA	SA	Europe	Africa	Asia	Oceania
NA	60%	10%	15%	2%	10%	3%
SA	35%	40%	12%	2%	8%	3%
Europe	40%	5%	40%	2%	10%	3%
Africa	40%	2%	30%	20%	5%	3%
Asia	30%	2%	10%	2%	50%	6%
Oceania	40%	2%	10%	2%	12%	34%

NA: North America, SA: South America

satellite, the appropriate detouring ratios for each traffic class need to be notified to them. The detouring ratios for each traffic class for different conditions can be given as summarized in Table I. By properly detouring traffic according to its priority, the mitigation of network congestion and the QoS provision can be achieved. While we assume that the traffic rate of class A is always less than the bandwidth of ISL, ILL, and GSL anywhere, additional resource reservation and/or admission control mechanisms are required if the volume of requested class A's traffic exceeds any link capacity.

IV. Performance Validation

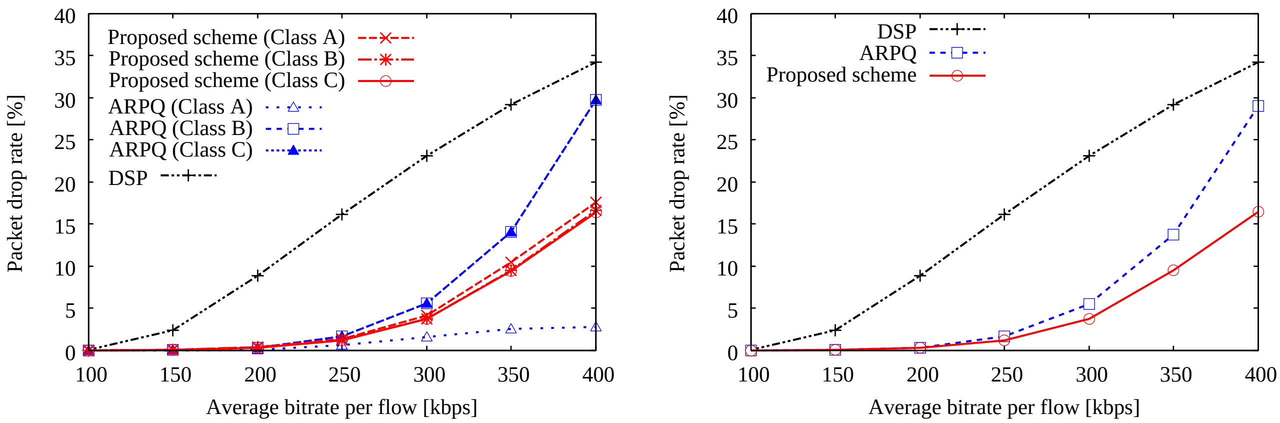
To validate the performance of the proposed scheme, we conducted extensive computer simulations by using Network Simulator version 2 (NS2) [26]. Here, we discuss the effectiveness of the proposed congestion prediction-based QoS-aware load balancing technique and review the results of our simulation experiments.

A. Network configuration and communication scenarios

We use a GEO/LEO hybrid satellite network with connectivity as depicted in Fig. 1. Three equally-positioned GEO satellites construct a ring constellation as the upper layer, while the Iridium constellation consisting of sixty-six LEO satellites is assumed as the lower layer. The bandwidth and buffer size of each link are set up according to Table II. The link bandwidth between GEO satellites and the bandwidth of ILL are sufficiently larger than those between LEO satellites because we assume that the GEO satellites have more capacity than the LEO satellites. The elevation mask of user terminals is set to 8.2 degrees by following the Iridium constellation's setup, and we accordingly set the elevation mask of LEO satellites for GEO satellites to the same value. User terminals and LEO satellites switch their connected LEO or GEO satellites when the currently used LEO or GEO satellites fall below their elevation masks, respectively.

Each user terminal generates ON/OFF traffic where the ON/OFF period follows Pareto distribution with the average ON/OFF durations equal to 200ms respectively. Simulation time is 6026.9 seconds which is equal to the rotation interval of the LEO satellites, and all the user terminals continue their communication during the simulation. One hundred user terminals are equally distributed to the top one hundred cities in the population. Half of the 200 traffic flows are established according to Table III which is determined by referring to [27], [28], and the other half of the traffic flows are randomly established. The traffic considered are User Datagram Protocol (UDP) streams and the packet size is equal to 1kB. The distribution of traffic classes A, B, and C, are 3%, 20%, and 77% according to [29], respectively.

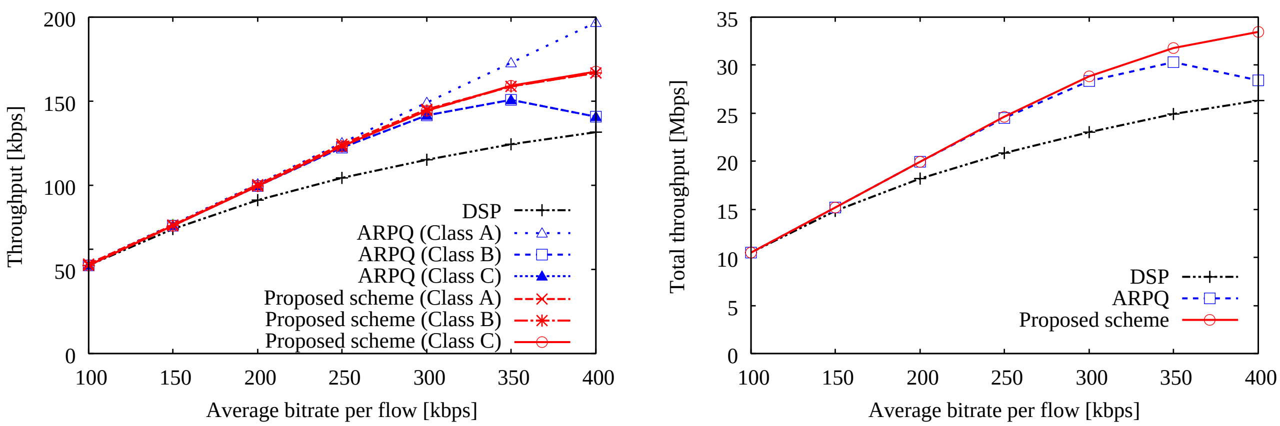
The parameters in the proposed scheme are set to as shown in Table IV. Since the proposed scheme is built on the Dijkstra Shortest Path (DSP) routing, where the summation of the propagation and queuing delays is utilized as a link cost, DSP



(a) Packet drops in each traffic class.

(b) Packet drops in the whole network.

Fig. 3: Packet drops experienced in DSR, ARPQ, and the proposed scheme in case of different traffic classes, and in the whole network.



(a) Throughput in each traffic class.

(b) Throughput in the whole network.

Fig. 4: Throughput experienced in DSR, ARPQ, and the proposed scheme in case of different traffic classes, and in the whole network.

TABLE IV: Parameters in the proposed scheme.

Threshold for starting detouring outside congestion area, λ	0.9
Threshold for starting detouring in a congestion area, μ	0.8
Threshold for finishing detouring, ν	0.5
Radius of congestion area	3000km
Measurement interval of traffic rate	20ms
Detouring ratio update interval	2s

is used in comparison. Also, ARPQ is used for comparison as a routing technique taking into account QoS provisioning. To fairly compare the three routing schemes including the proposed one, the interval of broadcasting of the information on queuing delay in each satellite is equally set to 3s, and the traffic rate measuring interval in ARPQ is the same as in the proposed scheme.

B. *Performance and discussion*

In performance evaluation, packet drops, throughput, and end-to-end delays are used as performance metrics. While the effectiveness of load balancing can be evaluated by reviewing packet drops and throughput, end-to-end delays demonstrate the QoS provisioning performance. The traffic bit rate of each flow, which is averaged over ON and OFF periods, is varied from 100kbps to 400kbps by 50kbps.

Fig. 3 presents the packet drop rates averaged in each routing scheme and in each traffic class. The highest drop rates in DSP imply that it is difficult to follow the dynamic changes in traffic volume by periodically recalculating all the shortest path information based on the queuing delay observed at each link due to the excessively long operational delay. In contrast, in the proposed scheme employing an intelligent congestion prediction mechanism over DSP, the packet drop rate is successfully decreased by half. Although ARPQ also succeeds in totally mitigating packet drops, there remains a significant difference between class A and other classes. This is because of the fact that ARPQ can mitigate the drops of class A's traffic, but it is unsure that the drop rates of classes B and C are consequently decreased by traffic detouring. Since ARPQ detours traffic of classes B and C at the constant ratio toward the LEO satellite having the minimum queue length regardless of the congested situations, the detouring ratio and the direction of detouring are not always appropriate.

Fig. 4a shows the total throughput of all the flows in each routing scheme and confirms the superior performance of the proposed scheme. The proposed scheme is able to achieve high throughput as a result of the reduction of packet drops by dynamically controlling traffic detouring based on the congestion prediction according to the traffic changes. On the other hand, it is clear from Fig. 4b which depicts the average throughput of each class that the throughput decrease in high traffic loads in ARPQ comes from the throughput reductions in classes B and C due to high drop rates because of the above mentioned drawbacks of traffic detouring algorithms.

Fig. 5 indicates end-to-end delays averaged in each class. Compared with DSP, the proposed scheme succeeds in decreasing the delay of class A by diverting the traffic of the other classes. The reason why class C's traffic has relatively longer delays while the delays of class B are just slightly increased by detouring is the difference of satellites used for traffic detouring, i.e., the class C's delay-robust traffic is detoured to GEO satellites. On the other hand, the delay of class A is increased as the traffic rate becomes larger in ARPQ where class A's traffic is detoured to GEO satellites when network congestion occurs. It is evident that the application of ARPQ, which is originally designed for MEO/LEO hybrid satellite networks, may not be adapted in the GEO/LEO hybrid satellite network because of the significant difference in link delays between GEO and MEO.

While the proposed scheme is designed for GEO/LEO hybrid satellite networks, the traffic detouring strategy can be adopted for other multi-layered satellite networks. However, in such usage, the congestion occurrence in ISLs at the upper layer needs to be addressed. For example, while comparing MEO/LEO and GEO/LEO combinations, the traffic passing through ISLs of the MEO layer is greater than that of GEO layer because each MEO satellite has a smaller coverage than that of a GEO satellite. In addition, the overhead of frequent handovers between upper and lower layers must be taken into account. This is one of the reasons why we focus on the GEO/LEO hybrid satellite networks.

V. Conclusion

Multi-layered satellite networks have attracted attention as a next generation satellite network in recent years. In this paper, we focused on the GEO/LEO hybrid satellite network, which is a typical example of two-layered satellite networks; the integration of GEO satellites having wide coverage areas at the upper layer and LEO satellites having small propagation delays at the lower layer. One of the notable advantages of GEO/LEO hybrid satellite networks is their ability to mitigate significant network congestion in each layer, especially in the LEO layer, that is caused by traffic convergence, which is

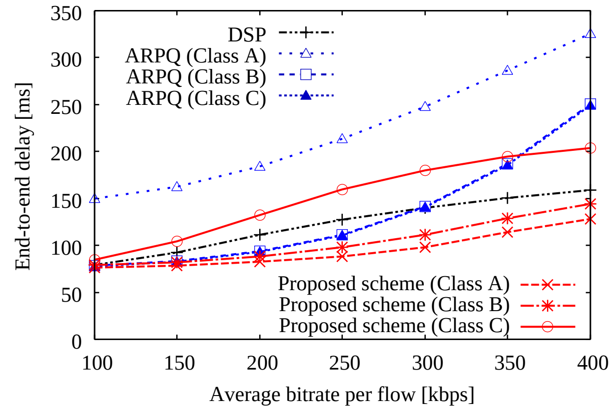


Fig. 5: End-to-end delays experienced in DSR, ARPQ, and the proposed scheme for each traffic class.

derived from geographically varying traffic demands. In other words, while the effect of traffic distribution in a single LEO satellites constellation is limited, detouring traffic to GEO satellites in GEO/LEO hybrid satellite networks can improve the load balancing performance. However, the QoS provisioning technique is necessary since there is a significant difference in propagation delays between GEO and LEO. Therefore, we proposed the QoS-aware load balancing scheme, which is developed through a congestion prediction-based traffic detouring method. Through extensive computer simulations, the superior performance of the proposed scheme is verified. Multi-layered satellite networks show a strong potential to solve the various problems remaining in the current existing satellite systems. Innovating the technologies to efficiently and fully utilize the advanced aspects of the multi-layered satellite networks will, indeed, lead to further challenges and open research issues.

References

[1] K. Naoto and S. Ryutaro, "Overview of the Wideband InterNetworking Engineering Test and Demonstration Satellite Project," Journal of the National Institute of Information and Communications Technology, vol. 54, no. 4, pp. 3–10, Dec. 2007.

[2] M. Agnew, "Ka-band Geosynchronous Satellites: 100Gbps and beyond," Proc. Ka and Broadband Communications, Earth Observation and Navigation Conference, Cagliari, Italy, Sep. 2009.

[3] H. Fenech, E. Lance, A. Tomatis, and M. Kalama, "The KA-SAT system," Proc. Ka and Broadband Communications, Earth Observation and Navigation Conference, Cagliari, Italy, Sep. 2009.

[4] W. Zhenyong, L. Dezhi, G. Qing, and L. Jianwei, "Hierarchical Satellite Network Design Based on Permanent Inter-Satellite-Links," Proc. Wireless Communications, Networking and Mobile Computing, Dalian, China, Oct. 2008.

[5] Z. Yuan, J. Zhang, and Z. Liu, "A Simplified Routing and Simulating Scheme for the LEO/MEO Two-Layered Satellite Network," Lecture Notes in Computer Science, Networking and Mobile Computing, vol. 3619, pp. 742–751, 2005.

[6] F. Alagöz, and Ö. Korcak, and A. Jamalipour, "Exploring The Routing Strategies in Next-Generation Satellite Networks," IEEE Wireless Communications, vol. 14, no. 3, pp. 79–88, Jun. 2007.

[7] K. Kimura, K. Inagaki, and Y. Karasawa, "Double-layered inclined orbit constellation for advanced satellite communications network," IEICE Transactions on Communications, vol. E80-B, no. 1, pp. 93–102, Jan. 1997.

[8] J. Hu, K.L. Yeung, and T. Li, "Routing and re-routing in a LEO/MEO two-tier mobile satellite communications system with inter-satellite links," Proc. IEEE International Conference on Communications 2000, New Orleans, LA, Jun. 2000, pp. 134–138.

[9] Iridium, http://www.iridium.com/

[10] I.F. Akyildiz, E. Ekici, and M.D. Bender, "MLSR: a novel routing algorithm for multilayered satellite IP networks," IEEE/ACM Transactions on Networking, vol. 10, no. 3, pp. 411–424, Jun. 2002.

[11] I.F. Akyildiz, E. Ekici, and G. Yue, "A distributed multicast routing scheme for multi-layered satellite IP networks," Wireless Networks, vol. 9, no. 5, pp. 535–544, Sep. 2003.

[12] F. Long, N. Xiong, A. V. Vasilakos, L. T. Yang, and F. Sun, " A sustainable heuristic QoS routing algorithm for pervasive multi-layered satellite wireless networks," Wireless Networks, vol. 16, no. 6, pp. 1657–1673, Oct. 2009.

[13] T. Wu, J. Hu, and S. Wu, "Convergent-reconstruction-based Inter-Layer Inter- Satellite Links Establishment Strategies for Two-tier LEO/MEO Satellite Networks," Proc. 2006 International Workshop on Satellite and Space Communication, Leganés, Spain, Sep. 2006, pp. 259–263.

[14] C. Chen and E. Ekici, "A routing protocol for hierarchical LEO/MEO satellite IP networks," Wireless Networks, vol. 11, no. 4, pp. 507–521, Jul. 2005.

[15] T. Taleb, Z. Md. Fadlullah, T. Takahashi, R. Wang, Y. Nemoto, and N. Kato, "Tailoring ELB for Multi-layered Satellite Networks," Proc. IEEE International Conference on Communications 2009, Dresden, Germany, Jun. 2009, pp. 5109–5113.

[16] Z. Zhao, J. Zhang, F. Sun, L. Li, and I. Chen, "Routing Strategy for Satellite-Based Networks with a Hierarchical Space Grid Architecture," Proc. The Fourth IEEE International Conference Information Communication Signal Processing, Singapore, Dec. 2003, pp. 1939–1943.

[17] Y. Zhou, F. Sun, and B. Zhang, "A novel QoS routing protocol for LEO and MEO satellite networks," International Journal of Satellite Communications and Networking, vol. 25, no. 6, pp. 603–617, Sep. 2007.

[18] S. Bayhan, G. Gür, and F. Alagöz, "Performance of delay-sensitive traffic in multi-layered satellite IP networks with on-board processing capability," International Journal of Communication Systems, vol. 20, no. 12, pp. 1367–1389, Jan. 2007.

[19] D. Kudoh, K. Kashibuchi, H. Nishiyama, and N. Kato, "Load balancing method based on congestion prediction for IP/LEO satellite networks," IEICE Transaction on Communications, vol. E92-B, no. 11, pp. 3326–3334, Nov. 2009.

[20] T. Taleb, D. Mashimo, A. Jamalipour, N. Kato, and Y. Nemoto, "Explicit load balancing technique for NGEO satellite IP networks with on-board processing capabilities," IEEE/ACM Transactions on Networking, vol. 17, no. 1, pp. 281–293, Feb. 2009.

[21] T. Taleb, N. kato, and Y. Nemoto, "Recent trends in IP/NGEO satellite communication systems: transport, routing, and mobility management concerns," IEEE Wireless Communications, vol. 12, no. 5, pp. 63–69, Oct. 2005.

[22] D. Yang and W. Liao, "On Multicast Routing Using Rectilinear Steiner Trees for LEO Satellite Networks," IEEE Transactions on Vehicular Technology, vol. 57, no. 4, pp. 2560–2569, Jul. 2008.

[23] J. Bai, X. Lu, Z. Lu, and W. Peng, "A Distributed Hierarchical Routing Protocol for Non-GEO Satellite Networks," Proc. The 2004 International Conference on Parallel Processing Workshops, Montreal, Quebec, Canada, Aug. 2004, pp. 148–154.

[24] Z. Li, D. Li, and G. Wang, "New dynamic routing algorithm based on MANET in the LEO/MEO satellite network," Frontiers of Electrical and Electronic Engineering in China, vol. 1, no. 2, pp. 169–170, 2006.

[25] E. Papapetrou and F.-N. Pavlidou, "QoS Handover Management in LEO/MEO Satellite Systems," Wireless Personal Communications, vol. 24, no. 2, pp. 189–204, 2003.

[26] "The Network Simulator - ns-2," http://www.isi.edu/nsnam/ns/

[27] M. Mahorcic, M. Werner, A. Sviqeli, and G. Kandus, "Adaptive routing for packet-oriented intersatellite link networks: performance in various traffic scenarios," IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 808–818, Oct. 2002.

[28] J. Hutcheson and M. Laurin, "Network flexibility of the Iridium global mobile satellite system," Proc. International Mobile Satellite Conference 1995, pp. 503–507, Ottawa, Canada, Jun. 1995.

[29] Cisco, "Cisco visual networking index: forecast and methodology, 2008-2013," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, Jun. 2009.

Hiroki Nishiyama received his M.S. and Ph.D. in Information Science from Tohoku University, Japan, in 2007 and 2008, respectively. He was a Research Fellow of the Japan Society for the Promotion of Science (JSPS) until finishing his Ph.D, when he then went on to become an Assistant Professor at the Graduate School of Information Sciences at Tohoku University. He has received Best Paper Awards from the IEEE Global Communications Conference 2010 (GLOBECOM 2010) as well as the 2009 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC 2009). He was also a recipient of the 2009 FUNAI Foundation's Research Incentive Award for Information Technology. His active areas of research include, traffic engineering, congestion control, satellite communications, ad hoc and sensor networks, and network security. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and an IEEE member.

Daigo Kudoh received the B.E. degree in information engineering and the M.S. degree in information science in 2008 and 2010, respectively, from Tohoku University, Sendai, Japan. His research interests were in the area of satellite networking, specifically traffic control. He is currently working in NTT East Corporation, Japan.

Nei Kato received his M.S. and Ph.D. Degrees in information engineering from Tohoku University, Japan, in 1988 and 1991, respectively. He joined the Computer Center of Tohoku University at 1991,and has been a full professor in the Graduate School of Information Sciences since 2003. He has been engaged in research on computer networking, wireless mobile communications, image processing, and neural networks. He has published more than 200 papers in journals and peer-reviewed conference proceedings. Dr. Kato currently serves as a chair of IEEE Satellite and Space Communications TC, and the chair of IEICE Satellite Communications TC, His awards include Minoru Ishida Foundation Research Encouragement Prize (2003), Distinguished Contributions to Satellite Communications Award from the IEEE Communications Society, Satellite and Space Communications Technical Committee (2005), the FUNAI information Science Award (2007), the TELCOM System Technology Award from Foundation for Electrical Communications Diffusion (2008), the IEICE Network System Research Award (2009), and many best paper awards including IEEE GLOBECOM . Besides his academic activities, he also serves on the expert committee of Telecommunications Council, Telecommunications Business Dispute Settlement Commission Special Commissioner, Ministry of Internal Affairs and Communications, Japan, and as the chairperson of ITU-R SG4, SG7, Japan.

Naoto Kadowaki received B.S. in Communications Engineering, Masters degree in Information Engineering and Ph.D. in Information Science from the University of Tohoku, Sendai, Japan in 1982, 1984 and 2010, respectively. From April 1984 to March 1986, he was with Mitsubishi Electric Corporation. He joined Communications Research Laboratory (CRL), currently renamed as National Institute of Information and Communications Technology (NICT) in 1986. He has been involved in research and development of high data rate satellite communication systems, mobile and personal satellite communications, computer networks, and communication protocols in NICT. From July 1990 to June 1991, he was a visiting researcher to AUSSAT that was reformed as OPTUS Communications, Sydney, Australia. After he was with Advanced Telecommunications Research (ATR) Institute International as the Head of Department of Autonomous Systems and the Head of Department of Smart Networks from July 2004 to December 2006, he was the Managing Director of Strategic Planning Department in NICT from January 2007 to June 2008. He is currently Executive Director of New Generation Wireless Communications Research Center and Director General of Yokosuka Research Laboratories in NICT. He is a member of the IEEE, AIAA and IEICE of Japan.