

新型的基于预测的卫星资源分配方法

于佳宗鹏

(南京航空航天大学航天学院,南京,210016)

摘要:针对传统的按需分配方式所造成的长时延问题,提出了一种新型的基于业务预测的按需分配方法。该方法通过预测业务在等待队列中的长度,将预测信息发送给网络资源控制中心,再由控制中心动态地分配带宽资源,达到服务质量的要求。在阐述了按需分配和业务预测的基本算法后,通过仿真分析了不同的业务类型采用不同的预测信息发送间隔时间的时延性能,分析中考虑了实时业务比非实时业务享有更高的优先权,并且比较了常规按需分配与基于业务预测的按需系统分配方法的时延性能。仿真结果表明利用业务预测,选择合适的预测信息发送间隔时间,能够有效地改善带宽资源分配的时延性能,相比一般的按需分配方式服务质量有了明显的提高。

关键词:卫星网络;按需分配;业务预测;时延

中图分类号: TN927 **文献标识码:** A **文章编号:** 1005-2615(2012)03-0333-07

Novel Resource Allocation Scheme Based on Traffic Prediction in Satellite Systems

Yu Jia, Zong Peng

(College of Astronautics, Nanjing University of Aeronautics & Astronautics, Nanjing, 210016, China)

Abstract: A novel dynamic bandwidth management mechanism based on the traffic prediction is presented to solve the problem of long delay caused by traditional demand assigned multiple access (DAMA). Firstly, the satellite terminals predict the queue length of the buffer, then send the reports to the network control center (NCC). Finally, the resources are allocated to the terminals by NCC. The algorithm performance is validated through the simulation for different types of traffic and different intervals between two consecutive prediction reports. The priority of real time traffic is considered. The delays of DAMA based on prediction and normal case are compared. The theoretical analysis and simulation results show that the new DAMA algorithm can efficiently tackle the problem of delay between bandwidth request and bandwidth assignment.

Key words: satellite network; demand assigned multiple access; traffic prediction; time delay

今后卫星通信的一个重要目标就是为多媒体和高数据速率的 Internet 应用提供无所不在的通信方式。可以认为,新一代卫星系统将面向个人提供电信服务,与上一代卫星网络的最大区别是承载的业务由低速数据和话音业务变为 Internet 和多媒体业务^[1]。由于卫星网络系统造价昂贵,且卫星带宽资源紧张,所以在保障宽带网络各类业务服务

质量的同时,如何有效地利用卫星链路的带宽资源,是卫星网络技术需要解决的重要课题^[2]。

多址接入的主要目的是最大程度和最有效地利用带宽资源。多址接入的方式主要有固定分配、随机分配、按需分配、按需分配与固定分配相结合^[3]、按需分配与随机分配相结合、按需分配与自由分配相结合这 6 种分配方式。固定分配方式是 6

种多址分配方法里最简单的、复杂度最低、适合于终端数量较少、且业务量恒定的系统^[4]。因为每当终端需要发送业务时,不需要发出请求和竞争信道容量,接入时间很快,时延很低。但是,当终端数量增大或是业务不是太均衡,此接入方法将不再适用,尤其是对突发性业务更不适用。因为终端不可以根据业务的活跃期和静默期而改变分配的信道的数量。因此,此接入方法对卫星网络资源将会带来巨大的浪费。随机分配接入方法的优点是在业务量较低时,有较快的接入时间和较低的传输时延,为突发业务服务,终端数量大,算法比较简单。但是,当业务量增大时,会发生碰撞,在卫星网络中,碰撞会造成传输时延的增大(尤其是同步卫星),继续重传导致了性能的降低。

按需分配的方式是通过地面终端事先发送预约请求,由卫星控制中心根据地球站请求分配带宽资源,以此达到最大程度地利用带宽资源目的。相比固定分配中预留过多的带宽和随机分配中的碰撞和重传来说,按需分配不会造成带宽的浪费。另外,控制中心完全控制带宽的分配,因此相比其他策略来说,网络的健壮性很高,公平性能够很容易实现。但是按需分配方法具有一个致命的缺点,那就是时延问题。因为每个终端需要单独占用上行链路的带宽资源来发送请求获得资源分配,这样一方面增加了资源的开销,另一方面会由于竞争原因造成申请失败。对GEO卫星系统来说,要经过一个高达250 ms的时延才能完成请求。

国内外曾经出现过许多按需分配的方法^[4-9],但几乎都是采用根据业务需求进行申请的方式,无法解决申请过程所经历的长时延问题。最近Bhathi B^[10]提出了一种基于业务预测的按需分配方法。在这种方法中,终端预测未来一段时间内的业务到达率,再根据预测情况来申请带宽,而不是等业务积累到一定程度再提出申请,这样对保持业务流能够均衡占有信道起到了一定作用。但仍避免不了由于申请竞争可能造成的时延问题。并且,由于该方法只预测了下一时刻新业务的到达率,并不考虑在缓存区中等待业务的容量,因此控制中心不能根据实际情况有效地给每个终端分配带宽。

本文提出了一种新型的基于业务预测的按需分配方法。其原理是通过预测未来一段时间的业务等待队列长度,并将预测结果独立地或嵌入在发送信息中上报给卫星网络控制中心,再由网控中心统一分配资源。本方法与一般的预约按需分配(Demand assigned multiple access, DAMA)不同之处

在于无需发送请求,而是报告业务流情况,把资源分配的决定权统一交给控制中心,对于每一个终端来说等效于直接面对网络,将业务放在缓存区中,由网络来提取。这样对业务的预测及资源分配进行统一管理,避免了各个站在盲目申请时产生竞争致使申请失败造成时延,以及频繁申请导致信道的浪费,因此有效地利用了卫星资源。

1 系统模型和基本定义

1.1 基本定义和限制条件

图1中的卫星网络系统由卫星网络终端(ST),卫星网络控制中心,用户终端(UT)组成,卫星网络控制中心(NCC)主要负责处理卫星终端的业务任务,每个用户终端,通过指定的卫星终端传到卫星上。为了提高卫星系统容量的利用效率,每个连接都有非常严格的时延要求。这里,卫星接入不采用固定分配的方法,而是根据需要,由NCC周期性分配带宽。

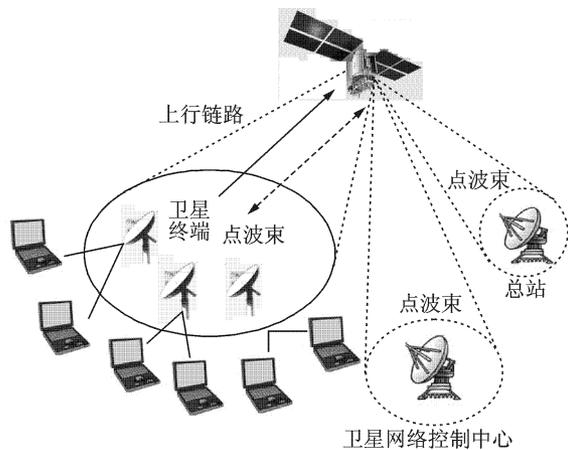


图1 卫星网络系统

为了分析方便,定义上行链路为卫星终端(ST)到卫星的链路, S 为波束内卫星终端的数量, $i=1,2,\dots,S$, i 代表一个与卫星正在建立连接的卫星终端, $C(i)$ 表示第 i 个卫星终端可以建立连接的链路数量的最大值, j 表示卫星终端正在与卫星连接的链路数量,因此可以用 (i,j) 来表示ST的上行链路连接状态。 j 的取值范围为 $[1,C(i)]$ 。

对于资源的分配通常根据申请或者观察判断来决定。而分析判断和执行要经过一定的时间间隔。在此, $t_h(h=1,2,\dots)$ 表示离散的时刻,时间间隔为 T_{short} ,下一个离散时间的时刻可以表示为 $t_{h+1}=t_h+T_{\text{short}}$ 。为了简便,下文中 t_h 均用 h 来表示。第 h 个时间间隔用 $[h,h+1]$ 来表示。

下面定义本文采用多址接入方式的两个概念

定义1 $D_{ij}(s)$ 表示连接状态 (i, j) 的接入时延,即业务在缓存区的等待时间,包括从UT发送包到ST,再等待从ST传出去这段时间。

定义2 $R_{ij}^{\text{in}}(h)$ 表示连接状态 (i, j) 中在建立连接的时间段里传送到上行链路的比特率,单位为b/s,可以表示为

$$R_{ij}^{\text{in}}(h) = \frac{\sum_{k=h-M}^h L_{ij}^{\text{in}}(k)}{M \cdot T_{\text{short}}} \quad (1)$$

式中: $L_{ij}^{\text{in}}(k)$ 表示在第 k 个时间间隔内 (i, j) 连接传送包的长度的总和,单位为bit; M 为观察间隔的次數,乘以间隔的长度 $(M \cdot T_{\text{short}})$ 即表示观察总时间。

为了计算在建立连接的这段时间内总共传送的比特率。多址接必入须满足以下条件:

(1)吞吐要求 所有连接传送的比特率不能超过波束的上行链路所有的可用容量 $R_{\text{up}}^{\text{tot}}$,如式(2)所示,其中 η 表示一个ST发送带宽请求的离散时间点, t_0 和 t_f 为连接的起始时间和终止时间。

$$\sum_{i=1}^S \sum_{j=1}^{C(i)} R_{ij}^{\text{static}}[\eta, k] \leq R_{\text{up}}^{\text{tot}} \quad t_0 \leq \eta \leq k \leq t_f \quad (2)$$

(2)时延要求 如果连接状态 (i, j) 正在传送实时业务(语音,语音会议等),就存在最大传输时延的问题,定义为 D_{ij}^{max} 为在缓存区中最大可以忍受的等待时间,如果等待时间 D_{ij} 超过 D_{ij}^{max} ,连接状态 (i, j) 的包将会被卫星终端丢弃。对实时业务来说,一定程度的丢包率是可以忍受的,但是必须最小化丢包率。

介于以上分析,实时业务的QoS参数为 R_{ij}^{static} 和 D_{ij}^{max} ,而非实时业务的QoS参数仅为 R_{ij}^{static} 。本文提出的算法目的在于最小化实时业务的丢包率和最大化非实时业务的吞吐率。假设丢包仅仅是因为等待时间超过 D_{ij}^{max} ,不存在排队的溢出问题。即每个队列都能处理所有到来的包,因此,即使在业务量繁重的时候,也不存在排队溢出的问题。

每个ST都被分配一个比特速率 R_{ij}^{static} , R_{ij}^{static} 与正在与此ST建立的连接有关。除此之外,保证动态的比特率由NCC根据ST报告的业务流情况决定,此决定建立在相应的按需分配算法上。令 $R_{\text{up}}^{\text{dyn}}$ 为可用的动态资源,动态资源为波束内没有分配的资源,可以用来动态分配给各个终端,表示为

$$R_{\text{up}}^{\text{dyn}}[\eta, k] = R_{\text{up}}^{\text{tot}} - \sum_{i=1}^S \sum_{j=1}^{C(i)} R_{ij}^{\text{static}}[\eta, k] \quad t_0 \leq \eta \leq k \leq t_f \quad (3)$$

令 $R_{ij}^{\text{dyn}}[\eta, k]$ 为在时间段 $[\eta, k]$ 内,分配给每个

连接状态 (i, j) 的动态资源,显然,在任何时间段内,都必须满足

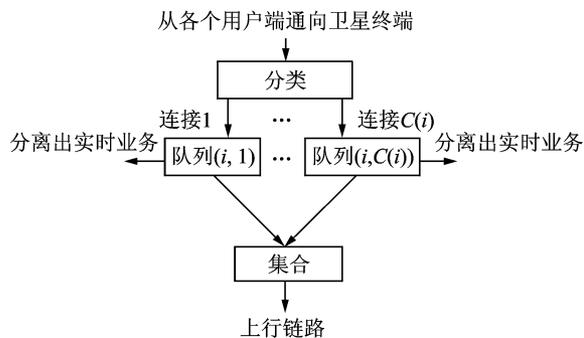
$$\sum_{i=1}^S \sum_{j=1}^{C(i)} R_{ij}^{\text{dyn}}[\eta, k] \leq R_{\text{up}}^{\text{dyn}}[\eta, k] \quad t_0 \leq \eta \leq k \leq t_f \quad (4)$$

式中: $R_{\text{up}}^{\text{dyn}}[\eta, k]$ 为上行链路动态资源的总和。

1.2 卫星终端结构

图2显示的是第 i 个ST的内部结构,ST提供 $C(i)$ 个先进先出(FIFO)队列缓存。每个缓存用来存储用户终端到来的包(等待传送上星),分类器用来分离去向各个终端的包(来自用户的),并把包派送往各个等待队列中。下面为了简洁起见,用 $queue(i, j)$ 来表示其中某个缓存器的状态, $q_{ij}(h)$ 表示第 h 个时间间隔内等待队列中的业务量。令 $\delta_{ij}^{\text{ass}}[\eta, k] \in [0, 1]$ 表示连接状态 (i, j) 在时间间隔 $[\eta, k]$ 内可用资源占动态上行链路资源 $R_{\text{up}}^{\text{dyn}}$ 的比例。在等待队列中,每个等待时间超过 D_{ij}^{max} 的包均被丢弃。对任意的时间间隔 $[\eta, k]$ 来说,必须满足以下条件

$$\sum_{i=1}^S \sum_{j=1}^{C(i)} \delta_{ij}^{\text{ass}}[\eta, k] \leq 1 \quad t_0 \leq \eta \leq k \leq t_f \quad (5)$$



2 按需分配的过程

图3是按需分配的过程示意图,在图中, d 表示一个来回的时延与时间间隔的比值,定义为 $2[D_{\text{prop}} + T_{\text{computer}}]/T_{\text{short}}$, D_{prop} 为从ST到NCC最大的传输时延, T_{computer} 为ST或者是NCC分配资源时的计算时间。为了简便,假设所有的ST同时发送请求,另外,还假设带宽请求周期性的进行。

令 T_{inf} 表示终端周期性上报业务情况的时间间隔,即为NCC动态分配资源的时间间隔,同样用与离散时间间隔的比值来表示, N 表示来回时延为 T_{inf} 的倍数, $N = d/T_{\text{inf}}$ 。图3是当 $N = 2$ 时的终端发送包的示意图。令 $R_{ij}^{\text{in}}[\eta, k]$ 表示第 i 个ST在 η 时

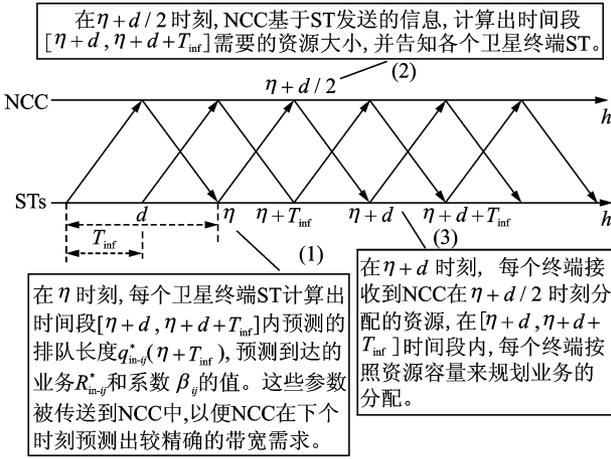


图3 卫星信道资源预测与分配管理

刻, 预测到下一时刻 k 到达缓存器 $queue(i, j)$ 的业务量。

在按需分配中, ST 只是将一些参数传送到 NCC 中, 由 NCC 统一计算带宽需求并分配带宽。这种中心控制的方法能够更公平, 更有效地分配带宽, 因此, 每当 ST 向 NCC 上报情况时, 需要提供以下信息:

(1) $q_{ij}^*(\eta+T_{inf})$ 在 η 时刻预测 $\eta+T_{inf}$ 时刻缓存器 $queue(i, j)$ 中的队列长度, 表示为

$$q_{ij}^*(\eta+T_{inf}) = q_{ij}(\eta) + \sum_{k=\eta}^{\eta+T_{inf}-1} R_{ij}^{in*}[\eta, k] \cdot T_{short} - \delta_{ij}^{ass}[\eta, \eta+T_{inf}] \cdot R_{up}^{dyn}[\eta, \eta+T_{inf}] \cdot T_{inf} \quad (6)$$

式中: $\eta+T_{inf}$ 时刻队列的长度等于 η 时刻队列的长度加上 $[\eta, \eta+T_{inf}]$ 这段时间到达队列的业务, 再减去 η 时刻 NCC 分配给 ST 的带宽。

(2) $R_{ij}^{in*}[\eta+T_{inf}, \eta+T_{inf}+d]$ 在时间间隔 $[\eta+T_{inf}, \eta+T_{inf}+L]$ 内到达缓存器 $queue(i, j)$ 的平均比特率。在 η 时刻, 第 i 个 ST 计算如下

$$R_{ij}^{in*}[\eta+T_{inf}, \eta+T_{inf}+d] = \sum_{k=\eta+T_{inf}}^{\eta+T_{inf}+d-1} \frac{R_{ij}^{in*}[\eta, k]}{d} \quad (7)$$

(3) β_{ij} 优先权系数, 为了给队列缓存器 $queue(i, j)$ 以优先权设立的参数, 主要为实时业务服务, 因为实时业务有时延限制。定义为

$$\beta_{ij} = 1 + K^{opt} \cdot \frac{B_{ij}^{loss}[\eta - T_{inf} \cdot \eta]}{B_{ij}^{out}[\eta - T_{inf} \cdot \eta]} \quad (8)$$

式中: $B_{ij}^{loss}[\eta - T_{inf} \cdot \eta]$ 表示在时间间隔 $[\eta - T_{inf}, \eta]$ 内因为时延问题从缓存器 $queue(i, j)$ 中丢弃的比特数量; $B_{ij}^{out}[\eta - T_{inf} \cdot \eta]$ 表示在此时间间隔内从 $\delta_{ij}^{ass}[\eta+d/2, \eta+d+T_{inf}]$ 接入卫星中的比特率和丢弃的比特率之和; K^{opt} 可以调节优先权的大小, 本

文取值 0.95。

从图 3 中可以看出, 根据从 ST 收到的信息, 在 $\eta+d/2$ 时刻, NCC 计算出每对 (i, j) 带宽分配 $\delta_{ij}^{ass}[\eta+d, \eta+d+T_{inf}]$ 的值, 主要目的是全部清空 $\eta+d+T_{inf}$ 时刻 ST 中等待队列中的业务。缓存器 $queue(i, j)$ 在 $\eta+d+T_{inf}$ 时刻的长度为

$$q_{ij}^*(\eta+L+T_{inf}) = q_{ij}(\eta+T_{inf}) + R_{ij}^{in*}[\eta+T_{inf}, \eta+T_{inf}+d] \times d - \{\delta_{ij}^{ass}[\eta+T_{inf}, \eta+d] + \delta_{ij}^{ass}[\eta+d, \eta+d+T_{inf}]\} \times R_{up}^{dyn}[\eta+T_{inf}, \eta+d+T_{inf}] \times d \quad (9)$$

式中: $\delta_{ij}^{ass}[\eta+T_{inf}, \eta+d]$ 表示在 $\eta+d/2$ 时刻之前, NCC 已经分配的带宽, 式(9)是假设 $d > T_{inf}$ 的情况下; $\delta_{ij}^{ass}[\eta+d, \eta+d+T_{inf}]$ 是在 $\eta+d/2$ 时刻分配的容量, 目的是清空 $\eta+d+T_{inf}$ 时刻队列中的业务, 只要使式(9)左边为 0, 即 $q_{ij}^*(\eta+d+T_{inf}) = 0$ 。因此可以算出

$$\delta_{ij}^{ass}[\eta+d, \eta+d+T_{inf}] = \frac{q_{ij}(\eta+T_{inf})}{R_{up}^{dyn}[\eta+T_{inf}, \eta+d+T_{inf}] \times d} + \frac{R_{ij}^{in*}[\eta+T_{inf}, \eta+T_{inf}+d]}{R_{up}^{dyn}[\eta+T_{inf}, \eta+d+T_{inf}]} - \delta_{ij}^{ass}[\eta+T_{inf}, \eta+d] \quad (10)$$

式(10)只有在满足以下条件时成立

$$\sum_{i=1}^S \sum_{j=1}^{C(i)} \delta_{ij}^{ass}[\eta, k] \leq 1 \quad t_0 \leq \eta \leq k \leq t_f \quad (11)$$

为了满足式(11)中的条件和考虑 β_{ij} , δ_{ij}^{ass} 的归一化计算公式如下

$$\delta_{ij}^{ass}[\eta+d, \eta+d+T_{inf}] = \frac{\beta_{ij} \cdot \delta_{ij}^{ass*}[\eta+d, \eta+d+T_{inf}]}{\sum_{i=1}^S \sum_{j=1}^{C(i)} \beta_{ij} \delta_{ij}^{ass*}[\eta+d, \eta+d+T_{inf}]} \quad (12)$$

根据以上的分析, 资源分配管理过程为以下几个阶段:

步骤 1 η 时刻: ST 根据式(6)计算出预测的 $q_{ij}^*(\eta+T_{inf})$ 长度, 根据式(7)计算出 $R_{ij}^{in*}[\eta+T_{inf}, \eta+T_{inf}+d]$, 根据式(8)计算出 β_{ij} , 然后把把这些参数上报到 NCC 中。

步骤 2 $\eta+d/2$ 时刻: NCC 接收到每个 ST 这 3 个参数的信息, $q_{ij}^*(\eta+T_{inf})$, $R_{ij}^{in*}[\eta+T_{inf}, \eta+T_{inf}+d]$ 和 β_{ij} , 然后根据式(9, 10)计算出 $[\eta+d, \eta+d+T_{inf}]$ 时间段的 $\delta_{ij}^{ass}[\eta+d, \eta+d+T_{inf}]$, 然后把对每个 ST 的带宽的分配发送给各个 ST。

步骤 3 $\eta+d$ 时刻: 第 i 个 ST 接收到来自 NCC 的带宽的分配 $\delta_{ij}^{ass}[\eta+d, \eta+d+T_{inf}]$ 。ST 可以根据每个连接分配的资源, 重新分配带宽, 分配的总带宽用 $\alpha_i[\eta+d, \eta+d+T_{inf}]$ 表示, 定义为

$$\alpha_i[\eta + d, \eta + d + T_{\text{inf}}] \leq \sum_{j=1}^{C(i)} \delta_{ij}^{\text{ass}}[\eta + d, \eta + d + T_{\text{inf}}] \quad (13)$$

3 业务预测

以上分析发现,按需分配中关键是使等式 $q_{ij}^*(\eta + d + T_{\text{inf}}) = 0$, 而 $q_{ij}^*(\eta + d + T_{\text{inf}})$ 等式中只有 $q_{ij}^*(d + T_{\text{inf}})$ 和 $R_{ij}^{\text{in}*}[\eta + T_{\text{inf}}, \eta + T_{\text{inf}} + d]$ 是未知的, 而由式(6,7)可知, $q_{ij}^*(\eta + T_{\text{inf}})$ 和 $R_{ij}^{\text{in}*}[\eta + T_{\text{inf}}, \eta + T_{\text{inf}} + d]$ 的计算关键是 $R_{ij}^{\text{in}*}[\eta, k]$ 的计算, 预测下一时刻 k 到达缓存器 $\text{queue}(i, j)$ 队列的业务到达率。

业务流量的预测方法比较多,有基于固定流量模型的,基于神经网络的,还有基于模糊理论的。本文为了降低终端的计算复杂度,采用的是计算量小,容易实现的归一化最小均方 Normalized least mean square (NLMS)算法^[11],对下一时刻 k 的业务到达率进行预测。

令 $R_{ij}^{\text{in}}(\eta)$ 为当前时刻到达队列的业务量,因为只在离散的时刻计算到达队列的业务量,间隔时间为 T_{short} , 因此 $R_{ij}^{\text{in}}(\eta)$ 为 η 时刻之前 $F \cdot T_{\text{short}}$ 时间内离散时刻预测缓存器队列中业务量值组成的矢量。 k 时刻的业务到达率的预测是基于 $R_{ij}^{\text{in}}(\eta)$ 的基础上算出,预测的队列长度

$$R_{ij}^{\text{in}*}(k) = \mathbf{w}_k^{\text{T}}(\eta) * R_{ij}^{\text{in}}(\eta) \quad (14)$$

式中: $\mathbf{w}_k^{\text{T}}(n)$ 为权矢量,尺寸为 $F \times 1$, 权矢量根据预测和真实的误差自适应地变化。

预测与真实值之间的误差为

$$\varepsilon(n) = R_{ij}^{\text{in}}(k) - \mathbf{w}_k^{\text{T}}(\eta) * R_{ij}^{\text{in}}(\eta) \quad (15)$$

式中 $R_{ij}^{\text{in}}(k)$ 为 k 时刻业务到达率的真实值。

权值通过下列方法更新

$$\mathbf{w}_q(n+1) = \mathbf{w}_q(n) + \mu(n) * \mathbf{q}(n) * \varepsilon(n) \quad (16)$$

在 NLMS 算法中, $\mu(n)$ 为步进值,可通过式(17)计算

$$\mu(n) = \frac{\mu_0}{1 + \mathbf{q}^{\text{T}}(n) * \mathbf{q}(n)} \quad (17)$$

先假定权值的初始值为 0, 然后根据上面的公式进行更新, $\mu(n)$ 和 F 分别设置为 0.01 和 10, F 值越小,复杂度越低。通过以上方法,每个用户终端根据预测的业务到达率来计算每个终端 $\eta + d + T_{\text{inf}}$ 时刻队列的长度,并根据队列的长度 NCC 安排分配给每个终端的带宽容量,因此避免了队列拥塞,并且降低了业务等待时延。

4 仿真分析

本文提出的基于预测的按需分配算法的卫星接入仿真模型是基于 OPNET 中的相关的模型。仿真场景如图 1 所示。其中的业务模型由 OPNET 的业务模型仿真产生,主要包括 5 种业务应用:两种实时业务,分别为话音和语音会议和 3 种非实时业务,分别为 FTP, Email 和 网页浏览。具体参数设置如表 1 所示。在一个卫星波束内,卫星终端数目设置为 4, 用户终端设置为 3, 且每个用户终端均同时进行 5 种业务的连接, 即 $i \in [1, 4], j \in [1, 5]$ 。对实时业务来说, 话音和语音会议最大忍受的等待时延分别为 50 和 100 ms^[12], 波束的总带宽设置为 10 Mb/s。

表 1 业务参数表

业务参数	数值
Email 间隔时间/s	exp(360)
Email 长度/kb	unif_int[1, 99]
FTP 间隔时间/s	exp(360)
网页浏览间隔时间/s	exp(60)
语音会议长度/kb	unif_int[2, 3]
话音活跃期长度	exp(0.352)
话音静默期长度	exp(0.65)

图 4 显示的是话音业务 12.5 s 的真实业务曲线和预测业务曲线, 共有 100 个离散点。 F 取值为 10, 即通过前 10 个点的值来预测下一个点的值。假设发送请求的间隔时间即为测量值的间隔时间, 当 $N=2$ 时, 测量间隔时间即 T_{inf} 为 125 ms, 从图中可以看出, 预测误差在 5% 之内。

图 5 显示的是不同 N 值情况下的等待时延的累积分布函数。 N 值的变化范围为 $[0.5, 8]$ 。 N 值越大, 表明上报需求越频繁。从图中可以看出, N 值越大, 等待时延越小, 对于实时业务, 超过最大忍受时延的概率越小, $N=4$ 以上才能满足 QoS 对时延的要求。但是由于 N 值过大, 发送上报需求过于频繁, 会造成系统开销过大。因此, N 值的选择需

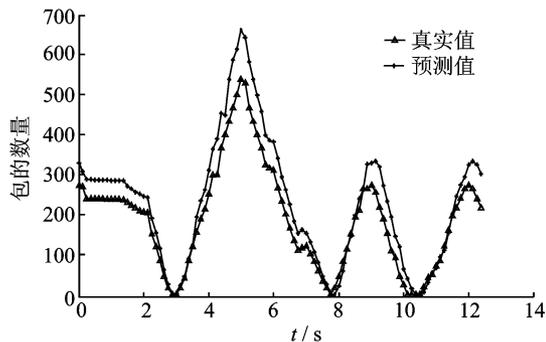


图 4 话音业务真实值与预测值对比图

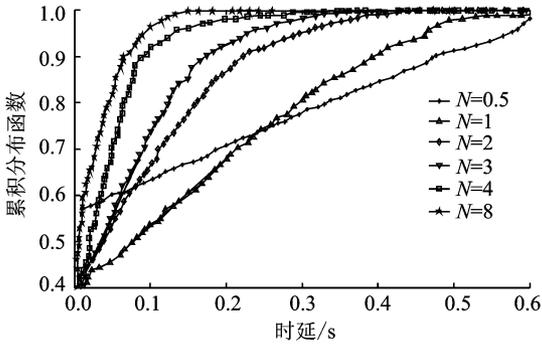


图5 不同 N 值的话音业务等待时延的累积分布函数

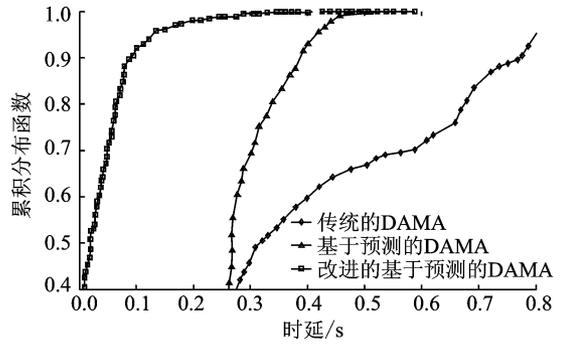


图7 不同DAMA方法的等待时延累积分布函数

要平衡时延与系统开销。

图6显示的是5种业务,包括实时和非实时业务的等待时延的累积分布函数。从图6中可以看出语音和语音会议的等待时延最短。因为语音和语音会议最大忍受的等待时延分别为50和100 ms^[9],超过该等待时延包就会被丢弃。由于语音会议的持续时间比语音长,并且要求所有用户全部接入,因此等待时延比语音长。其他3种非实时业务,由于没有时延限制,因此等待时延比实时业务长。

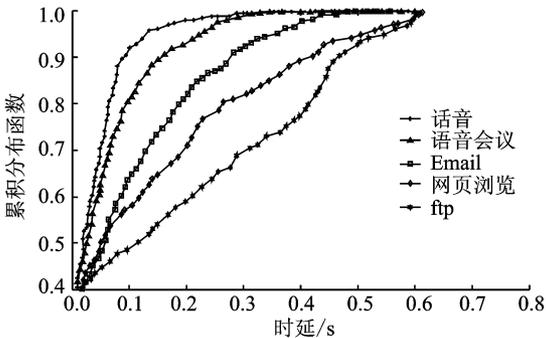


图6 5种业务等待时延累积分布函数

图7显示的是传统的DAMA,基于预测的DAMA^[10]和本文改进的基于预测的DAMA算法之间的时延性能的比较。从图中可以看出,传统的DAMA算法时延在0.25 s以上,因为终端需要得到控制中心的确认才能发送业务。基于预测的DAMA算法,有效地改进了时延问题,但由于每次都需要等待申请确认,因此时延均大于0.25 s。本文提出的通过预测业务等待队列长度来分配资源的算法,周期性地上报终端参数(等待队列长度,平均比特率,优先权等)到NCC,让NCC根据预测的情况提前分配资源给终端,避免了申请到申请确认这段时间所造成往返时延问题,有效地降低了业务在缓存区中的等待时延,并且根据终端的需要分配资源,有效地提高了资源的利用率。

5 结束语

尽管DAMA协议有许多变化,但是基本上都是由终端根据情况向网络控制中心发送预约请求,然后由网络控制中心NCC按照预约的容量分配给终端,传输结束后需要再一次预约。采用这种方法的关键问题是能否准确地判断何时预约和需求多少带宽,特别是对于申请时间的把握。由于不了解系统资源的分配情况,申请往往是盲目的。申请的不准确或频繁申请都会造成资源的浪费,因为每一次申请都要经历很长的时延,而且出现竞争后会导致更长的时延。本文提出的基于业务预测的按需分配算法,首先预测业务等待队列长度,之后不是采用申请的方式,而是采用报告的方式,由系统统一分配资源。仿真结果显示合理地选择报告的间隔时间,可以有效地解决长延时问题,尤其对具有优先级的业务有明显的改善。下一步工作还要研究业务预测信息传输带来的带宽开销对吞吐率的影响和如何进行优化,同时探讨独立报告方式和混合业务报告方式的适用条件和各自的性能,以完善预测按需分配算法。

参考文献:

- [1] Narula-Tam A. QoS considerations for future packet switched satellite communication systems with dynamic resource allocation[C]//IEEE Military Communications Conference. Atlantic City: Institute of Electrical and Electronics Engineers Inc, 2005:1727-1733.
- [2] Bharathi D. Differential availability and bandwidth management[C]//IEEE Military Communications Conference. Washington: Institute of Electrical and Electronics Engineers Inc, 2007: 7-15.
- [3] De Gaudenzi R, del Rio Herrero. Advances in random access protocols for satellite networks [C]//2009 International Workshop on Satellite and Space

- Communications. Piscataway: IEEE, 2009: 331-336.
- [4] Williams M, Nichols K. The effect of packet traffic prediction limitations on DAMA schemes for future satellite packet networks[C]//IEEE Military Communications Conference. Menlo Park: IEEE, 2009: 1-7.
- [5] Sheng Ming-Jye, Liebowitz B, Mak T. Analysis of DAMA performance for tactical radio and satellite [C]//IEEE Military Communications Conference. Piscataway:IEEE, 2009: 7-14.
- [6] Huan Yao, McLamb J, Mustafa M, et al. Dynamic resource allocation DAMA alternatives study for satellite communications systems [C]//IEEE Military Communications Conference. Piscataway: IEEE, 2009.
- [7] 于佳,宗鹏. 中国业务模型的建立和低轨卫星资源分配的研究[C]//第七届卫星通信新技术、新业务学术年会论文集. 北京:中国通信学会,2011:195-207.
Yu Jia, Zong Peng. The analysis of resource allocation in LEO satellite systems based on China traffic model [C]//The Seventh Satellite Communication Conference. Beijing: China Institute of Communications, 2011:195-207.
- [8] 李斗,王峰. 宽带卫星 Mesh 网多址接入信道预测分配方案研究[J]. 电子与信息学报,2008,30(4):763-767.
Li Dou, Wang Feng. The predictive multi-access channel allocation scheme in broadband satellite mesh network [J]. Journal of Electronics and Information Technology, 2008, 30(4): 763-767.
- [9] 李斗. 基于混沌预测的宽带 DVB-RCS 卫星接入信道动态分配方案研究[J]. 电子与信息学报,2008, 30(3):607-611.
Li Dou. The dynamic allocation of broadband DVB-RCS satellite access channel based on chaotic prediction [J]. Journal of Electronics and Information Technology, 2008, 30(3): 607-611.
- [10] Bharathi B. Predictive traffic estimation for uplink resource allocation in satellite communications[C]//Wireless and Microwave Technology Conference. USA: IEEE, 2009.
- [11] Haykin S. Adaptive filter theory [M]. 3rd Edition. New Jersey: Prentice-Hall, 1996.
- [12] Celandroni N. Suitability of DAMA and contention-based satellite access schemes for TCP traffic in mobile DVB-RCS[J]. IEEE Transactions on Vehicular Technology, 2009, 58(4): 1836-1845.