UMR: A utility-maximizing routing algorithm for delay-sensitive service in LEO satellite networks

Lu Yong, Zhang Jun *, Zhang Tao

School of Electronics and Information Engineering, Beihang University, Beijing 100191, China

Received 24 April 2014; revised 1 September 2014; accepted 24 December 2014
Available online 23 February 2015

KEYWORDS
Delay bound; LEO satellite networks; Micro-satellite; Network utility; QoS routing

Abstract This paper develops a routing algorithm for delay-sensitive packet transmission in a low earth orbit multi-hop satellite network consists of micro-satellites. The micro-satellite low earth orbit (MS-LEO) network endures unstable link connection and frequent link congestion due to the uneven user distribution and the link capacity variations. The proposed routing algorithm, referred to as the utility maximizing routing (UMR) algorithm, improve the network utility of the MS-LEO network for carrying flows with strict end-to-end delay bound requirement. In UMR, first, a link state parameter is defined to capture the link reliability on continuing to keep the end-to-end delay into constraint; then, on the basis of this parameter, a routing metric is formulated and a routing scheme is designed for balancing the reliability in delay bound guarantee among paths and building a path maximizing the network utility expectation. While the UMR algorithm has many advantages, it may result in a higher blocking rate of new calls. This phenomenon is discussed and a weight factor is introduced into UMR to provide a flexible performance option for network operator. A set of simulations are conducted to verify the good performance of UMR, in terms of balancing the traffic distribution on inter-satellite links, reducing the flow interruption rate, and improving the network utility.

© 2015 The Authors. Production and hosting by Elsevier Ltd. on behalf of CSAA & BUAA. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In recent years, the low earth orbit (LEO) satellite network has started developing into an integrated heterogeneous network with significant wireless characteristics on several links. Furthermore, increasing presence of micro-satellites and nano-satellites in several small-scale self-organizing LEO satellite networks has been observed, and these satellites will be employed to form constellations and interconnect with other satellite networks.1–3 As the real-time service is the core business of the LEO satellite networks and the foundation for providing multimedia applications in the future, effective transmission of the delay-sensitive flows with a strict end-to-end delay bound constraint via the multihop wireless satellite links has become one of the challenges that must be addressed in the development of the micro-satellite low earth orbit (MS-LEO) networks.

* Corresponding author. Tel.: +86 10 82317846 6272.
E-mail address: xlyong0828@gmail.com (J. Zhang).
Peer review under responsibility of Editorial Committee of CJA.

http://dx.doi.org/10.1016/j.cja.2015.01.001
1000-9361 © 2015 The Authors. Production and hosting by Elsevier Ltd. on behalf of CSAA & BUAA.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
The static resource reservation coupled with the QoS packet scheduling scheme is the conventional solution for the delay-sensitive flow transmission in the Internet. This method provides delay bound assurance by guaranteeing the required minimum resource supply according to the flow characteristics and works well in preconditions of the unchanged flow characteristics and link capacity. The weighted fair scheduling scheme such as the generalized processor sharing (GPS),\(^1\) the weighted round-robin\(^2\) (WRR) and the weighted fair queuing\(^6\) (WFQ) is a class of typical and advanced schemes for QoS packet scheduling.

However, in wireless networks, links often endure frequent on-off and congestion, which often makes the network fail to achieve sustained matching between the bandwidth supply and the bandwidth demand in flow duration, which leads to poor network utility in providing an integrated service for delay-sensitive flows.\(^7\) This problem is more serious in an MS-LEO satellite network, affected by its inherent features. First, the capacity of the inter-satellite links (ISLs) is limited, and both the source and the destination users in MS-LEO satellite network are in un-uniform distribution. The significant imbalance on the traffic flow distribution may lead to ineffective use of the insufficient bandwidth resource. Second, under the combined influence of the low-precision antenna, the limited power, the changes on the node distance and the environment interference, the link-up time jitter and the link capacity jitter are common on ISLs, leading to frequent congestion on the stature links, which account for a large proportion of the ISLs in use in the MS-LEO satellite networks. Third, the handovers of the inter-satellite links and satellite-ground links cause substitute traffic shift among ISLs, which may lead to severe burst congestion or even high interruption rate on communication hotspots. The frequent link-off and link congestion will cause repeated delay threshold breakthrough and eventually result in high delay QoS interruption rate and poor network utility.

The routing algorithm for congestion avoidance is one of the key solutions to this problem on the network layer in LEO satellite networks.\(^8\) The algorithm is primarily classified into two categories. One category is the adaptive routing schemes.\(^7\) The traffic on statured links is adjusted in real-time to peripheral idle links to avoid congestion. Most of these schemes do not support a strict delay QoS constraint. The other category is the load balancing routing algorithm.\(^10\)-\(^14\) The link utilization ratio is considered in the routing metric and priority is given to the route with the current or forecasted minimal traffic load level. Link residual bandwidth balancing plays a role in tolerating the link changes in the network, which reduces the congestion probability. The network utility is improved by more extensive use of ISLs and a lower congestion probability. Generally, the advantage of load balancing is considered based on the resource cost of a larger route hop number and the performance cost of a longer propagation delay. However, as we found in this paper, in the MS-LEO satellite network, when the load balancing routing algorithm is required to build a path that also provides strict end-to-end delay bound, the following three types of problems are encountered:

1. The bandwidth requirement for the delay bound guarantee increases significantly with the route hop number and the propagation delay. This problem not only causes a heavy cost on bandwidth resource but also increases the link congestion probability when the network is on heavy load.
2. Longer routes increase the probability of the delay bound breakthrough in a network with random link on-off and congestion.
3. Most of the load balancing routing algorithm is based on a uniform congestion avoidance principle. The algorithm is easily implemented but does not take into account the link characteristics and the flow characteristics in routing and fails to achieve network utility optimization.

To solve these problems, this paper proposes the UMR, a utility-maximizing routing algorithm for delay-sensitive flows in the MS-LEO satellite network. The UMR algorithm identifies a route that not only reduces the probability of delay bound breakthrough for the new call but also optimizes the total network utility. In principle, the objective is achieved by balancing the congestion probability of the new path and the influence of the new path on the congestion probability of existing paths. First, to calculate the probability of path congestion and evaluate the influence on the delay bound of the existing flows, a link state parameter, denoted by the link feasible probability (LFP), is defined to form an explicit evaluation on the link property for the delay bound guarantee. The LFP is formulated with consideration on the link characteristics and the flow characteristics. Based on LFP, the probability of congestion on a path can be formulated as a comprehensive consideration on the bandwidth resource cost, the route hop number and the propagation delay. Then, by adopting the path congestion probability, this routing problem can be considered as an utility-optimization problem in a flow utility self-interference system, where the route decision affects the variation trend of the total network utility by determining the interference relationship among the flow utilities. Finally, an on-demand distributed route-detecting algorithm is designed to find the quasi-optimal route solution with local link state information.

The remainder of the paper is organized as follows. In Section 2, the LFP is defined, and the objective of the UMR is formulated. In Section 3, the estimation model for the LFP in MS-LEO satellite network is formulated. In Section 4, the routing scheme is described. In Section 5, the performance of the UMR is evaluated through simulations, and a detailed analysis and discussion is provided. Finally, the paper is concluded in Section 6.

2. Definitions of LFP and the objective of the UMR

Considering a MS-LEO satellite network \( G = (V, E) \), where \( V \) is the node set, and \( E \) is the link set. Let \( T(t) \) be the existing flow set in \( G \) at time \( t \), in which each flow \( f \) traverses the network via a multi-hop route \( r_f \). Suppose each flow has a delay deadline \( D_f \) and can be characterized by a collection of known parameters \( \text{CH}_f = \{ t_f^s, A_f(t_f), \sigma_f, \rho_f, L_f^{\max} \} \), where \( t_f^s \) is the transmission start time, and \( A_f(t_f) \) is the distribution function of flow duration \( t_f \); \( \sigma_f \) denotes the maximum burst length, \( \rho_f \) denotes the average data arriving rate, and \( L_f^{\max} \) is the maximum packet length.
These parameters are readily accessible and can be obtained in the QoS negotiation or according to flow transmission statistics. The path \( p_f \) is formulated by a QoS routing algorithm. The resource reservation and a weighted fair packet scheduling schemes are adopted. The scheduling queue buffer is assumed to be infinite; therefore, the congestion control strategy will not be considered and we can only focus on the relationship of the routing schemes and the delay performance. \( d_s(t) \) is the end-to-end delay at time \( t \), and the LFP is defined below.

**Definition 1.** For link \( e \) on path \( p_f \) at time \( t \in [t'_e, t'_e + \tau_e] \), the LFP for flow \( f \), denoted by \( P'_f(t) \), is defined as the probability that link \( e \) continues to guarantee the delay bound for flow \( f \) in \([t, t'_e + \tau_e]\).

Suppose that each flow transmitting in the network is associated with a utility function \( U_f(p_j) \), which indicates the user satisfaction for transmitting flow \( f \) through route \( p_j \). In this paper, we adopt a simple and typical flow utility function \( U_f(p_j) = \mu_f \cdot o_f \) with practical significance, where \( o_f \) represents the value weight of flow \( f \), which can be formulated by a comprehensive consideration of the flow priority and the transmission rate; \( \mu_f = 1 \) if flow \( f \) completes the entire transmission process without breakthrough of delay bound; otherwise, \( \mu_f = 0 \). Assume that the links are independent of each other.

Then, \( U_f(p_j) \) is a random variable with its expectation at time \( t \in [t'_e, t'_e + \tau_e] \) given by

\[
E(U_f(p_j)) = \mu_f \cdot o_f \prod_{e \in p_j} P'_f(t_e) \tag{1}
\]

Suppose that the flow utility is additive; therefore, the aggregate utility can be regarded as the network utility. The greedy policy is employed and the routing objective is set to maximize the expectation of the total network utility after the path establishment for each flow.

Hence, \( t'_e \) is the initiation time of the path searching for flow \( f \), and \( t'_e \) is the adoption time of route \( p_j \) for flow \( f \). Assume that \( f \) is the only flow applying for a route \([t'_e, t'_e] \). Let \( \Gamma(t'_e) \) and \( \Gamma(t'_e) = \Gamma(t'_e) \cup f \) be the flow set in the network at \( t'_e \) and \( t'_e \), respectively. Then, the objective of the UMR algorithm is to find the \( p_j \) that maximizes \( U'_e(t'_e) \), the expectation of the total network utility at \( t'_e \), which according to Eq. (1) can be formulated by

\[
E(U'_e(t'_e)) = \sum_{\mathbf{i} \in \Gamma(t'_e)} \mu_f \cdot o_f \prod_{e \in p_j} P'_f(t_e) \tag{2}
\]

### 3. LFP formulation in a MS-LEO satellite network

In this section, the LFP parameter \( P'_f(t) \) is formulated in the network model of a MS-LEO satellite network. We address the problem of how to evaluate the probability of \( d_s(t_0) \leq D_f \) for \( t_0 \in [t, t'_e + \tau_e] \) at time \( t \in [t'_e, t'_e + \tau_e] \) according to the characteristics of link \( e \) given the flow set \( \Gamma(t) \), its parameter set collection \( \text{CH}_f(t) \), and the transmission path set \( \text{Path}(t) \).

#### 3.1. Bandwidth condition for delay bound constraint

The delay-bandwidth transfer relationship of the weighted fair scheduling schemes enables the evaluation of the end-to-end delay in a distributed way on links.

Let \( K_e \) denote the hop number of the path \( p_j \), on which the \( k \)th hop link \( e_k \) has its link capacity \( r_{e_k} \). \( |g'_{e_k}| \) is the bandwidth allocated on \( e_k \). \( g(p_j) \) is the minimum of the \( g'_{e_k} \) on \( p_j \). Let \( d_{\text{prop}}(p_j) \) be the propagation delay on \( p_j \). Then, according to Parekh and Gallager\(^\circ\) the relationship between the end-to-end delay and the bandwidth obtained on the path can be characterized by

\[
D_f \leq \frac{\sigma_f}{g(p_j)} + \sum_{k=1}^{K_e-1} \frac{r_{e_k}}{g_{e_k}} + \sum_{k=1}^{K_e-1} \frac{r_{e_k}}{r_{e_k}} + d_{\text{prop}}(p_j) \tag{3}
\]

With the reasonable assumption that \( L_{\max}/r_{e_k} \) is negligible (approximately \( 10^{-4} \) s for 1500 B and 25 M link capacity) and supposing that links on the path make consistent decisions on the minimum bandwidth resource reservation, the minimum bandwidth resource requirement to guarantee the end-to-end delay bound \( D_f \) on path \( p_j \), denoted by \( g'(p_j) \), can be given by

\[
g'(p_j) = \max \left( \frac{\sigma_f + (K_e - 1) \times L_{\max}}{D_f - d_{\text{prop}}(p_j)} , r_{e_k} \right) \tag{4}
\]

#### 3.2. LFP formulation

According to the Definition 1, in MS-LEO satellite network, \( P'_f(t) \) is formulated by

\[
P'_f(t) = P'_{\text{s},f}(t) \times P'_{\text{c},f}(t) \tag{5}
\]

where \( P'_{\text{s},f}(t) \) is the probability of the persist connection of link \( e \) during \([t, t'_e + \tau_e] \); \( P'_{\text{c},f}(t) \) is the probability of continuously providing bandwidth resources larger than \( g'(p_j) \) for flow \( f \) on link \( e \) during \([t, t'_e + \tau_e] \) under the condition of link connection.

#### 3.2.1. Formulation of \( P'_{\text{s},f}(t) \)

According to the link connection characteristics in the MS-LEO satellite network, \( P'_{\text{s},f}(t) \) is formulated by

\[
P'_{\text{s},f}(t) = P'_{\text{s},f}(t) P'_{\text{e},f}(t) \tag{6}
\]

where \( P'_{\text{e},f}(t) \) denotes the probability that flow \( f \) survives from the forced link disconnection in \([t, t'_e + \tau_e] \), which is induced by several factors such as the relative motion of satellites, the variations on satellite attitude and power, and the space environment interferences, \( P'_{\text{e},f}(t) \) contributed by the link-off determined by the link shutdown plans, \( r'_e \) is the link disconnection time of link \( e \).

The link-off time of the forced link disconnection cannot be accurately predicted. If the micro-satellites do not have directional antennas and precise control units, the link disconnection in handovers on ISVs should also be considered as forced link disconnection. However, as satellites move on periodic orbit, the forced link disconnection exhibits regional characteristics. Suppose that a segment of a curve on the orbit for the transmitting satellite of link \( e \), denoted by region \( R(e) \), is susceptible to link disconnection in which the probability of link disconnection, denoted by \( \varphi_{R(e)} \), can be obtained according to the empirical statistics. Assume that the satellite crosses \( R(e) \) at an approximate uniform speed and the link-off location in \( R(e) \) is in accordance with a uniform distribution. Let
\( P_{R_s}(t) \) be the probability of link disconnection in \( R(e) \) for link \( e \) estimated at \( t \) as shown by

\[
P_{R_s}(t) = \begin{cases} \frac{a(t)}{1 - A(t - t_s)} e^{\int_{t_m - t_o}^{t_s} \frac{a(t)}{1 - A(t - t_s)} dt} (t \in [t_s', t_{out}]) \\ 0 (t \notin [t_{fast}, +\infty)) \end{cases}
\]

(7)

where \( a(t) \) is the probability density function of the flow duration \( t_s, t_m \) and \( t_{out} \) denote the time that the transmitting satellite of link \( e \) enters into and moves out of \( R(e) \), respectively. \( t_s' = t \) if \( t > t_o \). Otherwise, \( t_s' = t_o \).

Thus, suppose that the transmitting node of link \( e \) plans to cross the region set \( \{ R_k(e) \} \) in \( [t, t_s' + t_e] \) when the regions are independent, the \( H_e(t) \) in Eq. (6) can be given by

\[
H_e(t) = \prod_{\{ R_k(e) \} = 1} \left( 1 - P_{R_s}(t) \right)
\]

(8)

In a MS-LEO satellite network, the ISL is often disconnected according to a preset time such as when there is an initiative link handover; when the ISL enters into a power saving period of the satellite; or when the satellite enters into a link-off region just like the polar cycle. As the link disconnection time is pre-generated and stored on-board, this time can be accurately predicted.

\( r_s' \) is the link disconnect time of link \( e \). Then, the probability that flow \( f \) terminates transmission before \( r_s' \), denoted by \( P_{r_s'}(r_s', t) \) as in Eq. (6), is obtained by Eq. (9), and \( P_{r_s'}(r_s', t) \) \begin{align*} \rightarrow 1 & (r_s' \rightarrow +\infty) \end{align*}. Thus, if \( r_s' \) does not exist, we consider \( P_{r_s'}(r_s', t) = 1 \) in the following analysis.

\[
P_{r_s'}(r_s', t) = \frac{A(t - t_f') - A(t - t_f)}{1 - A(t - t_f)}
\]

(9)

3.2.2. Formulation of \( P_{r_s'}(t) \)

In a MS-LEO network, primarily two factors contribute to \( P_{r_s'}(t) \) in Eq. (5). One factor is the initiative power adjustment according to a preset power saving plan. The other factor involves the capacity declination induced by the receiving power coast-down and the random interference. Suppose that according to the power adjusting time schedule, \([t, t_s' + t_e]\) is divided into discrete time intervals \( t_k = [t_k, t_{k+1}], k = 0, 1, 2, \ldots \) in which \( t_0 = t \). Then, \( P_{r_s'}(t) \) is evaluated by

\[
P_{r_s'}(t) = \sum_{k} \left( 1 - P_{r_s'}(t_k, t) \right) P_{r_s'}(t_{k+1}, t) P_{r_s'}(t_k, t)
\]

(10)

where the \( P_{r_s'}(t_k, t) \), which can be obtained by Eq. (9), is the probability that flow \( f \) terminates before the \( k \)th power adjusting start time \( t_k \), the \( P_{r_s'}(t_k, t) \) denotes the probability of keeping bandwidth supply for flow \( f \) larger than \( g^*(p_f) \) in \([t, t_s' + t_e]\) with preconditions \( t_s' + t_e \in [t_k, t_{k+1}] \).

Let \( r_s'(t) \) denote the minimum capacity requirement of flow \( f \) on link \( e \) at \( t \), according to the weighted faire scheduling scheme, and can be given by

\[
r_s'(t) = \frac{g^*(p_f) \sum_{n \in n_t(e)} \phi^*_n}{\phi^*_e}
\]

(11)

where \( I_e(t) \) is the flow set on link \( e \) at \( t \) and \( \phi^*_n \) and \( \phi^*_e \) are the scheduling weight factor for flow \( i,f \in I_e(t) \).

Let \( P_{r_s'}(t) \) denote the transmitted power and \( L_r(t) \) be the gain on link \( e \). \( \sigma \) is the thermal background noise. \( W_r \) is the total bandwidth on link \( e \). The stepwise reduction in link capacity often occurs due to the reduction plan of \( P_{r_s'}(t) \) because the limit on satellite node power is one of the main characteristics in the MS-LEO network. According to the power adjusting plan, each time interval \( t_k \) has a transmitted power \( P_{r_s'}(t_k) \). The gradual reduction of capacity is associated with the variation in \( L_r(t_k) \). The relative motion of the satellites is predictable; therefore, \( L_r(t_k) \) can be pre-calculated, and the downward trend of \( L_r(t_k) \) can be easily identified. Thus, let \( w_{s'}(t_k, t_{k+1}) \) be the minimum link capacity in \([t_k, t_{k+1}]\), according to the Shannon theorem, and can be given by

\[
w_{s'}(t_k, t_{k+1}) = \frac{W_r \times \log \left( 1 + P_{r_s'}(t_k) \times \min(L_r(t')) \right)}{\sigma}
\]

(12)

for \( t' \in [t_k, t_{k+1}] \).

Thus, the minimum link capacity in \([t, t_s' + t_e]\) with \( t_s' + t_e \in [t_k, t_{k+1}] \), denoted by \( w_{s'}(t, t_s' + t_e) \), can be given by

\[
w_{s'}(t, t_s' + t_e) \approx \min(w_{s'}(t_1, t_s'), w_{s'}(t_1, t_2), \ldots, w_{s'}(t_{k-1}, t_k), w_{s'}(t_k, t_{k+1}))
\]

(13)

In addition to the thermal noise, the link capacity may also be affected by other interference and cause a random capacity link fluctuation \( e_{t_k} \). Let \( w_{r_s'}(t_k) \) denote the link capacity at time \( t_k \). Suppose that the actual link capacity \( r_{s'}(t_k) \) can be measured by employing the cognitive radio technology, the \( e_{t_k} \) can be estimated by \( e_{t_k} = w_{s'}(t_k) - r_{s'}(t_k) \), and the probability distribution function \( F(e_{t_k}) \) can be formulated from the statistics of \( e_{t_k} \) in real-time.

Therefore, suppose that the flow set on link \( e \) remains unchanged during \([t, t_s' + t_e]\). Then, according to the definition, the \( P_{r_s'}(t_k, t) \) in Eq. (10) can be given by

\[
P_{r_s'}(t_k, t) \approx F[w_{s'}(t, t_s' + t_e) - r_{s'}(t)]
\]

(14)

3.3. Discussion on the LFP

Remark 1. The LFP is a function of time. Real-time updating enables an accurate assessment; however, the LFP is still beneficial for network utility even if not updated due to its predictive inference in definition.

Remark 2. The relationship between LFP and route choice is mainly determined by the minimum bandwidth requirement on local link as illustrated in Eq. (11). Therefore, a path choice for a new flow only affects the LFP of those flows whose paths share common links.

Remark 3. As illustrated by Eqs. (5), (10)(11) and (14), the LFP of flows will be decreased by any new route entry on common links, which forms a mutual interference relationship among the flow utilities similar to a self-interference system and enables the routing decision to maximize the network utility.
gaining on itself and the utility loss of the influenced flows. This idea has inspired the design of the routing scheme presented in the next section.

4. Routing scheme of UMR

4.1. Formulation of routing metric

Suppose the anterior \( k-1 \) hops of \( p_f \), denoted by \( p_f^{k-1} \), have been formed before link \( e \) in route searching. Then, according to the objective of the UMR, the routing metric for flow \( f \) on link \( e \), denoted by \( u'_f(p_f^e) \), is defined as the variation in the expectation of the network utility by selecting link \( e \) as the \( k \)th hop link of \( p_f \) and is approximately formulated by

\[
 u'_f(p_f^e) = o_f \prod_{m \in p_f^e} P_m^e(t_f^m) - \xi \sum_{m \in p_f^e \setminus \{f\}} o_f P_f^e(t_f^m) \left( 1 - \frac{P_m^e(t_f^m)}{P_f^e(t_f^m)} \right)
\]

with

\[
 P_f(i) = \prod_{m \in p_f^e} P_m^e(i)
\]

According to Eqs. (15) and (16), the \( u'_f(p_f^e) \) can be further given by

\[
 u'_f(p_f^{k-1}) + o_f(P_f^e(t_f^m) - 1) \prod_{m \in p_f^{k-1}} P_m^e(t_f^m) - \xi C_f(t_f^m) - \sum_{m \in p_f^e \setminus \{f\}} o_f P_f^e(t_f^m) \left( 1 - \frac{P_m^e(t_f^m)}{P_f^e(t_f^m)} \right)
\]

where \( C_f(t_f^m) = \sum_{i \in \mathcal{I}} o_f P_f^e(t_f^m)(1 - P_f^e(t_f^m)/P_f^e(t_f^m)) \) can be considered as the utility cost on link \( e \). Thus, the \( u'_f(p_f^e) \) can be calculated according to the LFPs on link \( e \) and the information of the detected anterior route.

As illustrated in Eq. (15), an interference weight factor \( \xi \) defined on \([0, 1]\) is introduced to the routing metric. The flow utility system has a self-interference characteristic as illustrated in Remark 3; therefore, a trade-off should occur between the new call blocking rate and the network utility. \( \xi \) is supposed to enable a flexible choice between the two parameters to extend the UMR to satisfy a wider user requirement.

Depends on the value of \( \xi \), a different optimization objective and admission principle will be formed in the UMR. For example, with \( \xi = 1 \), the utility is maximized without considering the blocking rate of the new call; a lower \( \xi \) means that the reliability of the new flow is valued more, and more importance is attached to the user perception of the blocking rate in first calling. Furthermore, if \( \xi = 0 \), the route that maximizes the utility of the new flow will be selected, and the admission constraint principle deteriorates, the selected route must at least have a probability of surviving during the flow duration, which is much looser than the case with \( \xi = 1 \).

4.2. UMR routing scheme

To reduce the overhead for information collection in the MS-LEO satellite networks and improve the accuracy of the information, the route flooding scheme is employed in the UMR. The objective of the UMR can be described as searching for a path that maximizes the total path routing metric, while subjecting to the following constraints:

(1) Providing sufficient bandwidth resource for the delay bound guarantee at present.

(2) Having a non-negative path routing metric. The second constraint can be considered as a route admission principle. The routing scheme on nodes is outlined in Fig. 1, and the routing process is presented in Fig. 2 (RDP means route detecting packet, RRP means route responding packet).

4.2.1. On source node

Once a route application is received, it generates a RDP, including flow characteristics and delay bound, and sends the RDP to all link neighbors to initiate a route detecting process. Once a RRP is received, it extracts the selected path and initiates flow transmission.

4.2.2. On forwarding nodes

Once a RDP is received on a forwarding node \( n \), a collection of the possible next-hop links \( E_n \) is formulated by adding all the

---

**Fig. 1** Routing scheme of UMR algorithm.
available output links on n except for the reverse link of which RDP is received. Then, the following routing steps are implemented for each \( e \in E_{n} \).

**Step 1.** Compute \( g'(p_f^i) \) according to Eq. (4). Check whether the path pruning constraint \( \eta_{p_f^i}(t_f^i) \geq g'(p_f^i) \) is satisfied, where \( \eta_{p_f^i}(t_f^i) = \min(\eta_{p_f^i}(t_f^i), \eta_e(t_e^i)) \) is the minimum link residual resource on \( p_f^i \). If the constraint is not satisfied, stop the operation and delete link \( e \) from \( E_{n} \).

**Step 2.** Extract flow characteristics from RDP. Use \( g'(p_f^i) \) to substitute the \( g'(p_f^i) \) approximately and compute the \( P_{f}^i(t_f^i) \) according to the LFP formulation model in Section 2 on the basis of the local information. Update \( P_{f}^i(t_f^i) \) for flow \( i \in \Gamma_e(t_e^i) \) based on the \( g'(p_f^i) \).

**Step 3.** Extract the detected anterior route information \( u_{f}^i(p_f^i) \) and \( \prod_{e \in \partial_f \Gamma_e} P_{m}^e(t_e^i) \) from RDP. Query \( P_{f}^i(t_f^i) \) and \( P_{f}^i(t_f^i) \) from the local flow information. Generate routing metric \( u_{f}^i(p_f^i) \) according to Eq. (17).

**Step 4.** Check whether the path pruning constraint \( u_{f}^i(p_f^i) > 0 \) is satisfied and if not satisfied, stop the operation and delete link \( e \) from \( E_{n} \).

**Step 5.** If link \( e \) is qualified, update the detected anterior route information in the RDP such as \( \prod_{e \in \partial_f \Gamma_e} P_{m}^e(t_e^i) \), \( u_{f}^i(p_f^i) \), and \( \eta_{p_f^i}(t_f^i) \) and send the RDP to the neighbor node on link \( e \) for a next-hop route detection.

Once a RRP is received from the neighbor node of link \( e \), the forwarding link \( e \) has been selected to constitute the \( p_f \). Then, the flow characteristics and the path characteristics such as \( P_{f}^i(t_f^i) \) and \( P_{f}^i(t_f^i) \) are stored; the scheduling weight is allocated; and the RRP is sent to the upstream neighbor according to the \( p_f \).

### 4.2.3. On destination node

Each RDP that arrived at the destination node is an alternative path complying with the routing constraints. Because \( u_{f}^i(p_f^i) \) can be considered as the equivalent measurement of the network utility, the network utility maximization is achieved by selecting a path with a maximal \( u_{f}^i(p_f^i) \). The path selected will be recorded on the destination node, and a RRP will be generated and feedback along the selected path for path confirmation and resource allocation.

### 5. UMR simulation

The simulation result illustrates the improved performance of the UMR algorithm in terms of the load balancing performance, the delay interruption rate, and the network utility. The simulation also reveals the operational principle of the utility-maximizing routing metric and the admission control principle in the UMR.

#### 5.1. Simulation environment and algorithms

The simulation environment was built based on the C++. The basic topology parameters are shown in Table 1, and the environment was a typical LEO Walker constellation with ISL number and basic capacity similar to the Iridium system. Several real and assumed models were defined to simulate the node behavior in a MS-LEO satellite network. The region

---

**Table 1** Network topology parameters in simulation.

<table>
<thead>
<tr>
<th>Constellation structure</th>
<th>Constellation altitude (km)</th>
<th>ISL number per node</th>
<th>ISL basic capacity (Mbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walker &lt; 14.9 ISL</td>
<td>1400</td>
<td>4</td>
<td>20</td>
</tr>
</tbody>
</table>
within the latitude of $5^\circ$ to the polar cycle was defined as the initiate link-off location. Each link had an extra preset link-off time, which was generated according to a uniform distribution in the simulation time. The handovers of ISLs were implemented in the region of a $5^\circ$ latitude after the equidistant of handover links. To simplify the simulation, a unified and unchanged maximum link capacity of 20 M was used, with a power saving plan initiated randomly on each link to reduce the link capacity to 12 M periodically. Link capacity jitter was triggered randomly in time interval $[k \times 10, k \times 10 + 2]$ s ($k = 1, 2, \ldots$); and the jitter amplitude was in accordance with the uniform distribution in $[2, 8]$ M; the jitter duration was 2 s.

Users were distributed according to the host distribution in Table 2. Users kept access to network via the nearest satellite in coverage. Source users and destination users were randomly selected for each call. Three flow types were adopted with a different characteristic and priority as shown in Table 3, occupying 0.3, 0.4, and 0.3 of the total flow. The delay constraint was 400 ms for a typical IP voice service. The target flow number (TFN) varied from 100 to 1300 to test the performance on different load levels, and calls were generated in a cycle manner to keep the TFN in network.

Three routing algorithms were involved in the simulations of the minimum hop routing algorithm\(^\text{16}\) (MHA), the shortest-widest routing algorithm\(^\text{17}\) (SWP), and the UMR algorithm. All the algorithms provided delay bound guarantee based on the ideal GPS scheduling scheme and identified route with the bandwidth constraint. The UMR algorithm was simulated with $\zeta = 0$ and $\zeta = 1$. The LFP and $P_f(t_f)$ were not updated in real-time.

5.2. Simulation result and analysis

5.2.1. Load distribution characteristics

Figs. 3–5 illustrate that the UMR algorithm had the lowest statured link proportion due to its lower load level and improved load balancing performance.

Fig. 3 shows the stature link proportion (SLP), where $x$ is the link utilization threshold of the network saturation judgment. The UMR algorithm had the lowest SLP. When $\text{TFN} < 400$, the UMR algorithm had an SLP similar to the SWP, and the MHA algorithm had an SLP larger than others. When $\text{TFN} \geq 400$, the UMR algorithm had the lowest SLP, whereas the SWP algorithm had an SLP approximately 10% larger than the MHA algorithm, which shows that the longer route requires significant additional resources for delay bound guarantee and forms a larger SLP when the network is almost statured. As expected, the UMR algorithm for $\zeta = 1$ had an SLP less than that for $\zeta = 0$.  

### Table 2 Internet host distribution by continent\(^\text{15}\) (January 2005).

<table>
<thead>
<tr>
<th>Continent</th>
<th>Number of hosts (10(^3))</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>223545.1</td>
</tr>
<tr>
<td>South America</td>
<td>6026.2</td>
</tr>
<tr>
<td>Europe</td>
<td>52947.1</td>
</tr>
<tr>
<td>Africa</td>
<td>5621.6</td>
</tr>
<tr>
<td>Asia</td>
<td>28511.4</td>
</tr>
<tr>
<td>Oceania</td>
<td>671.3</td>
</tr>
</tbody>
</table>

### Table 3 User type and characteristics.

<table>
<thead>
<tr>
<th>User type</th>
<th>Average data arriving rate (Mbit/s)</th>
<th>Maximum burst length (kB)</th>
<th>Maximum packet length (B)</th>
<th>Delay bound (s)</th>
<th>Priority level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.72</td>
<td>24</td>
<td>200</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.96</td>
<td>48</td>
<td>200</td>
<td>0.4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1.44</td>
<td>72</td>
<td>200</td>
<td>0.4</td>
<td>3</td>
</tr>
</tbody>
</table>

Fig. 3 Stature link proportion.

Fig. 4 Average utilization rate of link resources.

Fig. 5 Standard deviation of link traffic distribution.
Fig. 4 illustrates that the network obtained the lowest average utilization rate of link resources with the UMR algorithm. The stature load of the UMR algorithm was 10–20% less than the other algorithms for $\zeta = 1$ and was similar to the MHA algorithm for $\zeta = 0$, which suggests that a more stringent admission principle plays a significant role in imposing an extra restriction on the traffic load level.

The load balancing performance is indicated by the standard deviation of link traffic distribution in Fig. 5, where the following results were obtained:

1. When $\text{TSN} < 400$, the UMR algorithm had a load balancing performance better than the MHA algorithm and worse than the SWP algorithm.
2. The UMR algorithm improved the load balancing performance versus the SWP algorithm when $400 < \text{TSN} < 800$ because the admission principle affected the load distribution by refusing the flows that can only build routes on stature links.
3. The UMR algorithm for $\zeta = 1$ distributed routes in a less balanced manner than the other algorithms and remained nearly unchanged when $800 < \text{TSN} < 1300$ because the admission principle rejected most flows and the network entered into saturation as shown in Fig. 2.

5.2.2. Average hop number on the route

The average hop number of the path is illustrated in Fig. 6, suggesting that the UMR algorithm selected the flows and routes in a flexible manner based on the load level. When $\text{TFN} < 400$, the UMR algorithm formed routes that were longer in hops than the MHA algorithm and shorter in hops than the SWP algorithm. When $\text{TFN} > 400$, the UMR algorithm for $\zeta = 0$ had an average hop number of path similar to the MHA algorithm, and the UMR algorithm for $\zeta = 1$ had an average hop number of path that was clearly lower than the other algorithms, indicating that only a considerably shorter route satisfies the strict admission constraint when the network is nearly saturated.

5.2.3. Interruption rate and new call blocking rate

Fig. 7 illustrates that when $\text{TFN} > 300$, the UMR algorithm obtained the lowest delay QoS interruption rate and that the SWP algorithm obtained the largest. The UMR algorithm for $\zeta = 1$ had an excellent performance ($< 3\%$) on the interruption rate in the entire simulated load range, which is approximately 15–20% less than the MHA algorithm and 25–30% less than the SWP algorithm on a heavy load. Additionally, the UMR algorithm for $\zeta = 0$ had a 5–8% reduction for the interruption rate from the MHA algorithm. This better performance was achieved by reducing the probability of the routes being affected by an unstable link connection and capacity; (1) routes were distributed on more reliable links using the utility-maximal metric; (2) the statured link proportion was reduced; and (3) lower route hop numbers reduced the probability of the routes being affected.

Fig. 8 shows that the UMR algorithm for $\zeta = 1$ had the largest new call blocking rate as an inevitable cost for the strict route admission principle, and the UMR algorithm for $\zeta = 0$ had a new call blocking rate 5–10% less than the UMR algorithm for $\zeta = 1$. Thus, as shown in Figs. 5–7, a compromise was reached between the new call blocking rate and network utility using $\zeta$. The SWP algorithm had a new call blocking rate similar to the MHA algorithm, indicating that its higher resource cost will result in the degradation of the interruption rate and the new call blocking rate. Therefore, the classical load balancing routing approach similar to the SWP could not be combined with the weighted fair scheduling schemes to provide a delay bound guarantee in wireless network with unstable links because a better performance than the simple MHA algorithm may not be achieved.

5.2.4. Network utility

The network utility in Fig. 9 was expressed using the success flow number (SFN), and the priority weighted utility (PWN) was defined in Section 2. When $\text{TFN} > 300$, the UMR algorithm increased the SFN by 5–10% on the SFN for $\zeta = 0$ and by 10–20% for $\zeta = 1$, which increased the PWN by approximately 5–30% and 25–35%, respectively. The increase
in the SFN and the reduction of interruption rate did not completely conform when the load was heavy, which may be due to the high interruption rate and the new call blocking rate, causing the actual accessed flow number to always be slightly less than TFN and for the flow generation and routing to take longer in the simulation. However, the lower interruption rate increased the SFN, and the consideration on flow priority in the UMR algorithm resulted in a more significant improvement in PWN. A smaller $\xi$ reduced the network utility.

6. Conclusions

In this paper, a routing algorithm was proposed to improve the network utility for the delay-sensitive flows in the MS-LEO satellite network. The simulation results show that when the traffic load is heavy, the UMR algorithm greatly reduced the delay interruption rate and increases the network utility. The results also indicate that the utility advantage of the UMR algorithm is obtained at the cost of a higher new call blocking rate, and the interference weight factor provides a flexible choice between the two performances. Therefore, the UMR algorithm is considered a promising choice for providing the delay QoS integrated service in the MS-LEO satellite network. The simulation results show that when the traffic load is heavy, the UMR algorithm greatly reduced the delay interruption rate and increases the network utility.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Nos. 61171069, 61231013) and Innovative Research Groups of the Foundation of China (No. 60921001).

References


Lu Yong is a Ph.D. candidate of the School of Electronics and Information Engineering, Beihang University. Her main research interest lies in satellite communications network.

Zhang Jun is a professor and Ph.D. supervisor at School of Electronics and Information Engineering, Beihang University. His main research interest lies in space integrated network and air traffic management.

Zhang Tao received Ph.D. degree from School of Electronics and Information Engineering, Beihang University. His main research interest lies in satellite communications network.