Optimal Model and Algorithm of Cross-layer Dynamic Resource Allocation based on Multi-Period and Utility

Qing Shao* and Yang Su
School of Communication and Informatics Engineering, Xi’an University of Science and Technology, Xi’an, China

*Corresponding author email: shaoqing9566@163.com

Abstract. An optimal model and algorithm for dynamic resource allocation in satellite communication systems is proposed to further promote the system throughput and fairness based on user satisfaction. It takes the parameters of application layer, data link layer and physical layer into consideration and constructs a new kind of utility function by using the cross-layer design strategy and converts the resource allocation problem into a utility maximization model. The optimal solution of the model is solved by the combination of the Lagrange multiplier method and the Greedy algorithm, which maximizes the utility and ensures the optimal resources allocation of the satellite communication system in the multiple periods. The simulation results show that the proposed optimal model and algorithm effectively promote the throughput and fairness based on user satisfaction on the premise of ensuring service quality, therefore it can meet the requirements of the variety and the dynamic change of the service in the satellite communication system.

Keywords: Satellite communication; utility function; cross-layer design; dynamic resource allocation.

1. Introduction
One of the goals of satellite communications is to provide high-speed communication for rapidly transmitted Internet applications and multimedia technologies. But the cost of satellite network is expensive and the on-board resources are limited, which puts forward higher requirements for satellite resource management. The European Telecommunications Standards Association has developed standard for Satellite communications systems [1] (Digital Video Broadcasting-Return Channel by Satellite Second generation/Second generation Framing, DVB-RCS2/S2). The standard defines technical specifications for the physical layer, but does not specify resource allocation and management policies. Therefore, under the condition of limited resources, the dynamic resource allocation strategy of satellite communication system has become a research hotspot. Ruian W puts forward a method based on experience quality, and introduces economic link utility function for the first time[2]. The user experience quality is a combination of link cost and reward, which improves the user experience quality, but does not consider the reasonable utilization of resources. SU J T proposes a method based on stackelberg differential game [3], which makes a improvement in such aspects as service quality, but does not consider the improvement of throughput. Traditional algorithms do not consider the interaction between each layer in the network protocol layer and cannot achieve the optimization of the entire network performance. The cross-layer design thought consider the relevant information in each layer and adjust the resources adaptively according to the dynamic network, so as to realize the overall optimal allocation of network resources. Due to the
above advantages, a resource allocation algorithm based on nonlinear programming is proposed [4], which emphasizes the application layer on slot allocation, and improves user satisfaction by setting Quality of Service (QoS) requirements. On the basis of [4], [5] proposes an improved strategy, which adopts relative value coefficient to deal with link state and user service priority, effectively improving system throughput. However, the current research only focuses on the allocation of slot resources in a single hyperframe (i.e. a single period) without considering the impact of multi-period allocation.

In order to make the allocation of resources within multi-period more reasonable, this paper introduces the idea of cross-layer design, considers the parameter information between the application layer, the physical layer and the data link layer, constructs the utility function reasonably through the parameter information in each protocol layer, and transforms the resource allocation problem into the utility maximization model. Then, the maximum utility of the system is obtained by combining the Lagrange multiplier method with the Greedy algorithm, which ensures the overall optimal allocation of satellite resources in multiple periods.

2. DVB-RCS2/S2

DVB-RCS2/S2 defines Satellite communication system, whose overall architecture consists of satellites, Gateways, Network Control Center (NCC) and Return Channel Satellite Terminals (RCSTs)[6]. Each terminal is connected to the system by means of multi-Frequency Time Division Multiple Access (MF-TDMA). According to the requirements of DVB-RCS2/S2, the user terminal sends the resource request to NCC through the back link, NCC collects the request and executes the corresponding algorithm for optimal allocation of system resources. Then, the result is fed back to the corresponding terminal through the terminal burst time schedule, so that the terminal can transmit the data. This paper focuses on the resource allocation problems (especially the time slot request of each user) under the condition of resource constraint.

3. Cross-layer Dynamic Resource Allocation Algorithm

Due to the limited resources in the satellite communication system, operators need to improve the system throughput in order to obtain more returns, which will inevitably affect the service quality of users. In this paper, a cross-layer dynamic resource allocation algorithm is proposed based on Multi-Period and Utility (MPU), whose main purpose is to make a balance between the high system throughput and the high quality of service to users.

3.1. Cross-layer Design

The communication of traditional hierarchical network protocol happens between two adjacent protocols and the information is not allowed to be shared in other layers, which results in non-optimal and inflexible problems. In order to solve the contradiction between high demand of user terminal and limited available resources, the concept of cross-layer design is introduced in the allocation strategy. The basic idea is to make effective connections of the information in each layer through a cross-layer design module. This module can pass key parameters between different layers, allowing non-adjacent layers to directly communicate messages or share state variables, so that parameters can be selected, passed and optimized among different layers. The cross-layer idea adaptively makes global changes according to the practical application requirements and network state, which alleviates the contradiction between communication requirements and transmission environment restrictions[7]. The current cross-layer design includes interactions between the five main protocol layers, as shown in Figure 1. This paper focuses on the application layer, physical layer, data link layer, and constructs the cross-layer dynamic resource allocation scheme as shown in Figure 2.
3.2. Utility Maximization Model of MPU Algorithm

In order to verify the feasibility of the above scheme, Network Utility Maximization (NUM) model is introduced by mathematical modeling. Utility function can quantify the system gain brought by resource allocation, and the higher the value is, the better the performance of the system will be. In the MPU algorithm, the utility function in the model is reasonably constructed by comprehensively considering the number of time slot requests from each user terminal, the available resources of the system and the relevant parameters between each protocol layer.

- Set the service QoS level $q_i$ according to the application layer.
- Modulation order $M_i$ and coding rate are $R_i$ set according to the physical layer [8].
- Set $g_i$ according to the link layer queue state information. $g_i$ represents the number of slots in the allocation that terminal i is not satisfied with.
- Set the minimum guaranteed slots to meet the basic communication requirements of the users.

The utility function is defined as Equation (1):

$$u(x_i) = \left[ w_1 \frac{q_i}{\sum_{i=1}^{N} q_i} + w_2 \frac{M_i R_i}{\sum_{i=1}^{N} M_i R_i} + w_3 \frac{g_i}{\sum_{i=1}^{N} g_i} \right] \ln(x - d_i)$$

Where, $N$ is the total number of terminals, $u(x_i)$ is the utility function of terminal i, $x_i$ is the number of time slots obtained by terminal i, $w_1$, $w_2$ and $w_3$ are relative value coefficients and $w_1 + w_2 + w_3 = 1$. The value of $w$ coefficient can be flexibly changed to meet different user requirements. $w_1$ can be adjusted for bandwidth allocation based on the priority of users, $w_2$ can be adjusted for communication quality requirements, and $w_3$ can be adjusted for distribution fairness. In the simulation, $w_1 + w_2 + w_3 = 1/3$ is taken to indicate that the weight occupied by each layer is the same. According to the requirements of communication, the number of time slots allocated by terminal i is limited by the number of applied time slots $D_i$ and the number of minimum guaranteed time slots $d_i$, and the NUM model is shown as follows, where $B$ is the number of available slots in the system.

$$\begin{cases} \max \sum_{i=1}^{N} u(x_i) \\ d_i \leq x_i \leq D_i \\ \sum_{i=1}^{N} x_i = B \end{cases}$$
3.3. MPU Algorithm Implementation Process

**Step 1:** Determine whether available slots can satisfy all user's requests. If so, it means that resources are abundant at this time and can be allocated to each terminal in turn; If not, enter Step 2.

**Step 2:** Determines whether available slots meet the minimum guaranteed time slot requirements of all users. If not, it means that resources are extremely scarce, and the basic communication of high-priority service can be guaranteed in priority. If so, enter Step 3.

**Step 3:** Until this step, it means that the available slots cannot meet the request time slot of all users, but can meet the minimum guaranteed slots, so the resource optimization management strategy needs to be implemented.

First, define the utility function and build the NUM model according to the formula (1) and (2).

Then, the maximum value of utility is obtained according to the nonlinear integer programming theory.

In order to reduce the complexity of the algorithm, the Lagrange multiplier method is used to solve the problem, as shown in Equation (3).

\[
zi = \left[ w_1 \frac{q_i}{\sum_{i=1}^{N} q_i} + w_2 \frac{M_i R_i}{\sum_{i=1}^{N} M_i R_i} + w_3 \frac{g_i}{\sum_{i=1}^{N} g_i} \right] \left( B - \sum_{i=1}^{N} d_i \right) + d_i \quad (3)
\]

Finally, if there is a surplus time slot, the remaining resources is reallocated using the Greedy algorithm. As for the unsatisfied terminals, the utility increment value when allocating a single time slot is calculated respectively, as shown in Equation (4).

\[
\Delta u(x) = u(x+1) - u(x) = \left[ w_1 \frac{q_i}{\sum_{i=1}^{N} q_i} + w_2 \frac{M_i R_i}{\sum_{i=1}^{N} M_i R_i} + w_3 \frac{g_i}{\sum_{i=1}^{N} g_i} \right] \ln(1 + \frac{1}{x-d_i}) \quad (4)
\]

A single slot is allocated for the terminal with the largest utility increment value in the calculation result and its utility increment value is updated. Repeat the above steps and enter Step 4 after all slot assignments are completed. After each assignment, constraints in the NUM model should be considered to ensure the correctness of the solution.

**Step 4:** Each user judges whether the allocation result satisfies the number of request slots. If it does, the allocation will be terminated. If not, the difference between the allocation result and the number of request slots will be recorded to prepare for the construction of utility function next time.

4. Algorithm Performance Simulation and Analysis

In this chapter, the performance of MPU algorithm was verified by simulation. Max-Min algorithm [9] and cross-layer fairness water injection algorithm [10] were selected for simulation, and the feasibility of MPU algorithm was analyzed.

4.1. Simulation Scene

System simulation parameters are set in reference [10].

**Table 1.** Modulation mode and order.

| Modulation method | Modulation order |
|-------------------|------------------|
| BPSK              | 1                |
| QPSK              | 2                |
| 8PSK              | 3                |
| 16ASK             | 4                |

**Table 2.** QoS priority parameters.

| Service type      | QoS |
|-------------------|-----|
| VoIP              | 2   |
| Video streaming media | 1.75 |
| Telent            | 1.5 |
| Web               | 1.25 |
| FTP, SMTP         | 1   |
Table 3. Simulation parameters.

| User no. | $M_i$ | $R_i$ | $D_i$ | $d_i$ | $q_i$ |
|----------|-------|-------|-------|-------|-------|
| 1-2      | 1     | 1/2   | [15,16]| [2,0] | [1.75,1.25] |
| 3-6      | 1     | 2/3   | [9,19,14,5]| [0,1,2,0] | [1.5,2,1.25,1.75] |
| 7-13     | 2     | 3/4   | [17,13,4,5,13,13,8]| [1,2,2,0,1,1] | [1.75,2,1.5,2,1.25,1,5,1.25] |
| 14-20    | 3     | 4/5   | [12,10,2,2,7,1,8]| [1,0,2,2,1,2] | [1.5,1,5,1,75,1,75,2,1,25,1] |
| 21-24    | 4     | 7/8   | [14,3,2,13]| [0,1,2,3] | [1.5,2,1,1.75] |

4.2. Performance Parameters

- **Throughput.** The throughput is determined by the modulation order, coding code rate and the number of time slots allocated to all users, as shown in Formula (5).

$$T = \sum_{i=1}^{N} x_i \cdot M_i \cdot R_i$$ (5)

- **Absolute fairness.** The absolute fairness standard requires that the number of time slots allocated to each user should be the average of all resources to reflect the absolute fairness of the whole system, as shown in (6)

$$f = \frac{\left( \sum_{i=1}^{N} x_i \right)^2}{N \sum_{i=1}^{N} x_i}$$ (6)

- **Based on the fairness of user satisfaction.** From the perspective of relative fairness, considering users’ satisfaction with the distribution results, the fairness based on user satisfaction is defined, as shown in Equation (8).

$$S_i = \frac{x_i}{D_i + g_i}$$ (7)

$$S = \frac{\sum_{i=1}^{N} S_i}{N \sum_{i=1}^{N} S_i}$$ (8)

4.3. Simulation

4.3.1 **Time slot allocation results.** As shown in Figure 3, Max-Min algorithm distributes slots almost uniformly, ensuring absolute fairness. Fairness Water-filling algorithm only considers the service types in the application layer, so more slots are allocated to users with higher priority. The MPU algorithm not only guarantees the priority allocation of high-priority service, but also considers the channel state information of physical layer and allocates more slots for good channel conditions of 14–24. When the channel state is good, users can transmit more data within the unit symbol, so as to improve the throughput of the system.

![Time slot allocation results](image)

**Figure 3.** Result of slot allocation.
Table 4. Simulation result.

| Allocation algorithm | Throughput | Absolute fairness | Utility |
|----------------------|------------|-------------------|---------|
| MPU                  | 196.367    | 0.834             | 29.188  |
| Max-Min              | 170.033    | 0.922             | 27.286  |
| Fairness Water-filling | 174.866    | 0.850             | 28.364  |

4.3.2 System performance. As shown in Table 4, Max-Min algorithm pays attention to the average of allocation, so absolute fairness is the highest and throughput is the lowest. Compared with Max-Min algorithm, the throughput of cross-layer Fairness Water-filling algorithm is improved, but there is no significant increase. The system throughput and utility values of MPU algorithm are the highest, although at the expense of absolute fairness. In practical, it is not expected that the distribution of each service type is absolutely fair, so the algorithm is more in line with the actual communication requirements.

4.3.3 System throughput. In order to compare throughput more clearly, the number of allocable slots in the system is extended. Figure 4 shows that the MPU algorithm has the maximum throughput. MPU algorithm comprehensively considers parameters in the three protocol layers, and makes global adaptive changes according to the number of allocable slots and user requirements. Therefore, when resource allocation is relatively tight, adopting MPU algorithm can bring higher throughput.

4.3.4 Influence of multiple periods on the algorithm. In order to evaluate the multi-period optimization algorithm proposed in this paper, continuous slot allocation over 40 cycles is simulated. As shown in Figure 5, MPU algorithm improves fairness based on user satisfaction in multiple cycles. When MPU algorithm is adopted, each time slot resource allocation will consider the request of the user that has not been satisfied before, so as to ensure the relative fairness of the time slot obtained by the user. The value remains relatively flat in multiple cycles, which indicates that the overall resource allocation system reaches a relatively stable state.

5. Summary
This paper studies the DVB-RCS2 /S2 satellite communication system, analyzes the existing resource allocation strategy, and proposes the MPU algorithm which takes the parameters of application layer, data link layer and physical layer into consideration, and ensures the optimal allocation of satellite resources in multiple cycles. Then the system throughput, utility value, absolute fairness and fairness based on user satisfaction are simulated, which shows that the proposed MPU algorithm can not only
improve the throughput of the satellite communication system, but also improve the fairness based on user satisfaction, which makes the allocation of the limited resources more efficient and reasonable.

References

[1] Murillo Y, Beysens J, Bertrand, A. et al. Multidisciplinary learning through implementation of the DVB-S2 standard [J]. IEEE Communications Magazine Articles News & Events of Interest to Communications Engineers, 2017, 5: 124-130.

[2] Ruian W, Jinchun G, Gang X, et al. QoE-aware bandwidth allocation method with GEO satellites cooperation in distributed constellation network[C]// International Conference on Wireless Communications & Signal Processing. Nanjing, IEEE, 2015: 1-5.

[3] Jingtao Su, Shengsong Yang, Haitao Xu, et al. A stackelberg differential game based bandwidth allocation in satellite communication network[J]. China Communication, 2018, 15(08): 205-214.

[4] Zhang Di. A Cross-layer dynamic bandwidth allocation algorithm based on nonlinear programming in DVB-RCS [J]. Computer Engineering & Software, 2015, 36(09): 92-97.

[5] Zhou Weibing. Research of resource management and access technology based on satellite network [D]. Beijing: Beijing university of posts and telecommunications, 2018.

[6] Youssef B, Khalid S, Hajar C, et al. Measurement and test of a DVB-S2 satellite broadcast [C] // 2019 7th Mediterranean Congress of Telecommunications (CMT). 2019.

[7] Khalil A M, Zada K W, Tariq U, et al Editorial of cross-layer design issues, challenges and opportunities for future intelligent heterogeneous networks[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10: 4207-4208.

[8] Tao Kai, Sun Xiaoyun, Wang Yang, et al. System Modeling and Precoding Design for Multi-beam Dual-polarized Satellite MIMO System[J]. 2019, 13(4):374-381.

[9] Liu Hongying, Yang Zhiming, Cao Zhigang, et al. Max-Min rate control on traffic in broadband multibeam satellite communications systems [J]. IEEE Communications Letters, 2013, 17(7): 1396-1399.

[10] Zhou Wu, Zhang Qi, Tian Qinghua, et al. Cross-layer dynamic bandwidth allocation algorithm based on convex optimization theory in satellite communication system[C]// International Conference on Optical Communications & Networks. Hangzhou, IEEE, 2017: 1-3.