Received August 20, 2021, accepted August 31, 2021, date of publication September 3, 2021, date of current version September 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3109876

Estimation of Node Cache Occupancy in Satellite Storage Network

ERBAO WANG, QUNFENG DONG, YONGAN LI, AND YUYE ZHANG

School of Physics and Electronic Engineering, Xianyang Normal University, Xianyang 712000, China

Corresponding author: Erbao Wang (wangebxd@163.com)

The work of Erbao Wang was supported in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2021JM-517.

ABSTRACT The satellite network with cache function becomes the future development trend. How to efficiently use the node cache of the satellite network has become a research hotspot. A theoretical problem that needs to be solved in the latest research on effective utilization of cache is the estimation of cache occupancy. Based on the martingale encapsulation theory, combined with the channel conditions, access protocols, and queuing strategies in the satellite storage network, a satellite node occupancy estimation method and theoretical bounds are proposed in a two-layer satellite network of GEO and LEO. The accuracy of the martingale encapsulation method in estimating the cache occupancy is estimated by simulations. The simulation results show that our method is precise. The cache occupancy estimation method we proposed provides a theoretical basis for the optimal allocation of satellite cache resources.

INDEX TERMS Cache occupancy estimate, martingale theory, equivalent arrival martingale, service martingale, satellite storage networks.

I. INTRODUCTION

At present, the satellite storage network can be regarded as the further development of the satellite communication system, and finally use the information center network (ICN), the ground network, and the on-board cache [1]–[3] forms a multi-level, heterogeneous and three-dimensional storage network system. The satellite storage network can provide better technical support for satellite navigation, digital maps, ocean exploration, and data communication in remote areas. As shown in Figure 1, the satellite storage network is a distributed system that uses a space platform as a carrier to optimize the placement, processing and transmission of massive and multi-dimensional information.

The development of satellite storage networks is closely related to the development of the Internet and ground mobile communication networks. Earlier research on satellite storage networks is based on the Internet [4]–[8]. Early satellite storage networks used the broadcast content of high-orbit satellites to distribute data to reduce the delay of data reception. The satellite storage structure at this stage is only a single-layer satellite network structure, and it uses high-orbit satellite broadcasting. This form of broadcasting is likely to cause infection of other users' data. In the second stage, satcache is the representative of the satellite storage network [9]–[11] based on the Ground Information Center Network (ICN). Satcache proposes to estimate the potential interest of content and create the preferences of network users, and then store the content of other users' interest on the nearest satellite terminal. This method makes full use of

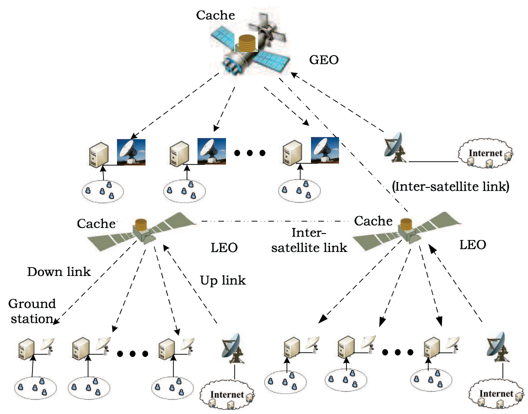

FIGURE 1. Structure diagram of the future satellite storage network.

The associate editor coordinating the review of this manuscript and approving it for publication was Chi-Tsun Cheng.

the satellite communication broadcasting and ground information center network ICN.

In the third stage, the combination of satellite cellular and terrestrial wireless communication networks allows users to seamlessly access rich multimedia content (such as TV broadcasting, VOD streaming and other services) on their devices. The fourth stage of development is based on on-board cache. In [3], a two-layer cache model is proposed, in which the cache placed on the ground station constitutes the first layer of cache, and the cache deployed on the satellite constitutes the second layer of cache. To minimize the bandwidth consumption of downlink and uplink satellites, a joint cache optimization [3] between satellite and ground station is proposed, in which the joint cache optimization problem is described as a nonlinear integer programming problem. In traditional satellite ground networks, it is assumed that only ground stations have caching capabilities. Each ground station caches the most popular contents in its local area to meet the needs of local users. In [12], a cache layout algorithm with fast convergence speed and low complexity based on the assumption that the satellite has a cache function is proposed. The two-layer network architecture of the above two references refers to satellites and ground users. The satellites here are only single-layer satellites, and there are no inter-satellite links between satellites of different heights. In addition, the analysis of cache occupancy is not fully considered when optimizing the cache.

Currently, single-layer LEO networks mainly include OneWeb, IridiumNEXT, and Starlink. IridiumNEXT, based on the first-generation Iridium satellite system, is currently developing rapidly. In September 2020, the twelfth batch of 60 satellites was successfully launched in the second-generation Iridium system. Therefore, we use the Iridium satellite system as the low-orbit satellite system, combined with the wide coverage of the high-orbit satellite to form the two-layer satellite network structure shown in Fig. 1.

This paper proposes a method to estimate the remaining cache of nodes using martingale theory, which provides a theoretical basis for the future design of the content placement research of satellite storage network nodes.

II. RELATED WORK

The cache occupancy[1] analysis of wireless networks mainly uses the queuing theory to obtain the average queue length. However, the arrival process is not necessarily the Poisson arrival, so the queuing theory cannot be used to obtain the average queue length. In addition, even if the arrival process is a Poisson process, the queuing theory can be used to obtain the average queue length. However, sometimes the focus is not on the average queue length but the maximum queue length. The earliest theory used to solve this type of problem is the network calculus [13]. In [13], the delay bounds is obtained in the Internet network by using the network calculus method, which is called the deterministic network calculus analysis method. This method for deterministic network calculation defines both business arrival and business departure as affine function or as a sum of multiple affine functions, and the cache backlog at any time can be calculated based on the arrival curve and the departure curve graph. However, in the satellite storage networks, the satellite service arrival and the channel access are random, so this theory is not suitable for the situation where arrive and leave services are random. In addition, this type of deterministic network calculation cannot take advantage of the statistical characteristics of the service, so random network calculations appear [14], [15]. Reference [15] discussed the application of the martingale theory to the analysis of the cache backlog boundary under several scheduling mechanisms. In [16], the upper martingale theory was used to solve the cache backlog bounds under several wireless access mechanisms. The advantages of the martingale theory are as follows. First, the martingale theory can convert multiplexing of streams into a multiplication of the corresponding martingales. Second, the martingale theory can convert flow scheduling into the martingale time shift. Compared with the queuing theory, this method does not assume that data arrival obeys Poisson distribution because the general data arrival does not necessarily follow the Poisson distribution, so the assumption of Poisson distribution is too harsh. In addition, the goal of queuing theory is to obtain the average queue length. However, the average queue length can hardly reflect the changing trend of the cache queue, that is, the instantaneous change in the cache queue is difficult to measure. Nevertheless, following the martingale encapsulation theory, it is possible to design the martingale encapsulation functions of the channel access process, channel fading process, and queuing process and use them to obtain the probability distribution function[2] of the cache occupancy and estimate the cache occupancy accordingly. In [17], a delay analysis of a multi-hop vehicle network based on martingale encapsulation theory is presented. In [18], under different scheduling strategies, multiple high-speed train network delay analysis results based on martingale encapsulation theory are presented.

The above two references use martingale theory to solve the problem of delay analysis of vehicle network. The main use environment of our article is a multi-layer satellite storage network, that is, the network environment is different. In addition, we mainly uses the martingale theory to estimate the problem of cache occupancy.

III. SYSTEM MODEL AND FUNDAMENTAL THEORY

A. SYSTEM MODEL

The inter-satellite link is highly invulnerable and can be independent of the ground for networking. In future

[1]Cache occupancy (backlog), that is, if there is only one queue in the node, then it is the length of the queue, otherwise, it is the accumulation of instantaneous data in the cache.

[2]The probability of violation is that given a cache backlog value, the data in the cache may be larger or smaller than this value in an actual environment. The probability distribution of the backlog in the cache can be obtained through statistical analysis, and the probability of violation can be obtained through this probability distribution.

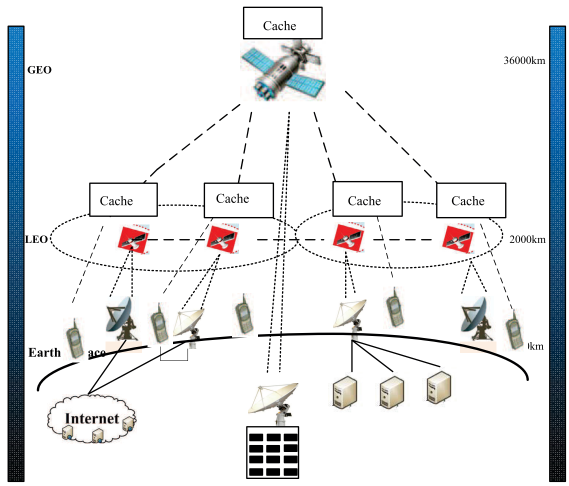

FIGURE 2. Schematic diagram of multi-layer satellite storage network.

development, most of the inter-satellite links will be used for networking. Therefore, in the system model of this part, this article assumes that the system has a GEO-LEO inter-satellite link. The advantage of this kind of inter-satellite link is to make full use of the inter-satellite links of different layers of satellites. That is, LEO and GEO establish a relay relationship through inter-satellite links, and then rely on GEO's coverage advantages to transmit data to low-latitude areas, and LEO satellites are mainly responsible for higher-latitude communication services.

Fig. 2 shows a multi-layer satellite storage network structure. In this network, a high-level node GEO satellite covers K LEO satellites. Due to the ideal electromagnetic environment at high altitude and little interference, the inter-satellite link of GEO-LEO satellites does not consider signal attenuation in this article, that is, GEO satellites communicate with LEO satellites through an ideal inter-satellite link. In addition, it is assumed that each satellite node is equipped with a cache, and it is assumed that the cache capacity of GEO satellites is larger, and that of LEO satellites is smaller. Since most of the current satellite storage networks are still single-layer networks for specific applications, for compatibility, we assumes that the cache of each layer of satellites is provided for the application of the layer first, and the remaining cache resources can be used by the services of lower-level nodes.

B. SEVERAL MARTINGALE PROCESSES

This subsection introduces the channel model, arrival process, service process, and Zipf distribution. The notations used in the following are listed in Table 1.

1) DOWNLINK WIRELESS CHANNEL MODEL

The transmission rate can be expressed as

$$R_{su} = w_{su} \log\left(1 + \frac{\sqrt{p_s}||h_{su}||^2}{n_0}\right), \tag{1}$$

TABLE 1. Summary of notation.

Notation	Description
$A(n)$	Accumulated amount of arrival data
$A^h(T)$	Accumulation of high-priority arrival
$A^f(T)$	Accumulation of low-priority arrival
$\alpha(n)$	Arrival curve
$B(n)$	Backlog
$\beta(n)$	Service curve
$D(n)$	Output process
f	Content index
F	Number of Content
H_{su}	Channel coefficient
$I(T)$	Indicator function
K	Number of LEO satellites
κ_h	High-priority arrival rates
κ_f	Low-priority arrival rates
κ_{sa}	Access service rate
κ_{sLRU}	LRU schedule service rate
κ_{sLFU}	LFU schedule service rate
κ_s	Queue service rate
κ_{ss}	Equivalent service rate
$M_{A^t}(T)$	Martingale-envelope of arrival process
$M_S(T)$	Martingale-envelope of service process
n_0	Zero-mean additive white Gaussian noise
p_s	Satellite transmitting power
p_{gb}	Transition probabilities from g to b
p_m	Probability of user's transmission
P^{us}	State transition matrix
R_a	Arrival rate
R_{su}	Downlink channel service rate
R_{us}	Uplink channel service rate
s	Coefficient of popularity distribution
$S^q(T)$	Queue service process
$S(T)$	Equivalent service process
$W(n)$	Virtual delay
w_{su}	Bandwidth of satellite-terrestrial link
λ_f	Arrival rate content f
$\sigma_{SP,L}$	SP based Backlog
$\sigma_{FIFO,L}$	FIFO based Backlog

where w_{su} is the bandwidth of a satellite-terrestrial link, and $||h_{su}||^2$ is the channel coefficient. Moreover, the probability density function (pdf) of h_{su}'s [19], [20] can be defined by

$$f_{||h_{su}||^2} = \frac{1}{2b_{su}}\left(\frac{2b_{su}m_{su}}{2b_{su}m_{su}+\Omega}\right)^{m_{su}} \exp\left(-\frac{x}{2b_{su}}\right) \cdot {}_1F_1\left(m_{su}; 1; \frac{\Omega x}{2b_{su}(2b_{su}m_{su}+\Omega)}\right), \tag{2}$$

where ${}_1F_1$ is the confluent hypergeometric function. For $m = 0$, Eq. (3) simplifies to the Rayleigh Probability density function (PDF). For $m = \infty$, it reduces to the Rice PDF. In Eq.(3), Ω is the average power of the line-of-sight (LOS) component, $2b_{su}$ is the average power of the multipath component, and m_{su} is the Nakagami-m parameter. Generally, compared with Rayleigh fading, Nakagami-m distribution can better reflect the real situation of wireless channel fading, and Rayleigh fading ($m = 1$) can be obtained according to the change of m parameters.

2) UPLINK WIRELESS CHANNEL MODEL

In general, the random access protocol is used in the uplink channel. Due to the low efficiency of S-ALOHA, a random access protocol has been used in the early satellite networks. Compared with ALOHA, the access efficiency of S-ALOHA is greatly improved. Therefore, we analyzes the access theory model of S-ALOHA as the uplink channel [21]–[25].[3] The mathematical expression of the successful access channel of a satellite user is as follows:

$$I(T) = \begin{cases} 1, & p_m(1-p_m)^{N-1} \\ 0, & 1-p_m(1-p_m)^{N-1}. \end{cases} \tag{3}$$

In Eq.(3), whether the satellite user can successfully access a channel is defined as an indicative function $I(T)$, where 1 represents successful channel access by a satellite user, and 0 represents the failure of the satellite user to access the channel. In Eq.(3), p_m represents the probability that the user sends the access channel in a slot m, and N represents the number of users accessing the channel at the same time. Next, we introduce two propositions that will be used later.

Proposition 1: Independent identically distributed process is a martingale process.

Suppose $X_1, X_2 \ldots$ is the data arrival process, and the data service process equals the data arrival process, i.e. $A(n) := S(n) := \sum_{k=1}^{n} x_k$. It can be proved here that both $A(n)$ and $S(n)$ are martingale processes, where $\theta > 0$, then $E(e^{\theta X_1}) < \infty$.

Proof: For $\theta > 0$, assume h_a and h_s are the constants and $h_a = h_s = 1$); κ_a and κ_s corresponds to $E(e^{\theta X_1}) = e^{-\theta\kappa_a}$ and $E(e^{\theta X_1}) = e^{-\theta\kappa_s}$ Then, according to the independence hypothesis, the following derivation can be obtained:

$$\begin{aligned} &E\left(h(x_{n+1})e^{\theta((n+1)\kappa_s - S(n+1))}|x_1 \ldots x_n\right) \\ &\quad = h(x_{n+1})e^{\theta(n\kappa_s - S(n))}E(e^{-\theta x_{n+1}})e^{\theta\kappa_s} \\ &\quad = h(x_n)e^{\theta(n\kappa_s - S(n))}e^{-\theta\kappa_s}e^{\theta\kappa_s} \\ &\quad = h(x_n)e^{\theta(n\kappa_s - S(n))}. \end{aligned}$$

Felix Poloczek [16] proposed several methods of channel access martingale. In this paper, slotted ALOHA [22], [23], competitive decomposition diversity slotted ALOHA [24], [26], and enhanced competitive decomposition diversity slotted ALOHA [25], [27] are represented as finite-state Markov chains. In addition, the martingale process access service is given by taking a two-state Markov chain as a model.

[3] In the references, the early S-ALOHA and the recent satellite network access protocol are provided. The latest protocols mainly consider how to improve the access success rate, so our analysis structure can be directly used without a need for many changes. Therefore, we adopt the S-ALOHA protocol.

Proposition 2: The random access process is a martingale process [16]. Let (x_n) be a Markov chain with two states, and s be the state space; it is assumed that $f : S \rightarrow R^+$ is a deterministic function, so $s(n) = \sum_{k=1}^{K} f(x_k)$ is a service martingale.

The state transition matrix can be defined as follows:

$$\mathbf{n}_{i,j} = P(x_{n+1} = j | x_n = i). \tag{4}$$

Accordingly, the exponential transformation matrix of Eq.(4) is expressed as follows:

$$\mathbf{T}^{\theta}_{i,j} = \mathbf{n}_{i,j}e^{\theta f(x_k)}. \tag{5}$$

For any $k > 0$, it holds that:

$$\begin{aligned} &E(h(x_{n+1})e^{-\theta((n+1)K - S(n+1))}|x_1 \ldots x_n) \\ &\quad = h(x_{n+1})e^{-\theta(nK - S(n))}E(h(x_{n+1})e^{-\theta x_{n+1}}|x_n)e^{-\theta K} \\ &\quad = e^{\theta(nK - S(n))}(T^{\theta}h)(x_n)e^{-\theta K} \\ &\quad = e^{\theta(nK - S(n))}sp(T^{\theta})e^{-\theta K}, \end{aligned}$$

where $(T^{\theta h})(x_n)$ is $T^{\theta h}$ of the vector x_n components, and $sp(T^{\theta})$ is the spectral radius of $\mathbf{T}^{\theta}_{i,j}$.

When the condition of $sp(T^{\theta})e^{-\theta K} < 1$ is true,

$$E\left(h(x_{n+1})e^{-\theta((n+1)K - S(n+1))}|x_1 \ldots x_n\right) \leq e^{\theta(nK - S(n))}. \tag{6}$$

Cache schedule based on SP In the SP schedule algorithm, the arrival flow can be of high priority $A^h(n)$ and low priority $A^f(n)$. So, the arrival process can be represented as $A^t(n) = A^h(n) + A^f(n)$.

Martingale Envelope: For $p \geq 0$, the monotonically increasing function $h_1, h_2, \cdots, h_p$, when $\theta > 0$, $(\Pi h, \theta, c)$ is called a martingale encapsulation process of data stream A, which can be expressed as Eq.(7).

$$\Pi\left(h(\overrightarrow{a_n})e^{\theta(A^r(m,n)\text{-}(n-m)c)}\right) \leq M_m(n). \tag{7}$$

Lemma 1: When the arrival process $A^t(n)$ is a supermartingale, the martingale encapsulation [14] can be expressed as:

$$\begin{aligned} M_{A^t}(n) = &\ h_{A^h}(a_n)h_{A^f}(a^f_n)h_{s^q}(s_n) \\ &\times e^{\theta(A^h(\kappa,n)\text{-}(n-\kappa)\kappa_f + A^f(n) - nk_h + n\kappa_s - S^q(n))}. \end{aligned} \tag{8}$$

Proof: Please refer to Appendix A.

In (8), $h_{A^h}(a_n)$, $h_{A^f}(a^f_n)$, $h_{s^q}(s_n)$ represent monotonically increasing functions, and κ_h and κ_f are high and low priority arrival rates on a satellite, respectively; $S^q(n)$ represents a queuing service process; $\theta > 0$, and κ_s depend on the queuing service rate.

Lemma 2: When the service process $S(n)$ is a supermartingale, the martingale encapsulation [14] can be expressed as:

$$M_S(n) = h_S(s_n)e^{\theta^*(\{(S(n) - n\kappa_{ss}\})}, \tag{9}$$

where $h_S(s_n)$ is a monotonic increasing function, $\theta^* > 0$ and κ_{ss} depend on R_{su}.

IV. SATELLITE CACHE OCCUPANCY ESTIMATION

In fact, the cache occupancy is closely related to the data arrival process, channel conditions, and cache service algorithm; the scheduling algorithm can be First In First Out (FIFO), Static Priority (SP), Least Recently Used (LRU), or Least Frequently Used (LFU). Therefore, this paper abstracts these services as a cascade of multiple services and estimates the cache occupancy using the martingale theory.

Theorem 1: Low-orbit satellite node cache backlog. Assume that the arrival process $A(n)$ and the service process $S(n)$ are two independent martingale processes, then Eq.(11) holds:

$$\theta^* := \sup\{\theta > 0 : k_a \leq k_s\}. \tag{10}$$

This queuing system is stable. For any $\sigma \geq 0$, the cache backlog bound can be expressed as:

$$p(B(n) \geq \sigma) \leq \frac{E(h_a(a_0))E(h_s(s_0))}{H} e^{-\theta^* \sigma}, \tag{11}$$

where,

$$H := \min\{h_a(x)h_s(y) : x - y > 0\}. \tag{12}$$

Proof: Please refer to Appendix B.

In order to support the principle of the martingale encapsulation estimation method, the derivation process of the upper bound of a virtual delay W(n) is given first, and then the upper bound of the virtual delay W(n) is defined.

Theorem 2: The upper bound of a virtual delay of business flow W(n) is expressed as:

$$p(\mathrm{W}(n) > \kappa) \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H} e^{-\theta_1^* \kappa \kappa_{ss}}, \tag{13}$$

where $R_a = \min\{\kappa_s, R_{us}\}$, $\kappa_{ss} = R_{su}$, $R_a < \kappa_{ss}$ and θ_1^*, any value satisfies:

$$\theta_1^* \in \left\{ \theta > 0 \middle| \frac{R_a \ln[sp(\mathbf{M}^\theta)]}{\theta} \leq \frac{\ln[e^{-\theta \kappa_{ss}}]}{-\theta} \right\}. \tag{14}$$

Proof: Please refer to Appendix C.

A. CACHE OCCUPANCY ESTIMATION BASED ON FIFO SCHEDULING ALGORITHM

The cache occupancy is not related only to the arrival process but also to the service rules in the cache, i.e., the scheduling algorithm. The current service rules include FIFO, SP, Last-In-First-Out (LIFO), and others. This paper estimates the satellite cache occupancy based on the FIFO and SP caching service rules.

Theorem 3: The upper backlog bound of a LEO satellite based on the FIFO scheduling algorithm can be expressed as:

$$p(B(n) > \sigma_{FIFO,L}) \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H} e^{-\theta_1^* \sigma_{FIFO,L}}, \tag{15}$$

where $R_a = \min\{\kappa_s, R_{us}\}$, $\kappa_{ss} = R_{su}$, $R_a < \kappa_{ss}$ and any value of θ_1^* satisfies the following:

$$\theta_1^* \in \left\{ \theta > 0 \middle| \frac{R_a \ln[sp(\mathbf{M}^\theta)]}{\theta} \leq \frac{\ln[e^{-\theta \kappa_{ss}}]}{-\theta} \right\}. \tag{16}$$

Proof: Please refer to Appendix D.

B. CACHE OCCUPANCY ESTIMATION BASED ON SP SCHEDULING ALGORITHM

Static priority is suitable for mixed services. For instance, voice service generally has high priority, and data service generally has low priority. Static priority was introduced in the field of Internet business in the early days, and the current land mobile networks have only refined this rule. Moreover, static priority is the basis of other priorities. Therefore, in this part, according to the martingale encapsulation theory, the estimation of cache occupancy in the case of static priority is obtained as follows.

Theorem 4: Given an amount of data consumed by a cache, $\sigma^t_{SP,L} \geq 0$. The upper bound of $p^h_{SP,L}$ is defined as:

$$\frac{E(M_{A^f}(0))E(M_{A^h}(0))E(M_S(0))}{H} e^{-\theta_2^*(\sigma^t_{SP,L} - \sigma^{a,h}_{SP,L})}, \tag{17}$$

where, θ^*.

The following conditions are satisfied:

$$\theta_2^* \in \left\{ \theta > 0 \middle| R_a \leq \frac{\kappa_{ss} \ln[sp(\mathbf{M}^\theta)]}{\theta} \right\}. \tag{18}$$

Proof: Please refer to Appendix E.

C. CACHE OCCUPANCY ESTIMATION BASED ON LRU SCHEDULING ALGORITHM

LRU principle: Given an independent cache of the size K, LRU strategy retains K recently needed objects. The location of the first cache is denoted as the most recently used (MRU), and the last cache is denoted as the least recently used (LRU). When a new demand arrives, there are two options, update and insert. If the required object is already in the cache, the policy will update to move the object to the MRU position. When the required object is not in the cache, it is inserted into the MRU position as a new object, and the object at the LRU position is evicted. In the LRU strategy, there are two processes, namely, update and insertion. When the update process conducts, the cache occupancy does not change, but when the insert process conducts, the cache occupancy changes due to the insertion. Therefore, it is necessary to estimate the cache occupancy in the case of object insertion. According to the martingale encapsulation theory, the estimation of cache occupancy in the case of LRU is as follows.

Theorem 5: Cache occupancy in the case of LRU is given by:

$$p(B(n) > \sigma_{LRU,L}) \leq \frac{E(M_{A^t}(0)) \prod_{i=1}^{L} h_{S_i}(s_n)}{H} e^{-\theta_G^* \sigma_{LRU,L}}. \tag{19}$$

Let $h_{S_i}(s_n) = 1$, then Eq.(19) can be simplified to:

$$p(B(n) > \sigma_{LRU,L}) \leq E(M_{A^t}(0)) e^{-\theta_G^* \sigma_{LRU,L}},$$

where

$$\sum_{i=1}^{L} \kappa_{S_i} = \kappa_{sa}(\tau_{sa}) + \kappa_{suc}(\tau_{suc}) + \kappa_{sLRU}(\tau_{sLRU}) + \kappa_{sdc}(\tau_{sdc}) = (sp(T^{\theta_{su}}) + sp(T^{\theta_{LRU}}))R_{us} + R_{us} + R_{ds} \tag{20}$$

$$\sum_{i=1}^{L} S_i(\tau_i) = S_{sa}(\tau_{sa}) + S_{suc}(\tau_{suc}) + S_{sLRU}(\tau_{sLRU}) + S_{sdc}(\tau_{sdc}) \tag{21}$$

$$\theta^*_{LRU,L} \in \left\{ \theta > 0 \middle| \frac{R_a \ln[sp(\mathbf{M}^\theta)]}{\theta} \le \frac{\ln[e^{-\theta \sum_{i=1}^{L}(\tau_i \kappa_{S_i} - S_i(\tau_i))}]}{-\theta} \right\}. \tag{22}$$

Proof: Please refer to Appendix F.

D. CACHE OCCUPANCY ESTIMATION BASED ON LFU SCHEDULING ALGORITHM

The LFU stands for the least frequently used. In the LFU strategy, data that are used the least frequently in a certain period of time are eliminated first. The LFU represents a caching algorithm used to manage computer memory. The main purpose of this algorithm is to record and track the number of times the memory block is used. When the cache is full and more space is needed, the system will clear the memory blocks with the lowest memory usage frequency. The simplest way to use the LFU algorithm is to allocate each block loaded into the cache counter A. Each time the block is referenced, the counter will be increased by one. When the cache reaches its capacity and a new memory block is waiting to be inserted, the system will search for the block with the lowest counter and delete it from the cache. Therefore, according to martingale encapsulation theory, the estimation of cache occupancy in the LFU case is as follows:

$$p(B(n) > \sigma_{LFU,L}) \le E(M_{A^t}(0)) e^{-\theta^*_G \sigma_{LFU,L}}.$$

where

$$\sum_{i=1}^{L} \kappa_{S_i} = \kappa_{sa}(\tau_{sa}) + \kappa_{suc}(\tau_{suc}) + \kappa_{sLFU}(\tau_{sLFU}) + \kappa_{sdc}(\tau_{sdc}) = (sp(T^{\theta_{su}}) + sp(T^{\theta_{LFU}}))R_{us} + R_{us} + R_{ds} \tag{23}$$

$$\sum_{i=1}^{L} S_i(\tau_i) = S_{sa}(\tau_{sa}) + S_{suc}(\tau_{suc}) + S_{sLFU}(\tau_{sLFU}) + S_{sdc}(\tau_{sdc}) \tag{24}$$

$$\theta^*_{LFU,L} \in \left\{ \theta > 0 \middle| \frac{R_a \ln[sp(\mathbf{M}^\theta)]}{\theta} \le \frac{\ln[e^{-\theta \sum_{i=1}^{L}(\tau_i \kappa_{S_i} - S_i(\tau_i))}]}{-\theta} \right\}. \tag{25}$$

E. ESTIMATION OF HIGH EARTH ORBIT SATELLITE CACHE OCCUPANCY

Different from the estimation of LEO satellite cache occupancy, a high-orbit satellite has not only its own business arrival but also the business arrival of LEO satellite users. In the estimation of high-orbit satellite cache occupancy, the arrival of LEO satellite services should be considered first, which should arrive after two hops of LEO satellite users. Therefore, this part analyzes the high-orbit satellite node cache occupancy for the case of two hops. Similar to the signal and system theory in circuit analysis, the service curve can be regarded as an impulse response of a linear system. In fact, in a multi-system structure, where systems are connected in series, the first-system output represents the second-system input. By analogy, the concatenation of multiple systems can be replaced by a single equivalent system $s(\tau, t)$, which consists of the minimum plus convolution of a single service process. Assume there are L servers in series, as given in Eq. (26).

$$S(\tau, n) = S_1(\tau, \tau_1) \otimes S_2(\tau_1, \tau_2) \otimes, \cdots, \otimes S_L(\tau_{L-1}, n) = \inf\{S_1(\tau_1) + S_2(\tau_2) + \cdots, S_L(\tau_L)\}. \tag{26}$$

Theorem 6: The upper bound of the high-orbit satellite cache backlog can be expressed as:

$$p(\mathrm{B}(n) > \sigma) \le \frac{E(M_{A^t}(0)) \prod_{i=1}^{L} E(M_{S_i}(0))}{H} e^{-\theta^*_G \sigma}. \tag{27}$$

where, $R_a = \min\{\sum_{i=1}^{L} \kappa_{S_i}, R_{us}\}$, $\sum_{i=1}^{L} \kappa_{S_i} = R_{su}$, $R_a < \sum_{i=1}^{L} \kappa_{S_i}$ and θ^*_G satisfied:

$$\theta^*_G \in \left\{ \theta > 0 \middle| \frac{R_a \ln[sp(\mathbf{M}^\theta)]}{\theta} \le \frac{\ln[e^{-\theta \sum_{i=1}^{L}(\tau_i \kappa_{S_i} - S_i(\tau_i))}]}{-\theta} \right\}. \tag{28}$$

$$\begin{aligned} p(B(n) > \sigma_{FIFO,G}) &= p(A^t(0, n-k) \ge D(n)) \\ &\le p(A^t(n,\kappa) \ge \min_{0 \le m \le n} \{A^t(m) + \sum_{i=1}^{L} S_i(\tau_i)\}) \\ &\le p(\max_{n \ge \kappa}\{A^t(n,\kappa) + (n-\kappa)(\sum_{i=1}^{L} \kappa_{S_i} - R_a) - \sum_{i=1}^{L} S_i(\tau_i)\} \ge 0 \\ &\le p(\max_{n \ge \kappa}\{A^t(n,\kappa) + (n-\kappa)R_a + n\sum_{i=1}^{L} \kappa_{S_i} - \sum_{i=1}^{L} S_i(\tau_i)\} \ge \kappa \sum_{i=1}^{L} \kappa_{S_i} \\ &\le \frac{E(M_{A^t}(0)) \prod_{i=1}^{L} h_{S_i}(s_n)}{H} e^{-\theta^*_G \sigma_{FIFO,G}}. \end{aligned} \tag{29}$$

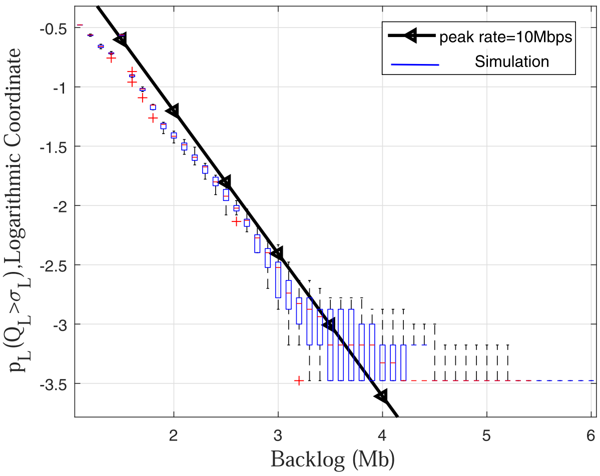

FIGURE 3. The probability of violation varies with the amount of cache occupied by FIFO scheduling.

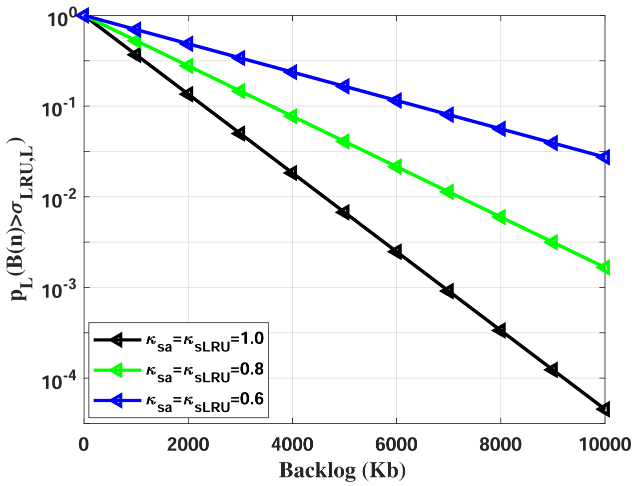

FIGURE 4. The violation probability changes with the cache usage based on LRU scheduling.

The upper bound of cache backlog for high earth orbit satellites can be derived as (29), shown at the bottom of the previous page.

Then, based $p_{FIFO,G} = p(B(n) > \sigma_{FIFO,G})$, the occupancy of the GEO satellite cache can be calculated by:

$$\sigma_{FIFO,G} = \frac{1}{\theta_G^*} \ln \frac{E(M_{A'}(0)) \prod_{i=1}^{L} E(M_{S_i}(0))}{Hp_{FIFO,G}}. \quad (30)$$

Proof: Please refer to Appendix G.

V. NUMERICAL RESULTS AND THE ANALYSIS

Several numerical simulations were conducted to validate the performance of the backlog estimation based on content popularity. We use STK tools to generate our simulation environment. The STK (Systems Tool Kit) tool can be used to generate an Iridium-like constellation which contained 66 LEO satellites evenly distributed over six orbital planes. According to the global geometries of GEO satellites, we assume that one GEO satellite covers 22 LEO satellites.

TABLE 2. Simulation parameters.

Notation	parameters	Description
p_k	10Mbps	peak rate
R_{on}	0.1	On Off rate
R_A	1Mbps	average arrival rate
κ_s	>1	$\frac{queuing\ \ service\ \ rate}{average\ \ arrival\ \ rate}$
R_{us}	<1	uplink data rate
κ_{ss}	=1	downlik service rate
R_{su}	<1	downlink data rate
κ_{sa}	1, 0.8, 0.6	access service rate

It was assumed that $\kappa_s > R_{us}$, and $\kappa_{ss} = R_{su}$. For the convenience of comparison, we assume that the user burst rate of the random access channel is used as a reference, and the other rates κ_s, R_{us}, R_{ss} and R_{su} are all normalized values, The notations used in the following are listed in Table 2. In our simulation, we assume that the random burst rate of the user data access channel is $10Mbps$. In order to describe the randomness of the data, we use the on-off modulation method for simulation, where the on-off ratio R_{on} is set to 0.1. At this time, the average data arrival rate is $10 * 0.1 = 0.1Mbps$. Combining user access ($\kappa_{sa} = 1$, $\kappa_{sa} = 0.8$ and $\kappa_{sa} = 0.6$), queuing rate($\kappa_{ss} = 1$) and downlink rate $R_{su} < 1$, we can use queuing theory to obtain the average value of cache occupancy that increases linearly with time. In addition, because of the different queuing strategies of the cache area, the downlink rate R_{su} is different. The estimation of cache occupancy under different queuing rules(LRU,LFU) is also simulated.

The relationship between cache occupancy (backlog) and violation probability of LEO satellite nodes was analyzed through the simulation with the FIFO cache scheduling. In order to verify the martingale encapsulation theory to estimate the cache occupancy, we use random burst data to pass through a random channel, and the cache backlog after the FIFO queuing strategy is used as a box plot and the encapsulation function estimate is simulated and compared. As shown in Fig. 3, using the martingale encapsulation function to estimate the cache occupancy is accurate.

In Fig. 4, the ordinate represents the probability of violation on a logarithmic scale. As shown in Fig. 4, as the backlog increased, the probability of violation decreased. In addition, given a violation probability that satisfied the need, an estimate of backlog could be obtained. According to Eq. (46), the probability of backlog violation was higher when the access probabilities were larger. In other words, under larger access and insert probabilities, the cache could accumulate content more easily. In addition, the LRU scheduling algorithm was used to simulate the violation probability of backlog of the LEO satellite cache.

In addition, three access service rates were compared using the LFU schedule algorithm, and the access service probability values were $\kappa_{sa} = 1$, $\kappa_{sa} = 0.8$ and $\kappa_{sa} = 0.6$. For $\kappa_{sa} = 1$, it corresponds to the situation of fixed access,

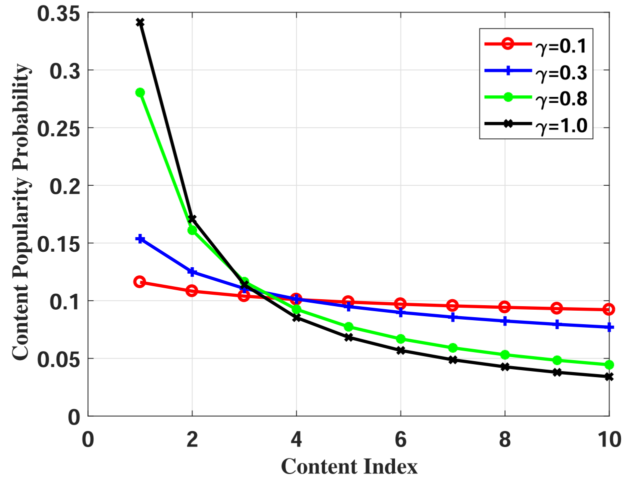

FIGURE 5. Zipf distribution (10 contents).

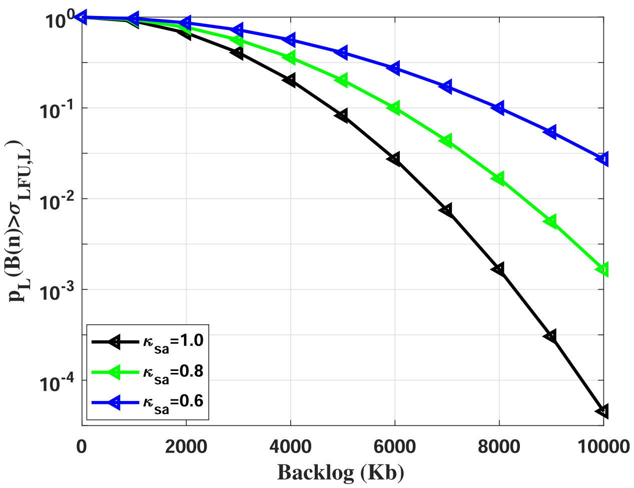

FIGURE 6. The probability of violation varies with the amount of cache occupied by LFU scheduling.

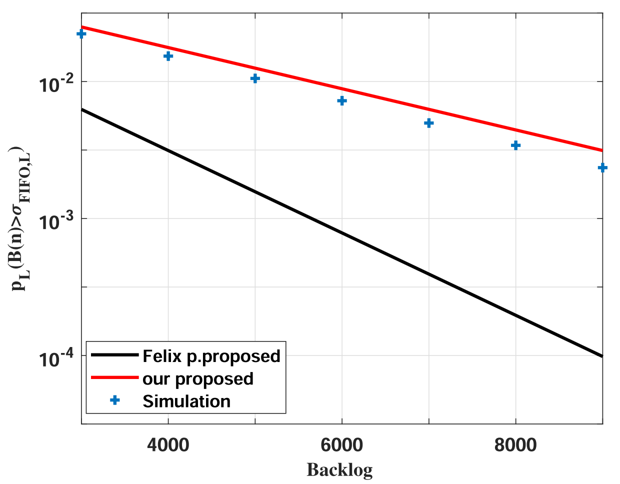

FIGURE 7. Comparison of two methods based on FIFO scheduling (Flex P. and our proposed method).

that is, the situation where the user access to the channel will not fail. For $\kappa_{sa} = 0.6$, it corresponds to the situation of the theoretical passing value of the current improved time S-ALOHA. For $\kappa_{sa} = 0.8$, it corresponds to the situation that the theoretical passing value is improved after random access is S-ALOHA.

In fact, with the increase in the backlog, the high access service probability decreased, while the probability of violation increased. As shown in Fig. 4, when the backlog was unchanged, the lower the access probability was, the greater the probability of backlog violation was, that is, the decrease in the access probability caused the cache backlog to increase. According to the LFU scheduling algorithm, the probability of content being inserted into a cache depends on the content's popularity. In this work, the Zipf distribution was used to describe the content's popularity, as shown in Fig. 6. As shown in Fig. 5, compared with the FIFO scheduling algorithm, the violation probability curve of the LFU was relatively flat, which was the reason why the popularity distribution of the content was considered.

In [16], Felix p. *et al.* proposed a service martingale under the random access protocol. Fig. 7 compares the cache occupancy estimates of our proposed method and the method proposed by Felix p. *et al.* in the case of FIFO queuing and S-AHLOA. It can be seen from Fig. 7 that in the two estimation methods, the simulation curve basically coincides with our theoretical curve, but has a large deviation from the estimation result by Felix p. *et al.*. The reason why our estimation is more accurate is that the cache occupancy estimation method we proposed fully considers the user's access protocol and queuing strategy, and we also consider the satellite channel conditions.

VI. CONCLUSION

In this article, we jointly considered satellite storage network access, channel conditions and queuing strategies to design a martingale encapsulation function, obtained the upper bound of the satellite node's cache occupancy, and gave an estimate of the cache occupancy. Compared with the queuing theory, the proposed method does not assume that the data arrival obeys the Poisson distribution, which is in accordance with the actual cases because the general data arrival does not necessarily follow the Poisson distribution. Then, Queuing theory can only give the average value of queue length, while the analysis of cache occupancy based on martingale theory can give the probability distribution of cache occupancy. Finally, the accuracy of the proposed martingale encapsulation method is verified by simulations. The results show that as long as the channel access protocol and transmission rate can be abstracted as martingale processes for general content placement network, the proposed method can be used to estimate the cache occupancy.

APPENDIX A
PROOF OF LEMMA 1

The arrival process $A^t(T)$ is a supermartingale, where $g(Y_{T\tau})$ can be written as

$$g(Y_T(\tau)) = \begin{cases} I(T)R_{us}, & good \\ 0, & bad \end{cases} \tag{31}$$

Satellite Channel Status Description: Satellite channels are greatly affected by the environment and weather. Therefore, two states are generally used to indicate this channel state, namely, 'good' means a good channel, and 'bad' means a bad channel.

Proof: Before the proceeding, the transition matrix $\mathbf{M}$ for x_N is defined by

$$\mathbf{M}_{i,j} = P(Y_{T+1} = j | Y_T = i). \quad (32)$$

Correspondingly, the exponential transformation along the columns of the matrix $\mathbf{M}_{i,j}$ [16] can be expressed as

$$\mathbf{M}^{\theta}_{i,j} = \mathbf{M}_{i,j} e^{\theta g(Y_j)}. \quad (33)$$

The channel service process can be expressed as

$$\begin{aligned} E&\left[e^{\theta(A^t(T+1)-g(Y(T+1)))}\middle| Y_1, Y_2, \cdots, Y_T\right] \\ &\leq e^{\theta(A^t(T)-R_{us}T)} E\left[e^{\theta Y_{T+1}}\middle| Y_T\right] \\ &\leq e^{\theta(A^t(T)-R_{us}T)} e^{-\theta p_m (1-p_m)^N R_{us}} sp\left(\mathbf{M}^{\theta}\right) \\ &\leq e^{\theta(A^t(T)-R_{us}T)}, \end{aligned} \quad (34)$$

where $sp(\mathbf{M}^{\theta})$ is $\mathbf{M}^{\theta}_{i,j}$ which is the spectral radius, and the inequality is workable in $e^{-\theta R_{us} p_m (1-p_m)^N} sp\left(\mathbf{M}^{\theta}\right) \geq 1$ case.

In this work, we assume the high-priority arrival $A^h(T)$, and low-priority arrives at $A^f(T)$. Then, the martingale envelope can be defined by

$$\begin{aligned} M_{A^t}(T) = {} & h_{A^h}(a_T) h_{A^f}(d^f_T) h_{s^q}(s_T) \\ & \times e^{\theta(A^h(\kappa,T)\text{-}(T-\kappa)\kappa_f + A^f(T) - Tk_h + T\kappa_s - S^q(T))}, \end{aligned}$$

where $h_{A^h}(a_T)$, $h_{A^f}(d^f_T)$, $h_{s^q}(s_T)$ are monotonically increasing functions, and κ_h and κ_f are the high-priority arrival and low-priority arrival rates of the LEO/GEO satellite, respectively, $S^q(T)$ is the queue service process, $\theta^* > 0$ and κ_s depends on the queue service rate. □

APPENDIX B
PROOF THEOREM 1

According to the definition of the martingale envelope given by Eq.(7), we consider the arrival supermartingale envelope $M_{A^t}(T)$ and service supermartingale envelope $M_s(T)$ are respectively expressed by

Proof: Suppose θ^* is given, and the parameters h_a, h_s, k_s and k_a are also fixed, so, according to the Independent and identically distributed hypothesis, similar to Proposition 2.2, can prove that the following process is a supermartingale

$$h_a(a_n) h_s(s_n) e^{-\theta^*(A(n) - nk_a + nk_s - S(n))}. \quad (35)$$

Define (36) as a martingale package

$$M(n) := h_a(a_n) h_s(s_n) e^{-\theta^*(A(n) - S(n))}. \quad (36)$$

Define N as when $A(n) - S(n)$ exceeds a stop of σ for the first time, the mathematical expression is as Eq.(37).

$$N := \min\{n : A(n) - S(n) \geq \sigma\}. \quad (37)$$

Note that $N = \infty$ is possible, then $p(B(t) \geq \sigma) = P(N < \infty)$. The stopping time theorem shows that a martingale process satisfies the condition $p(N < \infty) = 1$, $E(|M(n)|) < \infty$, $\lim_{n\longrightarrow\infty} E(|M(n) I_{\{n>N\}}|) = 0$, the expected value of any stop is equal to its expected value at time zero. Obviously, the martingale package $M(n)$ satisfies these three conditions. Applying the stopping time theorem to stopping time $N \wedge n := \min\{N, n\}$, there is the following derivation:

$$\begin{aligned} E&(h_a(a_0))E(h_s(s_0)) \\ &= E(M(0)) \\ &= E(M(N \wedge n)) \\ &\geq E(M(N \wedge n)|1_{\{N \leq n\}}) \\ &\geq E(h_a(a_N))E(h_s(s_N))E(e^{\theta^*(A(N)-S(N))})p(N \leq n) \\ &\geq He^{\theta^*\sigma} p(N \leq n) \\ &\geq He^{\theta^*\sigma} p(B(t) \geq \sigma). \end{aligned} \quad (38)$$

APPENDIX C
PROOF THEOREM 2

According to the definition of the martingale envelope given by Eq.(7), we consider the arrival supermartingale envelope $M_{A^t}(T)$ and service supermartingale envelope $M_s(T)$ are respectively expressed by

$$p(\mathrm{W}(n) > \kappa) \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H} e^{-\theta^*_1 \kappa \kappa_{ss}}, \quad (39)$$

where $R_a = \min\{\kappa_s, R_{us}\}$, $\kappa_{ss} = R_{su}$, $R_a < \kappa_{ss}$ and θ^*_1, any value satisfies:

$$\theta^*_1 \in \left\{\theta > 0 \middle| \frac{R_a \ln[sp(\mathbf{M}^{\theta})]}{\theta} \leq \frac{\ln[e^{-\theta\kappa_{ss}}]}{-\theta}\right\}. \quad (40)$$

Proof: According to Eq.(8), consider an arrival martingale package $M_{A^t}(n)$ and a service martingale package $M_s(n)$. These two martingale processes can be Expressed as follows:

$$\begin{aligned} M_{A^t}(n) &= h_{A^t}(a^f_n) e^{\theta^*(\{(A^t(n,\kappa)\text{-}(n-\kappa)R_a\})}, \\ M_S(n) &= h_S(s_n) e^{\theta^*(\{(S(n)-n\kappa_{ss}\})}. \end{aligned}$$

Here, this article assumes that the arrival process and the service process are independent. Therefore, the process $M(n)$ in the time domain $V := \kappa, \kappa+1, \ldots$ is also a martingale process. This random process can be expressed as follows:

$$\begin{aligned} M(n) = h_{A^t}(a^t_n) h_S(s_n) & e^{\theta^*(\{(A^f(n,\kappa)\text{-}(n-\kappa)R_a\})} \\ & \times e^{\theta^*(\{(S(n)-T\kappa_{ss}\})}. \end{aligned} \quad (41)$$

Therefore, according to the stopping time theorem, there is the following derivation

$$\begin{aligned} E(M(0)) &= E(M(\kappa \wedge n)) \\ &\geq E(M(\kappa \wedge n)|\{\kappa \leq n\}) \\ &\geq E(h_{A^t}(a^t_n) h_S(s_n)) \\ &\quad \times E(e^{\theta^*((A^t(n,\kappa)\text{-}(n-\kappa)R_a - S(n) + n\kappa_{ss}))}) p(\kappa \leq n) \\ &\geq He^{\theta^*\kappa\kappa_{ss}} p(\kappa \leq n), \end{aligned} \quad (42)$$

where, $H := \min\{h_{A^t}(a_n)h_S(s_n) : a_n - s_n > 0\}$

$$\begin{aligned} &p(\mathrm{W}(n) > \kappa) \\ &\quad = p(A^t(0, n-k) \geq D(n)) \\ &\quad \leq p(A^t(n,\kappa) \geq \min_{0\leq m\leq n}\{A^t(m) + S(m,n)\}) \\ &\quad \leq p(A^t(n,\kappa) - \min_{0\leq m\leq n}\{A^t(m) + S(m,n)\} \geq 0) \\ &\quad \leq p(\max_{n\geq\kappa}\{A^t(n,\kappa)\text{-}(n-\kappa)\kappa_{ss} + nR_a - S(n)\} \geq \kappa\kappa_{ss}) \\ &\quad \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H}e^{-\theta^*\kappa\kappa_{ss}}. \end{aligned} \tag{43}$$

APPENDIX D
PROOF THEOREM 3

Proof: According to Eq.(8), consider an arriving supermartingale package $M_{a^t}(n)$ and a service supermartingale package $m_s(n)$, these two martingale processes can be expressed as follows:

$$M_{A^t}(n) = h_{A^t}(a_n^f)e^{\theta^*(\{(A^t(n,\kappa)\text{-}(n-\kappa)R_a\})},$$
$$M_S(n) = h_S(s_n)e^{\theta^*(\{(S(n)-n\kappa_{ss}\})}.$$

In this paper, we assume that the arrival process and the service process are independent. Thus, the process $m(n)$ is also a supermartingale process in time domain $V := \kappa, \kappa+1, \ldots$. This stochastic process can be expressed as follows:

$$M(n) = h_{A^t}(a_n^t)h_S(s_n)e^{\theta^*(\{(A^f(n,\kappa)\text{-}(n-\kappa)R_a\})} \times e^{\theta^*(\{(S(n)-n\kappa_{ss}\})}. \tag{44}$$

Therefore, according to the stopping time theorem, there is the following derivation

$$\begin{aligned} &E(M(0)) \\ &\quad = E(M(\kappa\wedge n)) \\ &\quad \geq E(M(\kappa\wedge n)|\{\kappa\leq n\}) \\ &\quad \geq E(h_{A^t}(a_n^t)h_S(s_n)) \\ &\qquad \times E(e^{\theta^*((A^t(n,\kappa)\text{-}(n-\kappa)R_a - S(n) + n\kappa_{ss}))})p(\kappa\leq n) \\ &\quad \geq He^{\theta^*\kappa\kappa_{ss}}p(\kappa\leq n), \end{aligned} \tag{45}$$

where, $H := \min\{h_{A^t}(a_n)h_S(s_n) : a_n - s_n > 0\}$

$$\begin{aligned} &p(B(n) > \sigma_{FIFO,L}) \\ &\quad = p(A^t(0, n-k) \geq D(n)) \\ &\quad \leq p(A^t(n,\kappa) \geq \min_{0\leq m\leq n}\{A^t(m) + S(m,n)\}) \\ &\quad \leq p(A^t(n,\kappa) - \min_{0\leq m\leq n}\{A^t(m) + S(m,n)\} \geq 0) \\ &\quad \leq p(\max_{n\geq\kappa}\{A^t(n,\kappa)\text{-}(n-\kappa)\kappa_{ss} + nR_a - S(n)\} \geq \kappa\kappa_{ss}) \\ &\quad \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H}e^{-\theta^*\kappa\kappa_{ss}} \\ &\quad \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H}e^{-\theta^*\sigma_{FIFO,L}}. \end{aligned}$$

Then, based on $p_{FIFO,L}(p_{FIFO,L} = p(B(n) > \sigma_{FIFO,L}))$, The usage of this cache can be estimated, as shown in Eq.(46).

$$\sigma_{FIFO,L} = \frac{1}{\theta_1^*}\ln\frac{E(M_{A^t}(0))E(M_S(0))}{Hp_{FIFO,L}}. \tag{46}$$

APPENDIX E
PROOF THEOREM 4

Proof: In order to obtain the cache backlog of low priority services, which is the same as FIFO, this paper first deduces the whole cache backlog. According to the stopping time theorem, there is the following derivation

$$\begin{aligned} &E(M(0)) \\ &\quad = E(M(\kappa\wedge n)) \\ &\quad \geq E(M(\kappa\wedge n)|\{\kappa\leq n\}) \\ &\quad \geq E(h_{A^t}(a_n^t)h_S(s_n)) \\ &\qquad \times E(e^{\theta^*((A^t(n,\kappa)\text{-}(n-\kappa)R_a - S(n) + n\kappa_{ss}))})p(\kappa\leq n) \\ &\quad \geq He^{\theta^*\kappa\kappa_{ss}}p(\kappa\leq n), \end{aligned} \tag{47}$$

where, $H := \min\{h_{A^t}(a_n)h_S(s_n) : a_n - s_n > 0\}$

$$\begin{aligned} &p(B(n) > \sigma_{SP,L}) \\ &\quad = p(A^t(0, n-k) \geq D(n)) \\ &\quad \leq p(A^t(n,\kappa) \geq \min_{0\leq m\leq n}\{A^t(m) + S(m,n)\}) \\ &\quad \leq p(A^t(n,\kappa) - \min_{0\leq m\leq n}\{A^t(m) + S(m,n)\} \geq 0) \\ &\quad \leq p(\max_{n\geq\kappa}\{A^t(n,\kappa)\text{-}(n-\kappa)\kappa_{ss} + nR_a - S(n)\} \geq \kappa\kappa_{ss}) \\ &\quad \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H}e^{-\theta^*\kappa\kappa_{ss}} \\ &\quad \leq \frac{E(M_{A^t}(0))E(M_S(0))}{H}e^{-\theta^*\sigma_{SP,L}}. \end{aligned}$$

Then, based on $p_{SP,L}(p_{SP,L} = p(B(n) > \sigma_{SP,L}))$, we can estimate the usage of this cache as follows (48)

$$\sigma_{SP,L} = \frac{1}{\theta_1^*}\ln\frac{E(M_{A^t}(0))E(M_S(0))}{Hp_{SP,L}}. \tag{48}$$

Similarly, the backlog boundary of low priority services is shown as follows:

$$\begin{aligned} &p(B(n) > \sigma_{SP,L}^t) \\ &\quad \leq \frac{E(M_{A^t}(0))E(M_{A^h}(0))E(M_S(0))}{H}e^{-\theta_2^*(\sigma_{SP,L}^t - \sigma_{SP,L}^{a,h})}. \end{aligned} \tag{49}$$

Then, the utilization of low priority service cache can be expressed as follows:

$$\sigma_{SP,L}^t = \frac{1}{\theta_2^*}\ln\frac{E(M_{A^f}(0))E(M_{A^h}(0))E(M_S(0))}{Hp_{SP,L}} + \sigma_{SP,L}^{a,h}. \tag{50}$$

APPENDIX F
PROOF THEOREM 5

Proof: According to Eq.(8), consider an arriving supermartingale package $M_{a^t}(n)$ and a service supermartingale package $M_{s_i}(n)$ These two martingale processes can be expressed as follows:

$$\begin{aligned} &p(B(n) > \sigma_{LRU,L}) = p(A^t(0, n-k) \geq D(n)) \\ &\quad \leq p(A^t(n,\kappa) \geq \min_{0\leq m\leq n}\{A^t(m) + \sum_{i=1}^{L} S_i(\tau_i)\}) \end{aligned}$$

$$\leq p(\max_{n\geq\kappa}\{A^t(n,\kappa)+(n-\kappa)(\sum_{i=1}^{L}\kappa_{S_i}-R_a)-\sum_{i=1}^{L}S_i(\tau_i)\}$$
$$\geq 0$$
$$\leq \frac{E(M_{A^t}(0))\prod_{i=1}^{L}h_{S_i}(s_n)}{H}e^{-\theta_G^*\sigma_{LRU,L}}.$$

where,

$$\sum_{i=1}^{L}\kappa_{S_i} = \kappa_{sa}(\tau_{sa})+\kappa_{suc}(\tau_{suc})+\kappa_{sLRU}(\tau_{sLRU})+\kappa_{sdc}(\tau_{sdc})$$
$$= (sp(T^{\theta_{su}})+sp(T^{\theta_{LRU}}))R_{us}+R_{us}+R_{ds} \quad (51)$$
$$\sum_{i=1}^{L}S_i(\tau_i) = S_{sa}(\tau_{sa})+S_{suc}(\tau_{suc})+S_{sLRU}(\tau_{sLRU})+S_{sdc}(\tau_{sdc}) \quad (52)$$
$$\theta_{LRU,L}^* \in \left\{\theta>0\middle|\frac{R_a\ln[sp(\mathbf{M}^\theta)]}{\theta} \leq \frac{\ln[e^{-\theta\sum_{i=1}^{L}(\tau_i\kappa_{S_i}-S_i(\tau_i))}]}{-\theta}\right\}. \quad (53)$$

APPENDIX G
PROOF THEOREM 6

Proof: According to Eq.(8), consider an arriving supermartingale package $M_{a^t}(n)$ and a service supermartingale package $M_{s_i}(n)$ These two martingale processes can be expressed as follows:

$$M_{A^t}(n) = h_{A^t}(a_n^f)e^{\theta_G^*(\{(A^t(n,\kappa)\text{-}(n-\kappa)R_a\})},$$
$$M_{S_i}(n) = h_{S_i}(s_n)e^{\theta_G^*(\{(\tau_i\kappa_{S_i}-S_i(\tau_i))\})}.$$

Here, it is assumed that the arrival process and the service process are independent. Thus, the process $M(n)$ is also a supermartingale process in time domain $V := \kappa, \kappa+1, \ldots$. This stochastic process can be expressed as follows:

$$M(n) = h_{A^t}(a_n^t)\prod_{i=1}^{L}h_{S_i}(s_n)e^{\theta_G^*(\{(A^f(n,\kappa)\text{-}(n-\kappa)R_a\})}$$
$$\times e^{\theta_G^*(\{(\sum_{i=1}^{L}(\tau_i\kappa_{S_i}-S_i(\tau_i))\})}. \quad (54)$$

Therefore, according to the stopping time theorem, there is the following derivation where, $H := \min\{h_{A^t}(a_n)\prod_{i=1}^{L} h_{S_i}(s_n) : a_n - \prod_{i=1}^{L}h_{S_i}(s_n) > 0\}$.

APPENDIX H
FUNDAMENTAL THEORY OF MARTINGALE PROCESS

Martingale Process: Consider $X_0, X_1, X_2, \cdots, X_{n'}$ is a series of random variables; if a sequence $\{X_i\}$ satisfies Eq.(55), this sequence is called the martingale process [15].

$$E(X_{i+1}|X_1,X_2,\cdots,X_i) = X_i. \quad (55)$$

Furthermore, if the sequence $\{X_i\}$ satisfies Eq.(56), then this sequence is called a supmartingale.

$$E(X_i|X_1,X_2,\cdots,X_{i-1}) \leq X_{i-1}. \quad (56)$$

Stopping Theorem: Suppose T is a non-negative integer, and a stop-time condition of a sequence $\{Z_n, n>0\}$ is the event $T = n$ that relies solely on random variables $Z_1, Z_2, \cdots, \cdots Z_n$. In the time slot from 0 to n, the arrival data are expressed as $A(n) = \sum_{i=1}^{i=n}A_i(n)$, and the leaving data are represented as $D(t) = \sum_{j=1}^{j=m}D_j(n)$. Given a value of σ, the stop time N represents the time when the queue length $B(n)$ exceeds σ first.

REFERENCES

[1] O. Kodheli, E. Lagunas, N. Maturo, S. K. Sharma, B. Shankar, J. F. M. Montoya, J. C. M. Duncan, D. Spano, S. Chatzinotas, S. Kisseleff, J. Querol, L. Lei, T. X. Vu, and G. Goussetis, "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, Oct. 2021.

[2] W. Zhang, Y. Xue, J. Wu, and X. Xu, "Satellite as a service: A hybrid resource management framework for space-terrestrial integrated networks," in *Proc. IEEE 11th Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Oct. 2020, pp. 171–174.

[3] H. Wu, J. Li, H. Lu, and P. Hong, "A two-layer caching model for content delivery services in satellite-terrestrial networks," in *Proc. IEEE Globecom.*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[4] S.-G. Chang, "Caching strategy and service policy optimization in a cache-satellite distribution service," *Telecommun. Syst.*, vol. 21, no. 2, pp. 261–281, Dec. 2002.

[5] A. Armon and H. Levy, "Cache satellite distribution systems: Modeling, analysis, and efficient operation," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 2, pp. 218–228, Feb. 2004.

[6] B. J. Dutson, "Satellites: The evolution from TV broadcasting to multimedia," in *Proc. Int. Broadcast. Conv. Conf. Publ.*, Sep. 1996, pp. 560–564.

[7] H. Linder, H. D. Clausen, and B. Collini-Nocker, "Satellite internet services using DVB/MPEG-2 and multicast web caching," *IEEE Commun. Mag.*, vol. 38, no. 6, pp. 156–161, Jun. 2000.

[8] B. Klasen, "Sattorrent: Satellite-aided P2P content distribution for virtual environments," in *Proc. 6th Int. ICST Conf. Simul. Tools Techn.*, Brussels, Belgium, 2013, pp. 146–152.

[9] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 26–36, Jul. 2012.

[10] R. Ravindran, A. Chakraborti, S. O. Amin, A. Azgin, and G. Wang, "5G-ICN: Delivering ICN services over 5G using network slicing," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 101–107, May 2017.

[11] A. Seetharam, "On caching and routing in information-centric networks," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 204–209, Mar. 2018.

[12] S. Liu, X. Hu, Y. Wang, G. Cui, and W. Wang, "Distributed caching based on matching game in leo satellite constellation networks," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 300–303, Feb. 2018.

[13] L. Boudec, J. Yves, and T. Patrick, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*. Berlin, Germany: Springer-Verlag, 2001.

[14] F. Poloczek and F. Ciucu, "Scheduling analysis with martingales," *Perform. Eval.*, vol. 79, pp. 56–72, Sep. 2014.

[15] F. Ciucu, F. Poloczek, and J. Schmitt, "Sharp per-flow delay bounds for bursty arrivals: The case of fifo, sp, and edf scheduling," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2014, pp. 1896–1904.

[16] F. Poloczek and F. Ciucu, "Service-martingales: Theory and applications to the delay analysis of random access protocols," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 945–953.

[17] Y. Hu, H. Li, Z. Chang, R. Hou, and Z. Han, "End-to-end backlog and delay bound analysis using martingale for internet of vehicles," in *Proc. IEEE Conf. Standards Commun. Netw. (CSCN)*, Sep. 2017, pp. 98–103.

[18] Y. Hu, H. Li, Z. Chang, and Z. Han, "Scheduling strategy for multimedia heterogeneous high-speed train networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3265–3279, Apr. 2017.

[19] A. Abdi, W. C. Lau, M.-S. Alouini, and M. Kaveh, "A new simple model for land mobile satellite channels: First- and second-order statistics," *IEEE Trans. Wireless Commun*, vol. 2, no. 3, pp. 519–528, May 2003.

[20] K. An, M. Lin, W.-P. Zhu, Y. Huang, and G. Zheng, "Outage performance of cognitive hybrid satellite-terrestrial networks with interference constraint," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9397–9404, Nov. 2016.

[21] G. Choudhury and S. Rappaport, "Diversity ALOHA—A random access scheme for satellite communications," *IEEE Trans. Commun.*, vol. COM-31, no. 3, pp. 450–457, Mar. 1983.
[22] J. Bai and G. Ren, "Polarized MIMO slotted ALOHA random access scheme in satellite network," *IEEE Access*, vol. 5, pp. 26354–26363, 2017.
[23] H. Baek, J. Lim, and S. Oh, "Beacon-based slotted ALOHA for wireless networks with large propagation delay," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2196–2199, Nov. 2013.
[24] M. Lee, J. Lee, H. Noh, and J. Lim, "Stability of reservation-contention resolution diversity slotted ALOHA for satellite networks," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2012, pp. 1–6.
[25] M. Lee, J.-K. Lee, J.-J. Lee, and J. Lim, "R-CRDSA: Reservation-contention resolution diversity slotted ALOHA for satellite networks," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1576–1579, Oct. 2012.
[26] O. del Rio Herrero and R. De Gaudenzi, "A high-performance MAC protocol for consumer broadband satellite systems," in *Proc. IET Conf. Publications*, 2009, p. 512.
[27] E. Casini, R. D. Gaudenzi, and O. D. R. Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme-for satellite access packet networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1408–1419, Apr. 2007.

ERBAO WANG received the M.S. degree in control engineering from Xidian University, Xi'an, China, in 2007, and the Ph.D. degree in military communications from the State Key Laboratory of Integrated Services Networks, Xidian University, in 2019. He is currently with the School of Physics and Electronic Engineering, Xianyang Normal University. His current research interests include satellite networks, game theory, edge computing, and martingale theory analysis.

QUNFENG DONG was born in China, in 1977. He received the M.Sc. degree in radio physics from Xidian University, Xi'an, China, in 2006, and the Ph.D. degree in electromagnetic field and microwave technique from Northwestern Polytechnical University, Xi'an, in 2011. He is currently a Professor with the Department of Electrical and Electronic Engineering, Xianyang Normal University. His current research interests include electromagnetic theory, radio wave propagation, and scattering in random media.

YONGAN LI was born in 1961. He received the B.S. degree from Northwestern University, Xi'an, in 1983. He is currently a Professor with the School of Physics and Electronic Engineering, Xianyang Normal University, Xianyang. He has authored several international journal articles and has acted as a reviewer for several international journals. His research interests include analog signal processing circuits and current-mode circuits.

YUYE ZHANG received the master's degree in control science and engineering from Northwestern Polytechnical University, in 2009. She is currently an Associate Professor with Xianyang Normal University. She has been engaged in the studies of signal processing and integrated circuit design.