

第 3 章 基于多源业务特性预测的地面 资源动态分配

3.1 引 言

在空间信息网络中,如何设计协作机制,以达到高效的资源分配,并实现网络资源利用率和利用效能最大化,对提高网络高时效数据获取与传输性能尤为重要。第 2 章从认知与传输能力增强出发,通过对基于中继卫星的多接入协作传输系统进行性能分析和传输资源优化配置,实现了网络对高时效数据传输业务服务能力的提升。然而,随着卫星系统中各个应用领域传输业务需求日益增长,不同传输业务,特别是突发性业务、多媒体业务,对空间信息网络动态资源分配机制的高效性、自适应性以及传输质量提出了更严峻的挑战。以台风监测跟踪卫星系统为例,目前我国已部署了大量太阳同步轨道(SSO)卫星(如风云三号卫星 FY-3A/B)以及地球同步轨道(GEO)卫星(如风云二号卫星 FY-2G/E/F 和风云四号卫星 FY-4),用以获取台风在发生-发展-消散过程中的雷达或光学图像;此外,作为当今世界上 GEO 分辨率最高的对地观测卫星,我国部署于 GEO 的高分四号卫星(GF-4)搭载有 50 m 分辨率的秒级凝视成像设备,并具备准视频成像功能,能够对台风运动轨迹实施连续动态视频获取^[98]。

目前,我国不同系列的卫星分属于不同部门,卫星在获取上述这些图像或视频等多媒体数据后,需要及时回传至地面站,并通过地面站分发至相应的管理部门或不同数据用户。然而,在空间信息网络中,地面传输服务资源(如传输功率和服务速率)是有限的。同时,卫星与地面站无法建立持续稳定的连接也从传输资源可用性方面导致了传输资源受限的约束。

在这些约束下,针对不同业务需求,恰当的资源分配策略应当保证地面传输服务资源利用率的最大化以及每颗卫星回传数据延迟的最小化。因此,与第 2 章讨论对象类似,本章将继续围绕多颗信源卫星接入场景下空间信息网络资源动态优化分配问题展开研究。在第 2 章中,本书对中继卫星的传输资源进行基于认知协作的分配,从而实现了系统协作传输能力增强。本章将围绕不同业务传输需求,对地面传输服务资源进行优化分配,从而实现资源高效利用和传输服务时效性提高。在本章讨论场景中,地面站作为云处理服务中心,可以感知不同接入卫星的数据流量到达以及信道状态信息(channel state information, CSI),并为多颗接入卫星不同特性的业务数据提供处理和传输,进而将数据分至不同的数据需求部门。这里假设地面云处理的流量感知能力是合理的,这一感知能力可以通过现有流量感知技术得以实现,如动态流量检测^[99]、基于 KL 散度(Kullback-Leibler divergence, KLD)测量的数据包采样技术^[100]、基于感知序列置信区间的流量估计技术^[101]。

目前研究表明,对业务数据流量特性的学习和预测以及对未来到达数据的预服务(predictive service)机制能够有效提高传输系统性能^[35]。因此,本章将对卫星传输业务的数据流量特征进行挖掘,将训练得到的流量特征用于对未来到达数据量的预测,并基于该预测引入预服务机制,研究基于视频流量预测的地面传输服务资源动态优化分配策略。具体地,本章的研究内容和贡献主要为以下三点:

(1) 针对空间信息网络中的视频传输业务设计了基于云处理的地面站传输服务资源分配机制。在该场景中,云处理服务器能够感知卫星接入及接入卫星数据到达、预测未来视频流量到达以及实施基于预测信息的资源分配策略。

(2) 建立了基于多层小波分解及 BP(反向传播算法, backpropagation)神经网络的视频流量预测系统。在该系统中,到达的流量序列首先依据不同分辨率进行多层离散小波分解,分解得到的数据流量的低频和高频成分分别作为 BP 神经网络的输入进行预测,最终得到对数据流量未来到达的预测。

(3) 针对空间信息网络中的视频传输业务,本章基于数据流量的预测信息提出了基于未来流量预测的地面传输服务资源分配策略。该策略

引入背压原理及算法,将预测信息应用于当前资源分配决策中,并考虑空间信息网络的能量损耗及信道传输状态信息,对资源进行动态优化分配,以实现网络传输效率及资源利用率等方面的性能优化。

本章内容安排如下。3.2节首先建立了系统模型。3.3节设计了基于多分辨率小波分解及BP神经网络的流量预测系统,用以视频流量预测。在3.4节中,结合基于背压原理的预服务机制,详细分析并提出了面向空间信息网络的地面传输服务资源分配机制。3.5节通过仿真实验验证了所设计的预测系统性能以及资源分配协议对传输延迟的改进。3.6节为本章总结。

3.2 系统模型

本章讨论基于地面云处理的空间信息网络,对云处理资源的控制管理能力能够有效提高系统的规模可扩展性和能力可伸缩性。空间信息网络运行于高动态的网络环境中。一方面,由于高速变化的网络拓扑,星间链路(ISL)与星地链路(SGL)难以保持稳定和连续的连接;另一方面,随着空间信息网络基础设施建设的不断推进,网络中卫星和相应数据传输量的规模不断增长,并通过卫星的更新换代来实现网络获取和传输能力增强。这些卫星设施及功能的进步将对整个网络的优化管理控制带来新的挑战。云处理能够提供泛在的任务驱动的网络接入,实现网络资源配置的全局优化,并能适应网络规模、能力扩展。因此,本章引入地面云处理服务,对卫星接入信息和数据流量进行感知,并对其传输服务资源进行分配,通过基于云处理的网络架构,实现对网络的可扩展性和自适应性的增强。

如图3.1所示,本章建立基于云处理服务的空间信息网络多接入排队系统。在该系统中,地面处理服务器能够获取 $N < \infty$ 颗接入信源卫星的流量信息,这些卫星($i = 1, 2, \dots, N$)将使用地面服务器的资源实现数据处理和传输,地面服务器用 d 表示。在这项研究中, N 颗信源卫星可以部署在不同轨道,地面服务器作为云处理中心部署在地面站,是信源卫星传输数据的信宿节点,并能实现卫星数据流量感知、视频流量预测以及传输处理资源分配的功能。此外,本章使用时隙划分方式对系统运行过

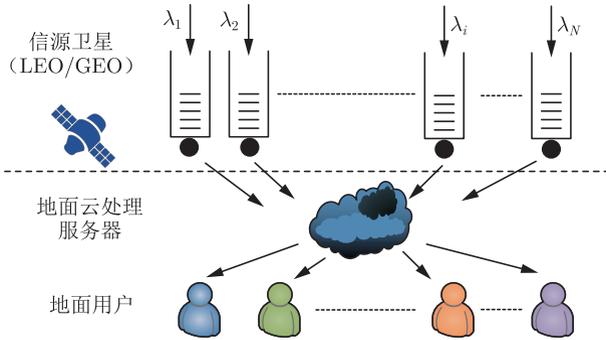


图 3.1 基于云处理的多接入卫星排队系统

程进行分析，用整数序列 $t \in \{0, 1, 2, \dots\}$ 表示各个时隙。

3.2.1 流量模型

将不同卫星的视频数据流量作为系统各个接入卫星的数据到达，使用 $A_i(t)$ 表示在时隙 t 到达卫星 i ($i = 1, 2, \dots, N$) 的新的数据包总量，以时隙长度作为单位时间长度，该数据包总量在本项研究中作为到达的数据流量。用 $\mathbf{A}(t) = [A_1(t), A_2(t), \dots, A_N(t)]$ 表示在时隙 t 所有接入信源卫星的数据流量向量。假设每颗卫星在不同时隙的数据到达是独立同分布的 (i.i.d.)，并使用 $\lambda_i = E\{A_i(t)\}$ 表示卫星 i 的数据到达率。这里假设不同卫星的数据到达过程是不相关的。此外，对所有信源卫星 i 及任意时隙 t ，假设存在 A_{\max} ，使 $0 \leq A_i(t) \leq A_{\max}$ 。

3.2.2 信道模型

在本章所讨论场景中，考虑卫星与地面站传输的视距 (LOS) 路径信号能量与其他散射信号相比在接收信号中占主要成分，因此，与 2.2.3 节的信道建模相似，本章仍使用莱斯衰落与加性高斯噪声 (AGWN) 信道对卫星与地面站之间的无线信道进行建模。地面站 d 在 t 时隙接收到来自信源卫星 i 的信号表示为

$$y_i^t = \sqrt{Gl_i^{-\gamma} h_i^t} x_i^t + n_i^t \quad (3-1)$$

其中， x_i ($i = 1, 2, \dots, N$) 表示第 i 颗卫星的发送数据， G 为传输功率， l_i 表示卫星 i 与地面站 d 之间的距离， γ 表示路径衰落指数；此外， n_i^t 表

示在 t 时隙, i 与 d 之间独立同分布的加性高斯噪声, 其均值为零, 方差为 N_0 ^[85-86]。与式 (2-10) 中定义相同, 在式 (3-1) 中, $h_i = X_1 + jX_2$ 建模为循环对称复高斯随机 (circularly symmetric complex Gaussian) 变量, 表示信道衰落系数。类似地, 可以推导得到卫星 i 与地面站 d 之间的传输信噪比 (SNR) 为

$$\text{SNR}_i = |h_i|^2 l_i^{-\gamma} G / N_0 \quad (3-2)$$

其中, $|h_i|^2$ 服从非中心卡方分布 (non-central Chi-square distribution, χ^2), 其概率密度函数如式 (2-13) 所定义。与 2.2.3 节分析方式类似, 本节定义了当 SNR 阈值 β 给定时, 卫星 i 与地面站 d 之间数据包成功传输接收的概率:

$$\begin{aligned} f_i \triangleq \Pr \{C_i\} &= \Pr \left\{ |h_i|^2 \geq \frac{\beta N_0 l_i^\gamma}{G} \right\} \\ &= \int_{\frac{\beta N_0 l_i^\gamma}{G}}^{+\infty} \frac{K+1}{\Omega} \exp \left[-K - \frac{(K+1)h}{\Omega} \right] \times \\ &\quad \text{I}_0 \left[2\sqrt{\frac{K(K+1)h}{\Omega}} \right] dh \end{aligned} \quad (3-3)$$

其中, C_i 表示卫星 i 与地面站 d 之间数据包成功传输接收事件。

3.2.3 预服务模型

本节依据图 3.1 所建立的基于云处理的多接入卫星排队系统, 讨论针对卫星到达数据流量的预服务机制。本章分配的资源是地面云处理器的传输服务资源, 以及功率和相应的传输服务率。因此, 本节对云服务器用于数据包处理的功率在不同时隙为各个接入信源卫星的预分配机制展开讨论。令 $P_i(t)$ 表示在时隙 t 为卫星 i 分配的用于数据包传输服务的功率, 则云服务器的功率分配向量可表示为 $\mathbf{P}(t) = [P_1(t), P_2(t), \dots, P_N(t)]^T$ 。

接下来分析卫星与地面站之间的传输链路状态。空间信息网络属于高动态系统, 信道衰落系数 h_i 以及来自不同卫星的不同数据传输服务请求等因素都会随时间动态变化。这种情况会对云服务器产生不同的功率和其他资源消耗, 以及不同服务速率请求。为刻画链路连接与传输状态的变化, 使用 $S_i(t)$ 表示在 t 时隙信源卫星 i 与地面站之间的链路

状态。因此, 系统中 N 颗卫星与地面站的链路状态可用向量 $\mathbf{S}(t) = [S_1(t), S_2(t), \dots, S_N(t)]^T$ 表示。假设 $\mathbf{S}(t)$ 在状态集合 $\{s_j\}_{j=1}^K$ 中取值, 相应的取值概率为 $p_{s_j}^i(t) = \Pr\{S_i(t) = s_j\}$ 。当链路状态为 s_j 时, 云处理器的功率分配向量 $\mathbf{P}(t)$ 在集合 $\mathcal{P}^{(s_j)}$ 中取值, 且对所有 $P_i(t) \in \mathbf{P}(t)$, 均存在常数 $P_{\max} > 0$, 使对所有卫星 $i = 1, 2, \dots, N$, 在任意时隙 t 均满足 $0 \leq P_i(t) \leq P_{\max}$ 。

假设云处理器能够感知链路状态并从相应的功率分配集合 $\mathcal{P}^{(s)}$ 中采取适合的功率分配策略。给定链路状态向量 $\mathbf{S}(t)$ 和功率分配向量 $\mathbf{P}(t)$, 定义云服务器在单一时隙内对每颗接入卫星所能提供的数据包传输总量, 即数据包服务率, 表示为

$$\mu_i(t) = \mu_i(S_i(t), P_i(t)), \quad \forall i = 1, 2, \dots, N \quad (3-4)$$

其中, $\mu_i(S_i(t), P_i(t))$ 是 $S_i(t)$ 和 $P_i(t)$ 的连续函数。假设存在约束 $0 \leq \mu_i(S_i(t), P_i(t)) \leq \mu_{\max}$, 对所有卫星 i 、时隙 t 以及 $S_i(t)$ 和 $P_i(t)$ 均成立。在这项研究中, 令 $S_i(t) \in \{1, 2\}$, 其中, $S_i(t) = 1$ 表示卫星 i 与地面站 d 之间的传输链路不存在, 此时, 将为该卫星到达的数据提供低的服务率; $S_i(t) = 2$ 表示该链路存在, 此时, 分配给卫星 i 的服务率将高于 $S_i(t) = 1$ 状态下分配的服务率。根据 3.2.2 节中的分析, 令

$$p_1^i(t) = 1 - f_i(t), \quad p_2^i(t) = f_i(t) \quad (3-5)$$

其中, $f_i(t)$ 由式 (3-3) 定义。本项研究中, 考虑云处理器的数据包服务率与链路状态 $S_i(t)$ 和分配功率 $P_i(t)$ 的关系为^[35]

$$\mu_i(t) = \lfloor \log(1 + S_i(t) P_i(t)) \rfloor \quad (3-6)$$

由于本章定义数据服务率为单时隙内云服务器可以服务的数据包数量, 因此, 式 (3-6) 中, $\lfloor x \rfloor$ 定义为不超过 x 的最大整数。

大多数关于多接入队列系统传输资源分配的研究主要基于不同队列当前或已到达的数据进行资源动态优化, 这种资源分配机制会由于业务数据流量的随机性导致系统性能降低, 如严重的传输延迟, 特别是在传输具有突发性特征的视频业务数据时, 系统中各队列中发生拥塞的概率将会提高。在本章中, 我们主要讨论由于排队系统的队列等待时间导致的传

输延迟。为了降低传输延迟、提高业务传输的服务质量, 本研究引入预服务机制, 即对未来可能到达的数据提供预测, 并提前分配传输资源, 从而实现预服务机制^[102-103]。

假设地面站云服务器能够对每颗卫星未来数据到达进行预测, 并基于预测的流量信息预先对这些队列分配功率资源实现预服务。令 $D_i \geq 1$ 表示对第 i 颗信源卫星数据流量进行预测的预测窗口长度。则在任意时隙 t , 地面云服务器对于每颗信源卫星 $i = 1, 2, \dots, N$ 可以预测的流量信息前向窗口可表示为 $\{A_i(t), A_i(t+1), \dots, A_i(t+D_i-1)\}$, 其中, $A_i(t+1), \dots, A_i(t+D_i-1)$ 表示在未来时隙预测的数据到达量, 即流量。假设在任意时隙的到达数据仅能在之后的时隙对其进行传输服务。在任意时隙 t , 令 $\mu_i^{(\tau)}(t)$ ($\tau = 0, 1, \dots, D_i - 1$) 表示在时隙 t 预先分配给未来到达数据 $A_i(t+\tau)$ 的服务率, 此外, 用 $\mu_i^{(-1)}(t)$ 表示分配给已经到达并在队列中数据的服务率。对所有 $\mu_i(t)$, 以下条件均成立:

$$\sum_{\tau=-1}^{D_i-1} \mu_i^{(\tau)}(t) \leq \mu_i(t) \quad (3-7)$$

3.2.4 排队模型

令 $Q_i(t)$ 表示 t 时隙在云服务器中来自于第 i 颗信源卫星的数据包队列长度, 则该队列长度的动态变化过程可表示为

$$Q_i(t+1) = \max \left\{ Q_i(t) - \mu_i^{(-1)}(t), 0 \right\} + A_i^{(-1)}(t) \quad (3-8)$$

其中, $A_i^{(-1)}(t)$ 表示第 i 颗卫星到达数据在经过一系列预服务过程后实际进入等待队列的数据包总量。具体来说, 使用之前预先分配的传输服务率对到达队列的数据包进行服务, 该处理过程可归纳为以下两种情况:

(1) $-1 \leq \tau \leq D_i - 2$:

$$A_i^{(\tau)}(t) = \max \left\{ A_i^{(\tau+1)}(t) - \mu_i^{(\tau+1)}(t - \tau - 1), 0 \right\} \quad (3-9)$$

(2) $\tau = D_i - 1$:

$$A_i^{(\tau)}(t) = A_i(t) \quad (3-10)$$

本研究将系统稳定性定义为系统中平均队列长度有限, 见定义 3.1。

定义 3.1 在 N -队列排队系统中, 令 $E\{Q_i(t)\}$ 表示队列 i 的平均队列长度, 当

$$\bar{Q} \triangleq \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{i=1}^N E\{Q_i(\tau)\} < \infty \quad (3-11)$$

时, 系统是稳定的。

通过上述讨论, 针对本章研究场景, 系统资源分配的优化目标是要在系统队列稳定的约束条件下, 找到一种功率分配和管理机制, 能够实现平均时间“代价”(成本, 延迟)最小。该优化问题可建模为

$$\min \bar{f}_c = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} E\{f_c(\mathbf{S}(\tau), \mathbf{P}(\tau))\} \quad (3-12a)$$

$$\text{s.t. } \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{i=1}^N E\{Q_i(\tau)\} < \infty \quad (3-12b)$$

其中, $f_c(\mathbf{S}(t), \mathbf{P}(t))$ 表示给定 $\mathbf{S}(t)$ 和 $\mathbf{P}(t)$ 条件下, 在 t 时隙由于功率损耗对云服务器产生的“服务成本”。在这项研究中, 将 t 时隙云服务器的功率损耗总和作为云服务器的服务成本:

$$f_c(\mathbf{S}(t), \mathbf{P}(t)) = \sum_{i=1}^N P_i(t) \quad (3-13)$$

假设存在 $f_{c \max}$, 对于所有 t 、 $\mathbf{S}(t)$ 和 $\mathbf{P}(t)$, $f_c(\mathbf{S}(t), \mathbf{P}(t)) \leq f_{c \max}$ 均成立。

3.3 基于离散小波分解的 BP 神经网络流量预测系统

视频流量具有突发特性 (burst), 表现为在短时间内随机集中产生大量数据包。同时, 视频流量的自相关呈现出双曲衰落的特征 (hyperbolic decay), 即长时相关特征。对这类突发、长时相关随机业务流量的准确预测是存在困难的。

针对这些特性, 本节引入离散小波变换 (DWT) 对视频流量的特征进行分解, 实现不同尺度上预测, 从而提高流量预测准确度。DWT 可将随机信号序列在不同维度和分辨率进行分解, 使用近似系数 (approximation

coefficients) 和细节系数 (detail coefficient) 实现表征, 近似系数对应于信号的大尺度、低频分量, 细节系数对应于信号的小尺度、高频分量。通过这种分解, DWT 能够同时实现对信号在时域和频域上的表征, 可以有效用于非平稳信号的分析处理。视频流量的长时相关特性可以通过这种多维度的分解过程实现有效提取和分析。同时, 对提取出的不同维度的流量特征, 可以通过人工神经网络 (ANN) 对其学习并完成预测。首先, ANN 提供了一种非线性的方法对模式进行学习, 获得输入与输出的隐含函数关系, 这种关系是不可知或不可解析刻画的。此外, ANN 可以对非平稳的动态过程进行建模, 进而解决视频流量强相关性带来的预测困难。因此, 本节将引入多维小波分解设计基于反向传播算法 (BP) 的 ANN 预测系统, 即 DWT-BP 预测系统, 用于视频流量的准确预测。DWT-BP 预测系统作为云处理中心的一个组成部分, 为下一阶段的功率及服务率分配提供流量预测信息。

3.3.1 多层小波分解

本节将简要介绍用于视频流量分解的 DWT 技术。小波变换使用小波母函数通过调整平移参数和尺度参数对信号进行分解。DWT 是通过对连续小波变换 (continuous wavelet transform, CWT) 经过离散化得到, 对连续信号 $x(t)$ 的 CWT 定义为

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \phi\left(\frac{t-b}{a}\right) dt \quad (3-14)$$

其中, $a > 0$ 是表征小波伸缩性的尺度 (scaling) 参数, $b > 0$ 是由中心位置决定的平移 (translation) 参数, $\phi(t)$ 表示小波母函数。离散信号 $x(k)$ 的 DWT 定义为

$$(m, n) = \frac{1}{\sqrt{2^m}} \sum_{k=0}^{T-1} x(k) \phi\left(\frac{k-n \cdot 2^m}{2^m}\right) \quad (3-15)$$

其中, T 表示信号长度, k 表示离散时间刻度, 整数参数 m 和 n 分别为离散尺度参数和平移参数 ($a = 2^m$, $b = n \cdot 2^m$) [104]。通过使用有效的滤波算法 [105], DWT 可以实现对离散信号在不同频率使用不同分辨率的分析。具体来说, 该算法使用低通和高通滤波, 将离散信号分解为近似分量

和细节分量，分别表示信号中的低频和高频分量。对离散信号 $x(k)$ ，这一处理过程可表示为

$$y_{\text{high}}(k) = \sum_{n=-\infty}^{+\infty} x(n) h(2k - n) \quad (3-16a)$$

$$y_{\text{low}}(k) = \sum_{n=-\infty}^{+\infty} x(n) g(2k - n) \quad (3-16b)$$

其中， $h(k)$ 、 $g(k)$ 分别表示高通滤波器和低通滤波器。通过信号分解，得到的近似分量反映了信号的慢变特征，并可以重复式 (3-16b) 所示分解过程进行多层分解。图 3.2 表示一个双层 DWT 信号分解过程，图中， A_1 、 A_2 表示第一层和第二层的近似分量， D_1 、 D_2 表示第一层和第二层的细节分量。通过降采样过程，如图中“ $\downarrow 2$ ”符号所示，信号的长度保持不变。

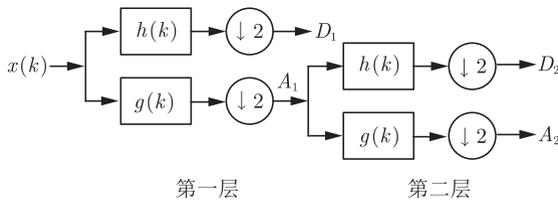


图 3.2 双层离散小波分解

3.3.2 BP 神经网络预测

反向传播算法 (BP) 是前向反馈 ANN 中用于监督学习的重要方法。BP 神经网络由输入层 (input level)、输出层 (output level) 以及若干层隐层 (hidden level) 组成，图 3.3 给出了典型的单隐层 BP 神经网络结构，其中，输入层接收外部输入向量，该向量通过有权连接传至隐层神经元 (neuron) 节点，然后在神经元节点计算相应的激活值 (activations)，并将激活值传送至输出层或下一隐层。在这一过程中，给定的输入向量可以视为在网络中前向传播，而在输出层形成的最终激活值作为输出向量。图 3.3 中，输入层、隐层、输出层的神经元数量分别为 n 、 l 和 m ， x_i ($i = 1, 2, \dots, n$)、 z_j ($j = 1, 2, \dots, l$) 和 y_k ($k = 1, 2, \dots, m$) 分别表示相应层的激活值； v_{ji} 表示从输入层神经元 i 到隐层神经元 j 的连接权重，