Journal of Communications and Information Networks
Vol.1, No.4, Dec. 2016
DOI: 10.11959/j.issn.2096-1081.2016.058

Review paper

A survey of routing techniques for satellite networks

QI Xiaogang[1], MA Jiulong[1], WU Dan[2], LIU Lifang[2], HU Shaolin[3]

1. School of Mathematics and Statistics, Xidian University, Xi'an 710126, China
2. School of Computer Science and Technology, Xidian University, Xi'an 710071, China
3. Key Laboratory of Spacecraft Faults Diagnosis and Maintance, Xi'an 710043, China

Abstract: Satellite networks have many advantages over traditional terrestrial networks. However, it is very difficult to design a satellite network with excellent performance. The paper briefly summarizes some existing satellite network routing technologies from the perspective of both single-layer and multilayer satellite constellations, and focuses on the main ideas, characteristics, and existing problems of these routing technologies. For single-layer satellite networks, two routing strategies are discussed, virtual node strategy and virtual topology strategy. Moreover, considering the deficiency of existing multilayer satellite network routing, we discuss the topic invulnerability. Finally, the challenges and problems faced by the satellite network are analyzed and the trend of future development is predicted.

Key words: satellite network, routing technology, virtual node, virtual topology, invulnerability

Citation: QI X G, MA J L, WU D, et al. A survey of routing techniques for satellite networks[J]. Journal of communications and information networks, 2016, 1(4): 66-85.

1 Introduction

With the rapid development of aerospace technology, scope of human activity has been extended to the space outside atmosphere. To enable the information to be effectively transmitted in space, space radio technology is being developed. With the development of the global network and information demand, the ground communication network has been unable to meet the growing demand for information acquisition and transmission. Satellite networks are getting an increasing amount of research attention. Satellite communication is an ideal long-distance communication technology, which not only can overcome the limitation of geographical conditions, but also can provide an inexpensive, constant, and reliable communication channel. The space information network is a network system for real-time acquisition, transmission, and spatial information processing, based on a space platform (such as synchronous satellite or medium, low orbit satellite, stratospheric balloon, and manned or unmanned aerial vehicle).

Manuscript received Nov. 26, 2016; accepted Dec. 1, 2016
This work is supported by the National Natural Science Foundation of China (Nos. 61572435, 61472305, 61473222), the Natural Science Foundation of Shaanxi Province (Nos. 2015JZ002, 2015JM6311), the Natural Science Foundation of Zhejiang Province (No. LZ16F020001), Programs Supported by Ningbo Natural Science Foundation (No. 2016A610035).

Satellite communication between two or more earth stations is, in essence, the use of an artificial earth satellite as a relay station to forward radio waves. The backbone of the space information network is generally composed of a multitude of satellites and its constellation, which can provide integrated communication services for various space missions. The satellite network is quite different from the traditional ground network. Significant transmission delay, transmission loss, and the dynamic change of the network's topological structure have brought great challenges to satellite network routing technology.

To use the satellite network, we must first solve the problem of inter-satellite routing, which has been the focus of research in network communication. With the help of GEO (Geostationary Earth Orbit) satellite relay, early satellite communication completes data forwarding between two points on the ground using Bent-Pipe. This form of data transmission is fixed, thus, there is no route of which to speak. In a satellite communications network, whose nodes are composed of a multitude of satellites, the optimal path can be selected according to a given link cost metric in the multiple paths between the source and destination satellites. The point-to-point routing problem in terrestrial networks has been solved, however, in a satellite network consisting of a large number of satellites, it is still a challenging problem because the satellite network has the following characteristics: limited on-board processing and storage capabilities, network topology with continuous high dynamic change, uneven distribution of data flow and high bit error rate, and long delay in the inter-satellite link.

The above characteristics of the satellite network make it impossible to directly apply ground network routing mechanisms, and we must study and design a new routing mechanism. In general, an efficient routing technique for a satellite network should have the following properties.

1) Adaptability of the dynamic change of network topology.

2) Invulnerability.

3) High efficiency.

4) Adaptability to changes in network traffic.

Routing problem is a fundamental problem in satellite networks. Currently, the research on satellite network routing is focused on solving the routing problem in the space segment. On-board routing achieves a certain optimal path from the source satellite to the destination satellite. This is part the point-to-point, and being the most basic routing problem in satellite networks, this is the basis for the realization of satellite network interconnection. The advantages and disadvantages of routing performance directly affects all aspects of the satellite network application, thus, it is necessary to find an effective solution.

2 Framework of satellite network routing

2.1 Satellite orbit

According to the angle between the satellite orbit plane and the Earth's equatorial plane, the satellite orbit can be classified as an equatorial orbit, polar orbit or inclined orbit. According to ground to satellite distance, there is GEO (Geostationary Earth Orbit), MEO (Middle Earth Orbit), and LEO (Low Earth Orbit). A GEO satellite is stationary, relative to the earth, and has some defects including orbit uniqueness, poor communication quality in high latitude regions of the earth, and interference from solar radiation when the satellite is in a straight line between the earth and the sun. Additionally, it is difficult to avoid the longer signal transmission delay and larger energy loss. On the contrary, LEO satellites have advantages of low transmission loss and shorter communication delay, while the link handover between satellites is frequent because of continuous high-speed satellites motion. Different from GEO and LEO, the MEO satellite system has the characteristics of both GEO and LEO, which is widely used in multilayer satellite networks.

2.2 Topology of satellite networks

In an LEO satellite network, $N_L \times M_L$ satellites are deployed in N_L orbital planes, each of which has M_L satellites that are uniformly distributed. For a polar orbit satellite network, the ascending nodes of the orbital planes have a fixed plane offset $\Delta\Omega=\pi/N_L$ along the equator in the range of π arcs, while for a walker constellation network, the values are $\Delta\Omega=2\pi/N_L$ and 2π, respectively. Each satellite node is usually assigned four satellite links: two intra-plane links between the adjacent satellites located in the same orbit plane and two inter-plane links between the neighboring satellites located in the left and right orbital plane. These two types of ISLs (Inter-Satellite Links) are known as intra-plane ISLs and inter-plane ISLs. It is noted that the satellites along the counter-rotating seam have only one inter-plane link because cross-seam ISLs are not used due to link acquisition and special antenna requirements. Moreover, intra-plane ISLs exist in the whole satellite orbital period, yet, inter-plane ISLs are not present when satellites are located above a given latitude threshold. Those open and closed behaviors of inter-plane ISLs may result in frequent changes of the LEO network topologys. A LEO network can be represented as α:h: $N_L \times M_L/M_L/F$, where α, h and F denote the inclination of the orbital plane, satellite orbit altitude, and phase factor, respectively. The comparison of structures of the walker constellation and polar orbit constellation is shown in Fig.1.

The typical representative of the multilayer satellite network is the TLSN (Three-Layer Satellite Network), which is composed of LEO, MEO, and GEO satellites. In the network, upper satellites can cover lower satellites. Satellite networks constitute an independent AS (Automatic System). The ground gateway is directly connected with the visible satellite, which is responsible for the address conversion and communication between the satellite network and ground network. The structure of a multi-layer satellite network is explained in Fig.2. Here, the LEO layer may have two types of structure, which are both illustrated in detail in Fig.1. In other words, Fig.1 is a detailed description of an LEO layer in the multilayer satellite network.

In multilayer satellite networks, the network

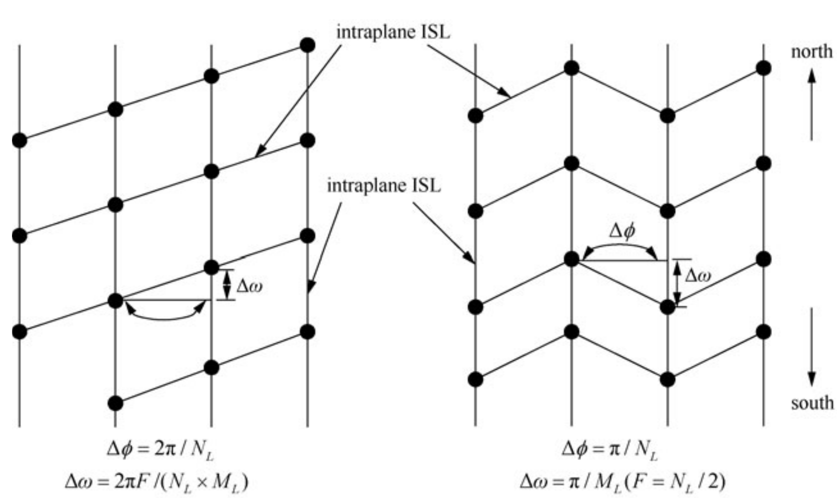


Figure 1 Comparison of two kinds of low orbit satellite network topology

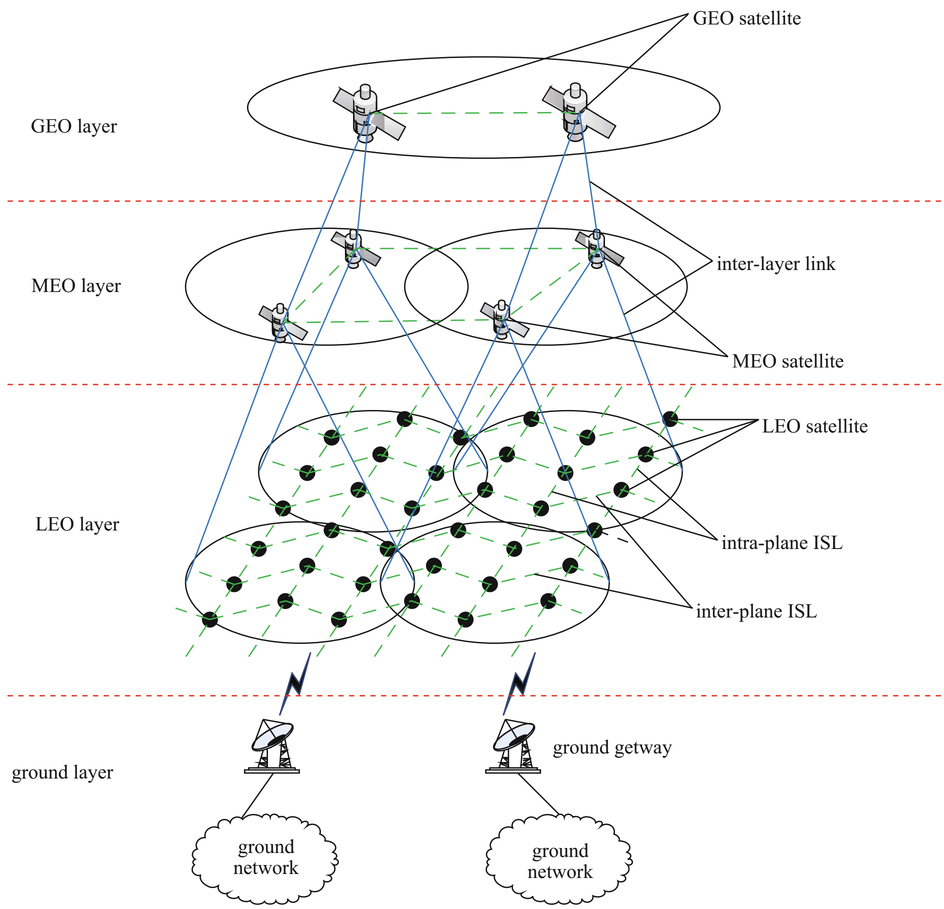


Figure 2 Structure of multi-layer satellite network topology

topology varies more frequently because the relative high speed motion between LEO and MEO satellites or GEO satellites is more frequent, except for the periodic ISL handover in an LEO satellite constellation. In this case, the network topology changes are affected by many factors, including constellation type, access strategy of ISLs, and initial relative positions of LEO, MEO and GEO satellite constellations.

2.3 Links of satellite networks

The satellite network consists of two types of full duplex links. Communication in the same layer is realized through ISLs, while different layers of the satellites communicate through IOLs (Inter-Obit Links). The ground gateway, which is covered by a LEO satellite, is connected to the LEO satellite via a UDL (User Data Link). A satellite can maintain UDLs to multiple terrestrial gateways. Similarly, a terrestrial gateway may be directly connected to multiple satellites in different layers.

2.4 Route of satellite networks

The main research goal of satellite network technology is to develop and design efficient routing algorithms and protocols for satellite network topology, providing a reliable data transmission path for satellite network users. In a satellite network composed of a multitude of satellites, the optimal path which is in accordance with the link metrics or QoS (Quality of Service) requirements is required to select a path from the source to the destination. Satellite network routing technology plays an important role in determining the efficiency and reliability of the whole network system.

According to the structure of the satellite network system, the network routing can be divided into three parts.

1) Boundary routing. The satellite network is regarded as an important part of NGI (Next Generation Internet). The integration of terrestrial and satellite networks requires the development of a satellite-to-ground boundary routing protocol between satellite and terrestrial networks. This routing protocol enables seamless interoperability between these networks.

2) Access routing. Access routing is responsible for satellite access for the ground mobile users and base station. Generally, when selecting the access satellite, it is necessary to consider the survival time, delay, and signal strength of the UDL.

3) On-board routing (routing in space segment). When a data packet is uploaded to the portal satellite, on-board routing is responsible for finding one or several paths from the source node to the destination node to find one or several paths to meet the QoS requirements. On-board routing is the most important and difficult part of satellite network routing technology, and at present, most research on this technology is focused on this part. To find a path that meets certain requirements from the source satellite to the destination satellite, we must first solve the problem of the satellite network topology. The technology of solving the satellite network topology change, whose aim is to find the correct path, is called the satellite network routing strategy.

3 Routing techniques for single-layer satellite network

In a single-layer satellite network, routing primarily includes the routing algorithm based on the virtual topology, and routing strategy based on the virtual node.

3.1 Routing based on virtual topology

The principle of the virtual topology strategy is to divide the system period of the satellite network into n time slices denoted as $[t_0, t_1],[t_1, t_2],\cdots,[t_1, t_n]$, during normal conditions. The topology in each time slice can be represented by one snap-shot. In this case, connection or disconnection of links occurs only at discrete time points $t_1, t_2, \cdots, t_n$. In the time interval $[t_i, t_{i+1})$, the dynamic topology of the satellite network can be modeled as a fixed topology. In each time slice, the cost change of each link is small enough so that it can be considered unchanged. Under the virtual topology strategy, the routing algorithm only needs to calculate the routing during the current time.

However, this method also produces a problem that significant storage space is required to store routing information because the system cycle is divided into a large number of time slices. Moreover, it has poor adaptability to the change of traffic flow, link congestion and real-time transmission.

3.1.1 DV-DVTR algorithm

Discrete-time DV-DVTR (Dynamic Virtual Topology Routing) is a connection-oriented routing algorithm based on the ATM mechanism, which is also the first routing algorithm based on virtual topology[1]. In the DV-DVTR algorithm, the system period of the satellite network is divided into time slices. In each time slice, satellite network topology structure is considered to be fixed. An example of dividing time slices in DV-DVTR is shown in Fig.3.

Therefore, the computation of satellite network routing is transformed into the routing calculation of n static virtual topologies with VP (Virtual Paths), and the optimization goal of the algorithm

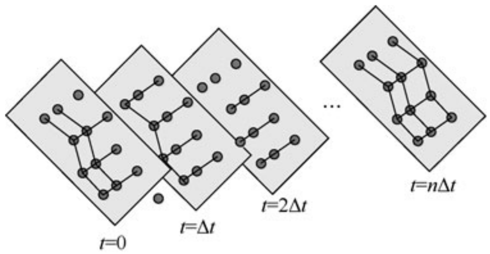


Figure 3 An example of dividing time slices in DV-DVTR

is to reduce the virtual path handover caused by the topology handover among different time slices. The DV-DVTR algorithm is divided into two stages: DT-VTS (Discrete Time Virtual Topology Setup) and DT-PSS (Discrete Time Path Sequence Selection). In the DT-VTS phase, the DT-DVTR algorithm discretizes the LEO satellite network topology, changing with time such that the LEO satellite network topology, in a small enough time interval is considered to be fixed. After discretization process, the dynamic changes of the LEO satellite network topology are represented as a series of static topology structures. In the DT-PSS phase, routing calculation can be achieved through the typical Dijkstra algorithm (Shortest Path Algorithm). The optimized route is uploaded to the satellite, which is modified at every time point.

DV-DVTR first proposed the idea that the frequent change of the satellite network topology can be discretized into a series of static topologies. However, it is not able to solve the problem of rerouting caused by link handover.

3.1.2 FSA algorithm

From the grouped link perspective, satellite network topology and network flow, FSA (Finite State Automata) routing algorithm aims to solve the routing problem in LEO satellite networks, whose optimization objective is to maximize ISL utilization[2]. Similar to the DV-DVTR algorithm, satellite network topology in the FSA algorithm is also divided into a finite number of time slices, and the algorithm is a connection-oriented routing algorithm. However, the difference is that the FSA algorithm takes the network topology of each time slice as a “state”, and the dynamic topology of the satellite network is modeled as a finite state machine. The satellite network system cycle is divided into a number of states that depend on the ISL connection state data, such that the network in each state has a fixed topology. Due to the periodicity of the satellite topology and limited number of states, the optimal path can be established in each fixed topology.

The FSA algorithm combines the ISL link assignment problem with the routing problem, and uses the dynamic ISL link allocation technique to solve the routing problem of the LEO satellite network. Thus, the satellite network routing problem is transformed into an optimization problem of the virtual network link assignment topology in a finite state machine. An overview of the proposed FSA-based link assignment scheme is shown in Fig.4. The link assignment for each state is determined from the visibility matrix and traffic requirements. This algorithm calculates a joint optimal solution for link and routing assignment using an iterative method, which provides a global optimal link allocation table, and lists the optimal communication paths in each state for each node.

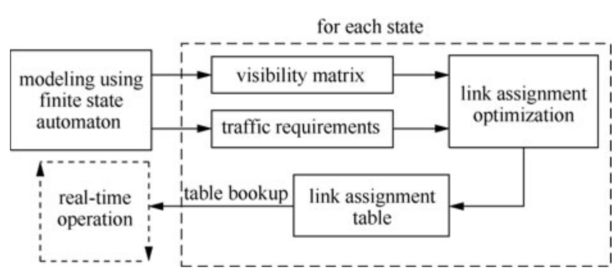


Figure 4 Overview of the FSA-based link assignment scheme

The advantage of an FSA algorithm lies in the optimization of network resources. The disadvantages are: first, the shortest path principle is not the path choice standard; second, dynamic link allocation technology is difficult for an LEO satellite network; and finally, the computational complexity of the routing algorithm is too high.

The DV-DVTR and FSA algorithms are both off-line routing algorithms, and are typical representatives of virtual topology routing algorithms in a connection-oriented satellite network. Both divide the satellite constellation system period into a finite number and equal length of time slices, and in each time slice, use a certain method to reduce the ISL handover and improve the ISL utilization rate. Unfortunately, in these two algorithms, all the routing calculations are based on the law of motion of the satellites in the off-line state, such that there are the following deficiencies: 1) Failing to adjust the routing scheme, according to the statistical characteristics of traffic in the current network, to achieve the optimal assignment of resources in the whole network and to guarantee the quality of service. 2)Network survivability is not good, because the routing tables of satellite nodes are calculated in advance by an off-line routing algorithm. When a certain satellite link fails, the algorithm can not dynamically observe the current state of the network. 3) The delay performance of the calculated path cannot be guaranteed. 4) Only when the connection request is established, can the ISL handover process be optimized, and the rerouting problem caused by the handover is not effectively solved.

3.1.3 CEMR routing algorithm

A CEMR (Compact Explicit Multi-path Routing) algorithm is the first algorithm to study the multi-path routing problem of a mobile satellite network[3]. The basic idea of the CEMR algorithm is to encode the path, using a compact path identifier, such as PathID. PathID is a global representation of the sequence that is composed of the inter-satellite link interface identifier. In this method, the intermediate nodes through the path can directly transmit packets according to the information described by PathID. Moreover, the CEMR algorithm contains a PathID verification algorithm that ensures the accuracy of PathID as well as the consistency, at any moment in time.

Based on the concept of dynamic virtual topology, the CEMR algorithm is studied and the satellite network is modeled, on a series of discrete time snapshots over a system cycle. When the time interval is small enough, the cost of each satellite link is considered to be constant during the time interval. Moreover, the time interval shift is defined as the behavior of the network state transition from the current time interval to the next time interval. CEMR multi-path routing consists of three parts: route discovery, route maintenance, and traffic assignment. The CEMR routing algorithm takes two parameters that have an important influence on the routing performance as a measure of the routing cost. The time delay is considered as the sum of propagation delay and queuing delay. The CEMR algorithm uses the global path to identify the idea of PathID, which realizes the explicit multi-path routing in the LEO satellite network. Meanwhile, the CEMR algorithm can reduce the delay of the LEO satellite network, increase the throughput, and balance the load, because the traffic is divided towards two or more feasible adjacent paths.

3.1.4 ELB routing algorithm

Although the data packet queuing delay is considered in the CEMR algorithm for a given satellite, it is not possible to predict the probability of dropping the current data packets for the next hop.

The ELB (Explicit Load Balancing) routing algorithm considers these factors, achieving the load balance of all satellites, and avoids data packet loss as far as possible. This is a type of routing protocol in which explicit communication congestion information between neighboring satellites is exchanged, and the protocol requests that these neighboring satellites to reduce the data forwarding rate[4]. In response, the nearby satellites reduce their data transmission and find other alternative paths of satellites that are not included. In this method, a better flow distribution between satellites is determined, effectively avoiding the congestion and resulting data loss, such that the proposed scheme is called the load balancing mechanism.

In ELB, the satellite uses three parameters to represent the congestion status and data transmission reduction ratio, respectively, for the two queue rate thresholds, and transmission rate ratio. Based on the simple mathematical model, ELB proposes a method for the systems parameters configuration. An important characteristic of the ELB routing algorithm is the traffic flow cascade problem, which may occur during the traffic detour. In fact, in order to avoid satellite congestion, the neighboring satellite is not included in the backup path that is used to forward part of the data. At this point, the system should ensure that the reselection of the path transmission does not affect the transmission of the previous destination path.

3.1.5 PAR routing algorithm

There may be more than one shortest path between two satellites in the nonsynchronous satellite constellation. PAR (Priority-based Adaptive Routing) takes into account the link utilization and historical information in order to achieve a uniform load distribution[5]. PAR sets the distributed path to the destination, using a priority mechanism that relies on the historical utilization and cache information of the ISLs. In order to avoid unnecessary data traffic and obtain higher link utilization, the improved PAR algorithm is formed.

The satellite network is usually modeled as a grid network. In a source and destination, there is probably more than one shortest path. When a satellite receives a data packet, it will search for the destination node. If the destination node is just in the same latitude or longitude, then there will be only one shortest path, otherwise, there will be two possibilities. In this case, the direction choice depends on the routing algorithm. In order to achieve this goal, four different adaptive shortest path algorithms are defined, the details of which are found in Ref.[5].

PAR and ePAR (enhanced PAR) is the two routing algorithms for NGEO (Non-geosynchronous Orbit) satellite networks. The PAR algorithm selects the output link when the packet of each hop node is to be sent. The selection principle is based on the priority mechanism, which is biased towards the use of fewer links. The ePAR algorithm delivers some improvements in providing the same links for data packets with the same source and destination. Although the PAR algorithm is designed to set the destination with a distributed minimum hop path, it can be extended to handle the ISL link length and further minimize the end-to-end delay. Therefore, it should be considered to merge the ePAR and DVTR to conform to the periodic characteristics of the topological changes.

3.1.6 DDRA routing algorithm

The DDRA (Dynamic Detection Routing Algorithm) is a typical routing algorithm based on a virtual topology strategy, which is an improvement of the time virtual routing scheme[6].

The idea is to determine whether a link is normal by the ACK (Acknowledge) confirmation and periodic detection of the number of data packets in the satellite outgoing queue. Moreover, it can detect a link's unexpected situation and make appropriate adjustments in time. The primary algorithm steps are shown in Fig.5.

The primary advantage of the algorithm is the

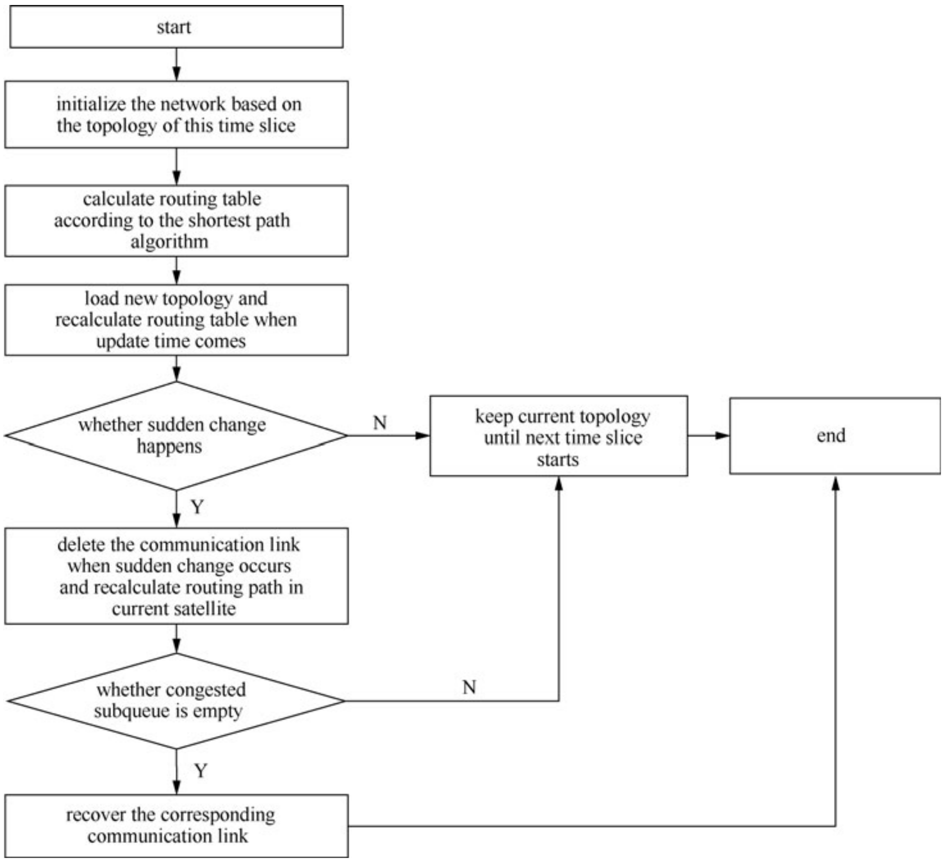


Figure 5 DDRA program

characteristics of low time delay. When an unexpected network condition occurs, the algorithm can promptly adapt to this emergency situation by sending queue detection and returning ACK, thus, overcoming the poor adaptability link congestion and link failure in the time virtualization schedule.

Compared with the method of virtual nodes, virtual topology strategy has good expansibility, which is suitable for establishing a multilayer satellite network. Routing tasks can be created off-line by the ground gateway, but also can be created by a real-time satellite located in the upper level. In this method, the load of LEO satellites will be greatly reduced. Moreover, this method and the way is also more flexible. However, in the virtual topology, a large number of time slices may be generated, and the computation of the routing algorithm will increase and become more complicated.

3.2 Routing algorithm based on virtual node

The virtual node strategy is formally proposed in Ref.[7]. It can be described as follows: The geographical location of a satellite S is given by $[lon_S, lat_S]$, indicating the longitude and latitude of the location of satellite S, respectively. Assume that the entire earth surface is covered by logical satellite locations of satellites, which are filled by the nearest satellite. Each logical location is embodied at any given time by a certain physical satellite. As a satellite disappears over the horizon, its corresponding position becomes represented by the next satellite passing overhead and the state information (such as routing table entries or channel allocation information) is transferred to it. Hence, the identity of satellite S is not permanently coupled with its logical location, which is taken over by the successor satellite in the same plane. The routing is performed basically by considering these logical locations as hops and a fix topology is formed. A routing decision is made based on this fixed virtual topology, and consequently, the network layer is isolated from the satellite constellation dynamics. By this way, it is not necessary to consider the satellite movements.

3.2.1 LZDR algorithm

The LZDR (Localized Zone Distributed Routing) is a type of connection-oriented satellite network routing algorithm based on a virtual node strategy. T.H. Chan modeled the polar orbit satellite network as a regular MSN network and proposed localized zone distributed routing[8]. The LZDR does not divide the routing zone according to a single virtual node, but instead merges the adjacent virtual nodes into one zone. The calculation includes intra-zone and inter-zone routing. For intra-zone routing, LZDR defines a virtual node in a zone for the routing controller in the zone and determines the packet transmission between zones in accordance with the minimum hops metric using the binary system in MSN. For inter-zone routing, the nodes in the zone need to exchange information with each other and the packet should be transmitted through the shortest path according to the time delay.

The LZDR algorithm attempts to reduce the communication overhead in the routing process by hierarchical routing. However, the algorithm only describes the method for calculating inter-zone routing according to the minimum hop path. It fails to point out the division of the regional boundary and network load information is not exchanged between zones, thus, it cannot guarantee the optimality of the inter-zone routing. Moreover, the LZDR algorithm only applies to the polar orbit of the LEO satellite network, and does not apply to the inclined orbit satellite constellation.

3.2.2 DRA algorithm

The DRA (Datagram Routing Algorithm) is a typical virtual node routing algorithm[7,9]. In DRA, it is assumed that the entire earth surface is covered by logical locations of satellites. These logical locations can not move and are always filled by the nearest satellite. Thus, the identity of every satellite is not permanently coupled with its own logical location, which will be taken over by the successor satellite in the same orbit. The logical location of a satellite S is given by $<p, s>$ where p is the orbit number and s is the satellite number. The routing is basically performed basically by considering these logical locations as hops. In this manner, it is unnecessary to consider the movement of the satellite.

For any set of parameters describing the satellite network, the DRA is likely to find the paths with minimum propagation delay between all source destination pairs that are outside the polar regions. This routing algorithm produces the paths in different ways, i.e., the satellites process every incoming packet independently, assuring that the packets will be forwarded according to minimum propagation delay routing as the result of their collective behavior. The next hop on the path is determined in three phases. In the direction estimation phase, possible next hops on the minimum-hop path are determined by assuming that all the ISLs have equal length. With this assumption, a minimum-hop path also becomes a minimum propagation delay path. However, the direction enhancement phase that the inter-plane links have different lengths and the decision made in the first phase needs be refined regarding the next hop. The DRA has a congestion processing function, which can find and avoid congestion by monitoring the link queue buffer occupancy. Once the buffer queue reaches a threshold value, it will back up the warning signal, and notify the source satellite to reroute, in order to avoid the occurrence of network congestion.

The virtual node strategy was proposed in DRA for the first time. The design and implementation of the algorithm is relatively simple. Because the satellite does not exchange any topological information, the algorithm does not bring any extra communication overhead. The constellation topology information is derived from the precalculation of the ground base station that is bound to the constellation. When a satellite node or link fails, performance of the algorithm will be drastically reduced, and the network robustness will be poor. The DRA algorithm takes a relatively simple and symmetrical LEO polar orbit constellation as the application model, which makes it necessary to improve the DRA if it is to be extended to the complex model. Although the DRA algorithm has many defects, its concept regarding the virtual node is influential, providing inspiration for research on the satellite network routing. On this foundation, a variety of satellite network routing algorithms, based on the virtual node, were presented.

3.2.3 LCRA algorithm

The LCPR (Low-complexity Probabilistic Routing Algorithm) is an improvement of the DRA routing algorithm, and is used to solve the problem of next hop selection when a satellite receives a packet[10,11]. By using the position information of the source node and destination node, the distributed computing method is used to obtain the optimal path. The specific routing algorithm is given in Ref.[10] and the queue delay is considered in Ref.[11].

Each node can dynamically select the next hop by informing its neighbors of the congestion state. Thus, the average queue time delay is reduced, and the packet loss rate is lower. Simulation results show that LCRA has better network performance than both the Dijkstra algorithms and DRA algorithm. The process of the algorithm is shown in Fig.6.

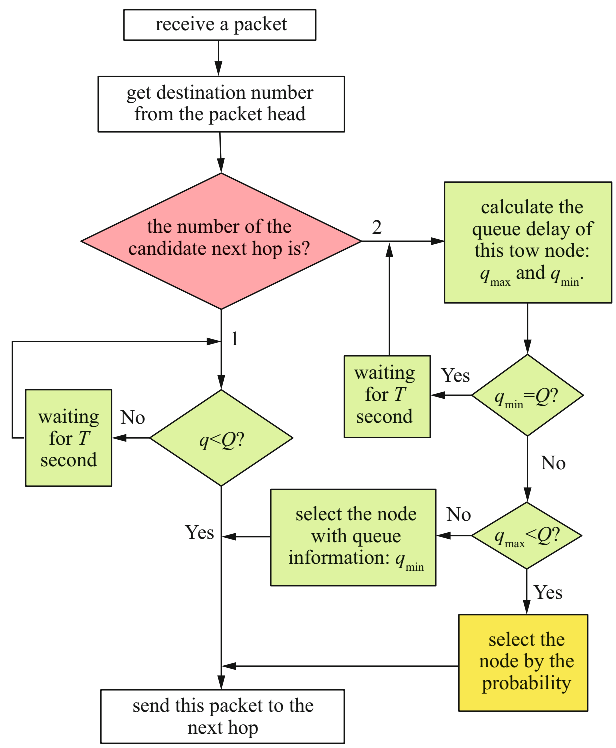


Figure 6 LCRA algorithm

3.2.4 DODR algorithm

In Ref.[12], the problem of repairing the network when there is a failed node or link in the LEO satellite network is discussed. DODR (Destruction-resistant On-Demand Routing) can be viewed as a variant of the on-demand distance vector routing algorithm. The prime goal of the LACR protocol is to minimize end-to-end delay and efficiently route the packet in a scenario where a failed link or node exists. Its network topology can be regarded as a mesh grid, as shown in Fig.7.

In the topology above, there may exist two kinds of link failure as shown in Fig.8. DODR protocol, whose primary process is described in the Fig.9, can solve this problem perfectly.

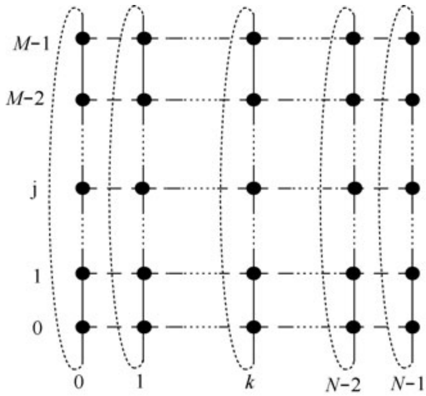


Figure 7 Network topology

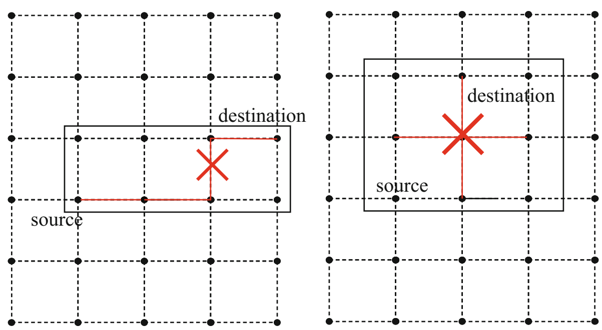


Figure 8 One link and four link failure caused by node failure

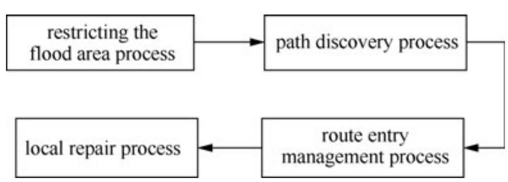


Figure 9 DODR routing process

4 Multilayer satellite network routing technology

A single-layer satellite network cannot meet the high survivability and efficiency of the satellite network. Moreover, damage resistance is poor. The concept of the multilayer satellite network was created to resolve this problem. The multilayer satellite network arranges the satellites in a two or three-layer orbit plane at the same time, using IOL to build up a three-dimensional network. Compared with the single LEO satellite network, a multilayer satellite network has higher spatial frequency spectrum utilization, flexible networking, and strong survivability. It is able to achieve the advantages of balancing various orbital satellite constellations, become an ideal method for future development of the satellite network. Algorithms generally adopt the master-slave mode, with MEO as the back-bone, and LEO for the access satellite.

4.1 MLSR algorithm

Based on the structure of a three-layer network composed of LEO, MEO, and GEO satellites, a MLSR (Multilayer Satellite Routing) algorithm was proposed in Ref.[13], where the SGGM (Satellite Group and Group Management) routing strategy was used for the first time. The MLSR algorithm is based on the IP multilayer satellite network routing algorithm, which is suitable for the three layer satellite network composed of the GEO, MEO, and LEO satellites. There are links between the satellites of each layer, and links between layers, such that a strong interconnection network structure is formed. The purpose of this algorithm is to reduce the computational complexity and communication overhead of the network topology.

The satellite network is organized hierarchically. In this hierarchy, satellites are grouped and their management is a given satellite in the upper layer. The hierarchical organization is used for routing table calculations. The data packets are forwarded independent of this hierarchy. The link delay information is regularly reported to the upper satellite.

In the MLSR algorithm, satellites are grouped according to the coverage area of satellites in the upper layer, implementing a hierarchical network topology for the collection of information. The routing table is periodically calculated by the upper satellites as the lower satellites are subject to change of the relationship between the group members. The routing metric is only the link transmission delay and the routing table is updated only when the group membership changes, i.e., the layer-to-layer coverage changes only when the update cycle arrives. Moreover, the algorithm lacks the ability to adapt to the sudden change of traffic in the network.

4.2 SGRP algorithm

The SGRP (Satellite Grouping and Routing Protocol) was proposed to further develop the concept of satellite grouping, suitable for the double layer satellite network composed of MEO satellites and LEO satellites[14]. In order to reduce the overhead of the system, it is based on a satellite grouping method, with calculated end to end delay, and consideration of link congestion.

The primary idea of the SGGR routing algorithm is to transmit data packets by the shortest delay path. The SGRP assigns the calculation of an LEO satellite's routing table to the MEO satellites. The low orbit satellite is grouped according to the coverage area of the middle orbit satellite of each snapshot period. The corresponding MEO satellite, which covers a group of low earth orbit satellites, becomes the group management. The group management is responsible for the collection and exchange of the link delay information of the low orbit satellite layer, and the calculation of the routing table for the low orbit satellite group members.

The detailed design of the SGRP includes three phases: delay report from the LEO satellite to the MEO layer, delay exchange in the MEO layer, and routing table calculation. Additionally, the SGRP routing algorithm provides a mechanism to solve congestion and satellite failure to avoid packet loss. The SGRA routing process is shown in Fig.10.

The SGRP routing algorithm enables the individual satellite network constellations to cooperate with each other, such that the calculation of the routing table is transferred to the middle orbit satellite, and the power consumption of the low orbit and middle orbit satellite is effectively assigned. In addition to the management functions, the satellite can be used for other purposes, such as the routing and forwarding of data packets, etc. Simulation results show that the end to end delay of the SGRP routing algorithm does not increase rapidly with the packet transmission rate, and shows little difference from the Bellman shortest path algorithm. The SGRP algorithm performance is similar to the Bellman shortest path algorithm performance in regards to satellite failure and link congestion.

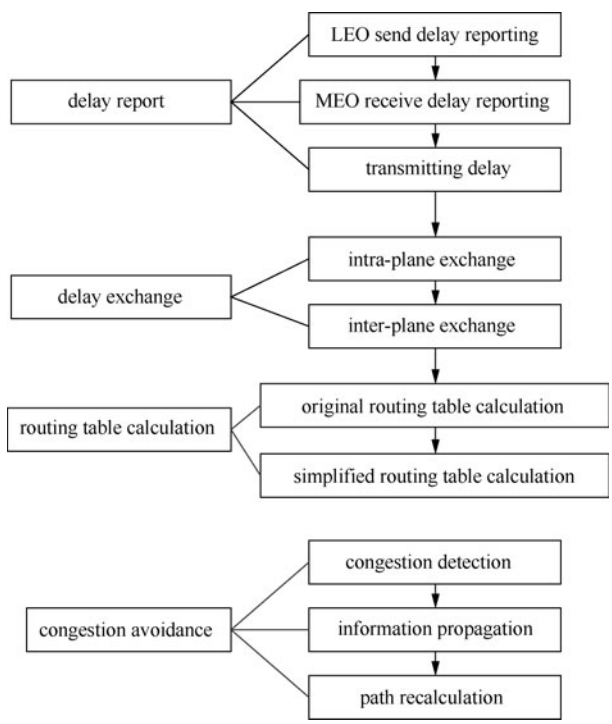


Figure 10 SGRA routing process

4.3 HSRP algorithm

HSRP (Hierarchical Satellite Routing Protocol) is appropriate for a SOS (Satellite Over Satellite) network for LDD (Long Distance Dependent) transmission and is a QoS-adaptive and hierarchical routing protocol for multimedia data[15].

SOS consists of a lower level constellation, and one or several constellations that are relatively higher than the low one. SOS is a combined network with multilayer constellations. These include a constellation with lower orbit altitude, and constellations with higher orbit altitude. Satellites in the lower layer are clustered to form the second satellite layer, which, in turn, are clustered into the third satellite layer, etc., up to the last. The altitude of the orbit and number of layers can be varied in the SOS network, based on the performance and type of services that are provided.

The HSRP is the first step toward achieving user-to-user delay guarantees. It identifies paths that meet delay constraints, and selects the one that leads to high overall resource efficiency. The primary flow of HSRP is shown in Fig.11.

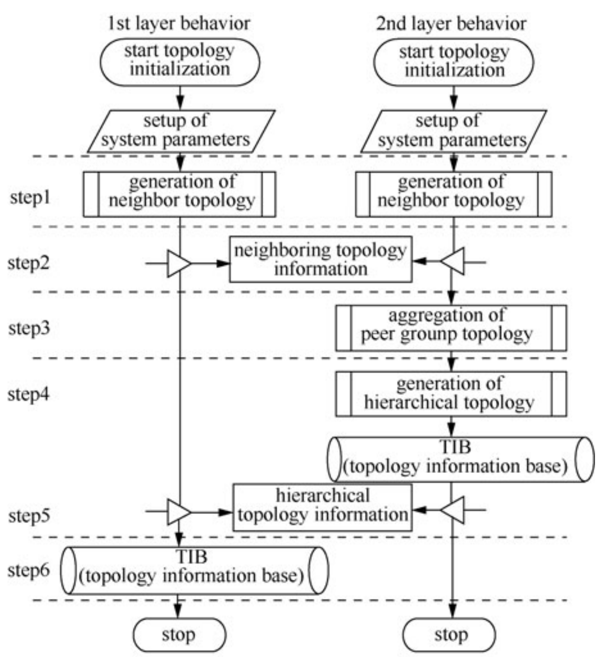


Figure 11 The primary flow of HSRP

4.4 SARA algorithm

The SARA utilizes the connectivity rules of ISL to partition network topology and compute the routing table[16]. GEO satellites can recount the routing table for the failure region when nodes fail, and this makes the network autonomous.

When satellite failure occurs, a new routing table should be computed with the algorithm. The steps are as follows:

First, the F-SA (Failure Satellite Advertisement) is formed by the satellite in the adjacent orbit, and then, the satellite forming the F-SA begins to transmit the F-SA until it reaches the manager satellite GEO. Then, the manager GEO satellite receives the F-SA, and recalculates a new routing table for the sub-failed region in every time slice. The RUI (Routing Update Information) is then formed and distributed. The manager GEO satellite transmits the RUI to the adjacent manager satellite by ISL, and LEO satellite by IOL. The steps of the algorithm are given in Fig.12.

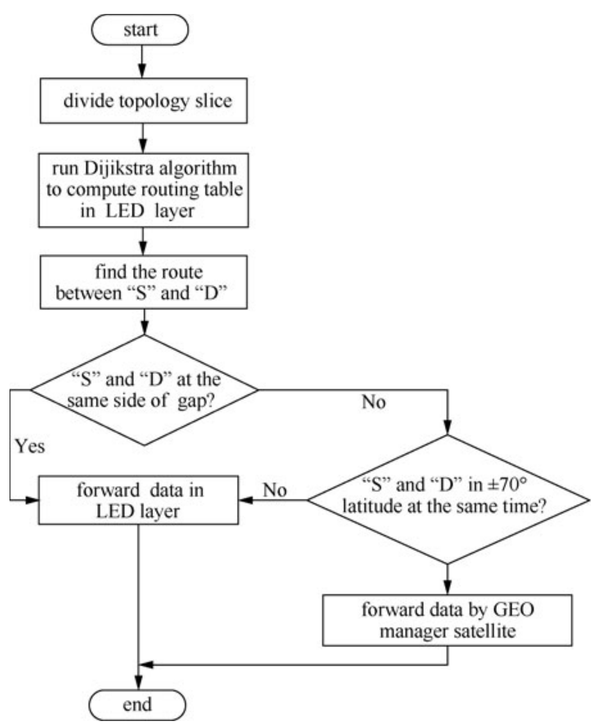


Figure 12 The primary algorithmic procedures of SARA

4.5 Routing algorithm based on data-driven

In Ref.[17], a double layer satellite network consisting of GEO and LEO satellites was established. The network itself has a great advantage in routing design. First, an LEO layer uses virtual a node strategy to deal with the problem of satellite movement, and GEO is static relative to ground. In this way, the topological dynamics of the satellite network are completely shielded, which is very helpful for the routing design. Second, the network takes into account the advantages of the GEO and LEO network. This simple network model is shown Fig.13.

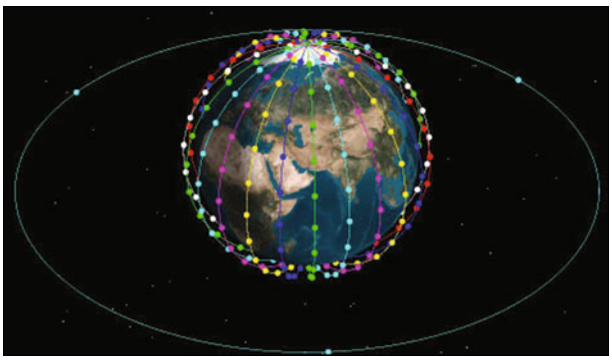


Figure 13 LEO/MEO joint constellation

In this model, a data-based inter-layer routing mechanism is presented. Using this method, the network traffic is balanced and the delay caused by queuing is decreased by classifying the packet into two types and using modified short-hop paths based on region partitions. On the basis of satellite grouping, the network increases the data driving mechanism, where the concept of plane speaker is used. First, the satellite, whose position is over the top of the user at the moment will be chosen. This satellite will then choose other speakers at different planes having its the same latitude with the one that has been chosen. The speaker is viewed as the place where the status messages of satellites are collected in the same plane as their speaker. The users will pass a signal to the control mechanism when there are packets to be sent. Then, the satellite speaker above will be called to collect the information. After all plane speakers receive all information from the network, the routing tables will be calculated. Then, only the necessary tables will be passed to their partners in the same plane. Finally, the user will be notified to pass the real traffic.

4.6 ERRS algorithm

A DTN (Disruption-Tolerant Network) can provide store-and-forward enabled routing strategies for the satellite network with frequent intermittent links. However, the current DRSA (Dynamic Route Selection Algorithm), which includes the CGR (Contact Graph Routing) algorithm, cannot find an end-to-end route over serial link segments with time-disjointed contacts. The routing algorithm, ERRS (Expanding Range Route Selection) is presented, which is able to find the EDT-optimal route by searching at each snapshot of time-varying topology[18].

Different from traditional TCP/IP protocol, the DTN architecture adds a new layer between the transport and application layer, known as the bundle layer. Real end-to-end data reliability across a hybrid network is provided by the bundle layer. This protocol architecture is shown in Fig.14.

The proposed algorithm is composed of two parts. The first part establishes the routing table on the ground, using time slices to describe the dynamic network topology and record all the accessible routes. In each time slice, the network topology remains unchanged. The second part is the routing algorithm. Simulation results show that the proposed algorithm can not only reduce the end to end transmission time, but also significantly increase the throughput of the network.

5 Routing with invulnerability based on multilayer satellite network

Invulnerability is an important factor to consider when designing a routing algorithm in a satellite network.

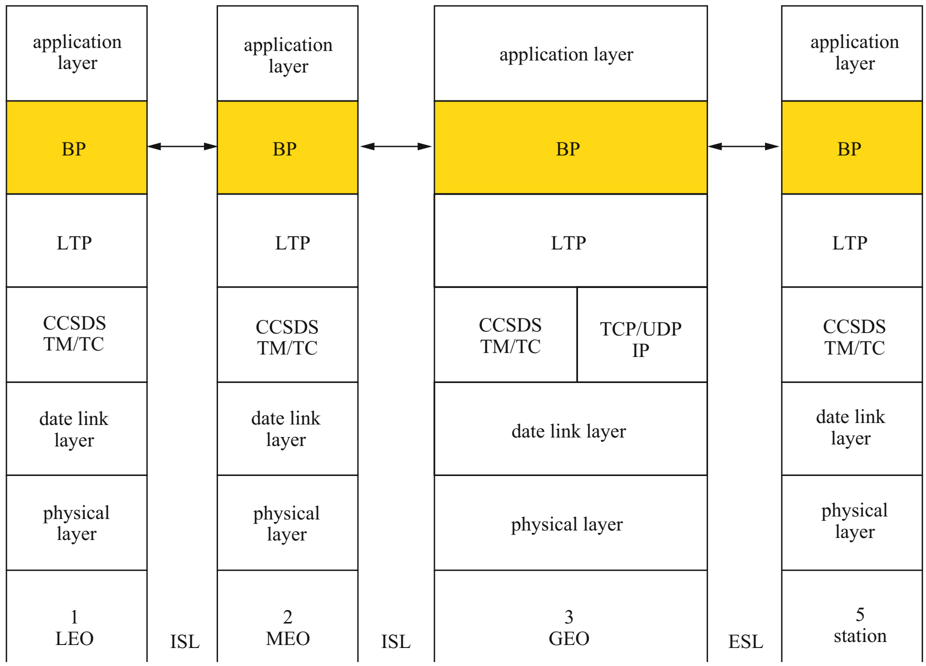


Figure 14 Protocol architecture

Although several multilayer satellite network routing technologies have been proposed, few studies take into account satellite network invulnerability.

Satellite grouping is one of the basic routing strategies in multilayer satellite networks. Unfortunately, the satellite group, due to the frequent movement of the LEO satellite, may produce a large number of uneven time slices. Moreover, the upper satellite can only capture the change of the lower satellite group, and cannot capture the variation of the satellite topology in the lower layer group. In this case, the network invulnerability is very weak. In Ref.[19], a formal model is given to clarify the dynamics and permanence of the virtual topology based on the virtual node method for LEO satellite networks. In this model, logical regions are stationary and a virtual topology is constructed, called FVT (Footprint-based Virtual Topology). In this manner, a virtual network between the satellite layer and earth is established.

Being inspired after having fully utilized the characteristics of the topology in Ref.[19] and having got inspiration from it, we predict that several routing schemes with invulnerability will be presented based on the multilayer satellite network after combining both centralized and distributed routing strategies.

A specific example of invulnerable routing follows[20]: The LEO layer is formalized as the FVT, and then the MEO layer collects the link delay and calculates the routing table for the LEO every $2\pi/m$ snapshot period, where m is the number of LEO satellites. As a result, when there are no LEO satellite failures, the MEO layer satellites are able to optimize the routing calculation for LEO satellites at any time because the LEO layer topology keeps constant. Even if LEO satellites failures exist, a few uniform snapshots of LEO layer are still guaranteed, which can improve the performance of routing, especially in invulnerability. In this manner, the topology of the satellite group is fully captured. The scheme, to some degree, can solve the defects of other multilayer

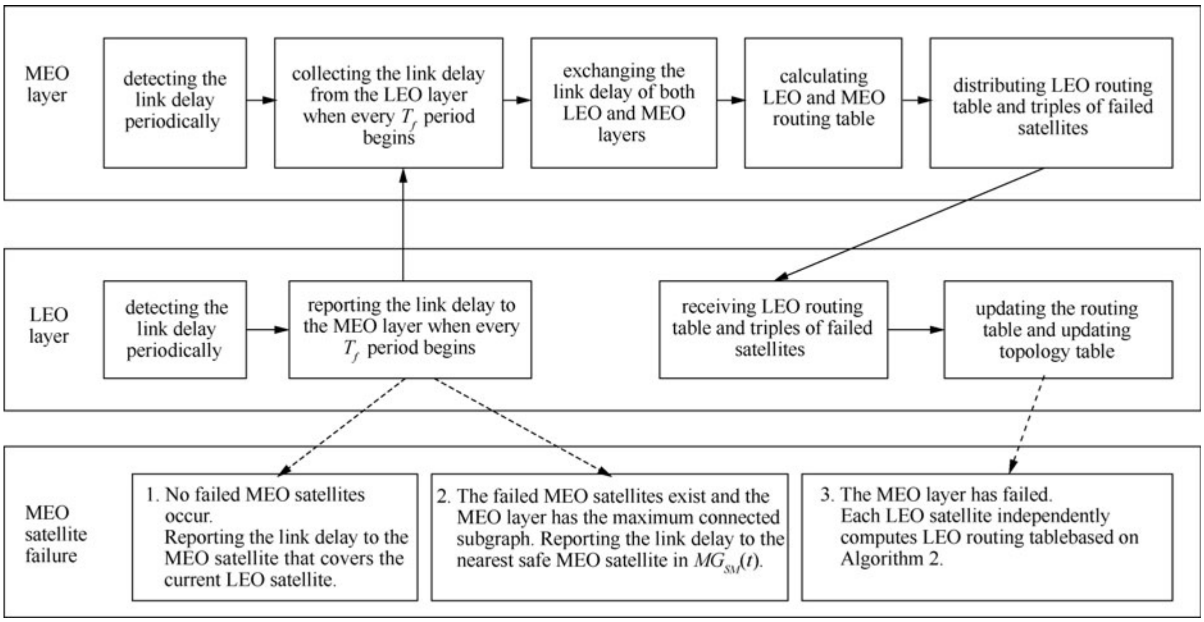


Figure 15 The route scheme with invulnerability

satellite network routing algorithms. Fig.15 shows the framework for this scheme.

Most multilayer satellite network route methods, previously mentioned, need to obtain link delay information from the lower layer and calculate all possible end-to-end paths. However, these methods are unable to adapt to the network node or link failure in a timely manner.

The ASOR (A-Star algorithm based on the On-demand Routing) algorithm is based on the on-demand routing protocol for hierarchical LEO and MEO satellite networks. This is proposed to solve the vulnerability of network[21], inspired by Ref.[13]. The protocol uses the A-star algorithm and an estimated cost function $f(n)$. The function $f(n) = g(n)+h(n)$ is the estimated cost of the shortest route through node n, where $g(n)$ is the cost function from source to the current node n, while $h(n)$ is the heuristic estimate cost function of the path from node n to the destination node. Among all the candidate nodes, the current node n, which has the smallest estimated cost $f(n)$, is the first node to be checked. The network model used in the protocol is shown in Fig.16. The primary process of the ASOR protocol is shown in Fig.17.

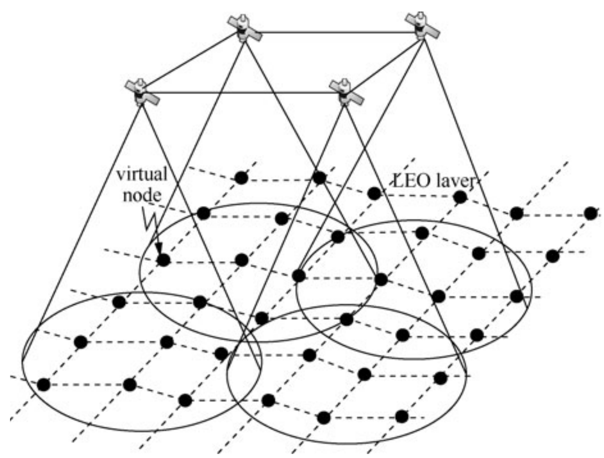


Figure 16 LEO/MEO joint constellation

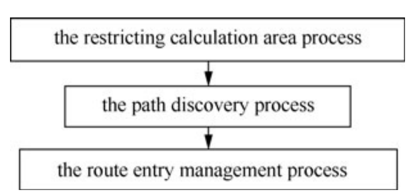


Figure 17 The primary process of the ASOR protocol

The protocol includes three processes:

1) The restricting calculation area process: In this

process, the MEO satellites can send the RREQ (Route Request) packets to obtain the state information of the LEO links. A request area formation process can be used in order to reduce the overhead before the next process.

2) The path discovery process is shown as follows.

Step 1 Initialization. The source LEO satellite caches the received packets in the buffer. Next, it identifies the master MEO satellite for the corresponding MEO satellite. This master MEO satellite instructs the slave MEO and other MEO satellites to obtain the link state information of the LEO satellites that are in the restricting area.

Step 2 The master MEO satellite calculates the optimal path and its lifetime. Next, it sends this information to other MEO satellites. Thus, they are able to easily calculate the entries according to the path information.

Step 3 The MEO satellites send the calculated route entries to the LEO satellites, and then, notify the source LEO satellite that the path discovery process has been completed.

Step 4 The source LEO satellite sends the packets using the optimal path.

3) The route entry management process: This process is employed to follow the tracks of the current activated route entries.

In spite of these processes, the invulnerability of the satellite network still needs to be strengthened. This is because the satellite network environment is complex, and the damage caused by satellite nodes and links is extremely serious. Thus far, there is no mature invulnerable routing protocol that has been put into use. We firmly believe that routing technology with strong invulnerability, resistant of the destruction of nodes and links, will emerge.

6 Future research trends of satellite network

Owing to cyclical changes in the satellite network topology, designing routing technology has always been a challenging problem. Although, there are a variety of satellite routing protocols and algorithms as previously described, a systematic satellite network routing algorithm is still lacking. Moreover, there is no fully accepted standard for satellite routing protocols, thus, routing technology must be further researched. Satellite network routing will, undoubtedly, become a hot topic in future space research, primarily including the following aspects:

1) Studying the route with invulnerability. The future space environment of the satellite network is more complicated. Other than the complexity of the space environment, the network may also suffer from human factors, such as military attacks. Several existing routing schemes have a common shortcoming that the performance of the network will decline sharply when the node or link in the satellite network is destroyed. Currently, most routing protocols rarely consider the survivability of satellite network routing. Some single-layer network survivability protocols have been proposed, however, for upper satellite failure in a multi-layer network condition, there is no superior routing technology to deal with satellite failure. Therefore, it is urgent to design a route with strong invulnerability. In the future, designing invulnerable routing will be a problem that should be widely discussed.

2) Meeting the multi-QoS routing technology. Among the existing satellite routing technologies, many are based on a certain satellite network index, such as delay, packet loss rate, and throughput. In the future, multi-QoS routing technology needs to be improved.

3) The future space networking mechanisms of satellite networks will become more flexible, and all types of flight devices may establish temporary or permanent links with satellites in order to form a more complex spatial network level. This may provide more flexible and abundant data transmission

services for the ground.

4) Optimizing network architecture. The satellite network, as a spatial network, has large scale features, concerning many aspects, such as complicated structure, various operations, strong scalability, and dynamic topology. Thus, scientific design and reasonable optimization of network architecture, will be the future development direction of Internet space.

5) Using DTN protocol. Because of certain differences between satellite and ground network, the satellite network is not suited to the traditional TCP/IP protocol. However, the DTN protocol is quite suitable for a space communication environment. Thus, it is likely to be the focus of future development direction[22].

7 Conclusion

In the process of establishing a satellite system network, the primary problem is the routing technology. Thus far, there is no practical dynamic satellite network routing system available in China. The global coverage of the satellite network has greatly expanded the communication space of the network, and promoted the development of the next generation of Internet and space networking. Future development and application of space technology will promote the seamless integration of satellite and terrestrial networks, which can bring more extensive communication services.

In this paper, satellite network routing technologies were briefly summarized from classification perspective. Core mechanisms, primary features, and existing problems of different representative routing technologies were discussed to provide information and ideas for related research. Throughout these discussed methods, the problem of invulnerability routing in the satellite network was rarely considered. In the future, designing a routing scheme with high invulnerability will be a key point in satellites networks technology. Moreover, we will continue to further study satellite routing in the aspect of invulnerability. Finally, the challenges and problems faced by the satellite network were analyzed and future research direction of satellite network routing technology was examined.

References

[1] WERNER M. A dynamic routing concept for ATM-based satellite personal communication networks[J]. IEEE journal on selected areas in communications, 1997, 15(8): 1636-1648.

[2] CHANG H S, KIM B W, LEE C G, et al. FSA-based link assignment and routing in low-earth orbit satellite networks[J]. IEEE transactions on vehicular technology, 1998, 47(3): 1037-1048.

[3] BAI J, LU X, LU Z, et al. Compact explicit multi-path routing for LEO satellite networks[C]//The Workshop on High PERFORMANCE Switching and Routing, 2005: 386-390.

[4] TARIK T, DAISUKE M, ABBAS J. ELB: an explicit load balancing routing protocol for multi-hop NGEO satellite network[J]. Proceedings of IEEE, 2006: 1-5.

[5] KORCAK O, ALAGOZ F. Analysis of priority-based adaptive routing in satellite networks[C]//The 2nd International Symposium on Wireless Communication Systems, 2005: 629-633.

[6] TAN H, ZHU L. A novel routing algorithm based on virtual topology snapshot in LEO satellite networks[C]//IEEE 17th International Conference on Computational Science and Engineering (CSE), 2014: 357-361.

[7] EKICI E, AKYILDIZ I F, BENDER M D. Data-gram routing algorithm for LEO satellite networks[C]//The19th Annual Joint Conference of the IEEE Computer and Communications Societies, 2000, 2: 500-508.

[8] CHAN T H, YEO B S, TURNER L. A localized routing scheme for LEO satellite networks[C]//AIAA 21st International Communications Satellite Systems Conference and Exhibit. 2003: 2357-2364.

[9] EKICI E, AKYILDIZ I F, BENDER M D. A distributed routing algorithm for datagram traffic in LEO satellite networks[J]. IEEE/ACM transactions on networking, 2001, 9(2): 137-147.

[10] LIU X, YAN X, JIANG Z, et al. A low-complexity routing algorithm based on load balancing for LEO satellite networks[C]//IEEE 82nd Vehicular Technology Conference (VTC Fall), 2015: 1-5.

[11] LIU X, JIANG Z, LIU C, et al. A low-complexity probabilistic routing algorithm for polar orbits satellite constellation networks[C]//IEEE/CIC International Conference on Communications in China (ICCC), 2015: 1-5.

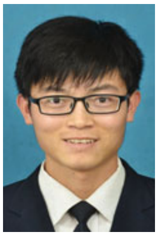
[12] JI X, LIU L, ZHAO P, et al. A destruction-resistant on-demand routing protocol for LEO satellite network based on local repair[C]//The 12th International Conference on Fuzzy Systems and

Knowledge Discovery (FSKD), 2015: 2013-2018.
[13] AKYILDIZ I F, EKICI E, BENDER M D. MLSR: a novel routing algorithm for multilayered satellite IP networks[J]. IEEE/ACM transactions on networking, 2002, 10(3): 411-424.
[14] CHEN C, EKICI E. A routing protocol for hierarchical LEO/MEO satellite IP networks[J]. Wireless networks, 2005, 11(4): 507-521.
[15] LEE J W, LEE J W, KIM T W, et al. Satellite over satellite (SOS) network: a novel concept of hierarchical architecture and routing in satellite network[C]//The 25th Annual IEEE Conference on Local Computer Networks, 2000: 392-399.
[16] LIANG J, ZHANG J W, XIAO N. Research and simulation on an autonomous routing algorithm for GEO-LEO satellite networks[C]// International Conference on Intelligent Computation Technology and Automation. 2011: 657-660.
[17] FENG X, YANG M, GUO Q. A novel distributed routing algorithm based on data-driven in GEO/LEO hybrid satellite network[C]// IEEE International Conference on Wireless Communications and Signal Processing (WCSP), 2015: 1-5.
[18] WU Y, YANG Z, ZHANG Q. A novel DTN routing algorithm in the GEO-relaying satellite network[C]//The 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN), 2015: 264-269.
[19] LU Y, SUN F, ZHAO Y. Virtual topology for LEO satellite networks based on earth-fixed footprint mode[J]. IEEE communications letters, 2013, 17(2): 357-360.
[20] LU Y, ZHAO Y, SUN F, et al. A survivable routing protocol for two-layered LEO/MEO satellite networks[J]. Wireless networks, 2014, 20(5): 871-887.
[21] JI X, LIU L, ZHAO P, et al. A-Star algorithm based on-demand routing protocol for hierarchical LEO/MEO satellite networks[C]// IEEE International Conference on Big Data (Big Data), 2015: 1545-1549.
[22] CAINI C, CRUICKSHANK H, FARRELL S, et al. Delay-and disruption-tolerant networking (DTN): an alternative solution for future satellite networking applications[J]. Proceedings of the IEEE, 2011, 99(11): 1980-1997.

About the authors

QI Xiaogang was born in Baoji, Shaanxi. He is now a Ph.D. tutor and professor. His current research interests include system modeling and fault diagnosis. (Email: xgqi@xidian.edu.cn)

LIU Lifang was born in Lanzhou, Gansu. She is now a Ph.D. tutor and professor. Her current research interests include data processing and intelligent calculation. (Email: liu@xidian.edu.cn@126.com)



MA Jiulong was born in Yan'an, Shaanxi. He is now a graduate student at Xidian University. His current research interest includes routing technology of satellite network. (Email: majiulong@stu.xidian.edu.cn)

HU Shaolin was born in Anqing, Anhui. He is now a Ph.D. tutor and researcher. His current research interests include system security, process monitoring, data mining, data musion and aerospace measurement and control system engineering. (Email: hfkth@126.com)

WU Dan was born in Laiwu, Shandong. She is now a graduate student at Xidian University. Her current research interests include the optimization and survivability for delay tolerant satellite network. (Email: danwufan@foxmail.com)