

文章编号: 2095-4980(2013)06-0969-08

基于 Markov 切换空间的分布式协同接入控制模型

刘兴华, 奚宏生

(中国科学技术大学 信息科学技术学院自动化系, 安徽 合肥 230026)

摘 要: 基于 3 层架构的映射框架, 本文以视频业务为主要研究对象, 研究了分布式协同接入控制系统的建模和调控机理。不失一般性, 假设视频业务需求近似服从泊松分布, 而服务时间服从指数分布, 将一个资源节点上的业务接入状态的演化用 Markov 过程来描述, 则所有的 N 个资源节点的分布式协同接入控制可以由一个 Markov 切换空间控制过程来阐述。针对构建的 Markov 切换空间控制系统, 基于系统灵敏度的观点对系统性能进行优化, 本文提出了基于策略梯度的随机逼近算法; 同时, 证明了该随机逼近算法的收敛性。此外, 本文给出了一个仿真实例来说明此随机逼近算法的有效性。

关键词: Markov 决策过程; Markov 切换系统; 性能势; 策略迭代

中图分类号: TN915

文献标识码: A

doi: 10.11805/TKYDA201306.0969

Distributed collaborative admission control model based on Markov switching space

LIU Xing-hua, XI Hong-sheng

(Department of Automation, School of Information Science and Technology, University of Science and Technology of China,
Hefei Anhui 230026, China)

Abstract: The modeling of distributed collaborative admission control system and mechanism are studied based on the three-layer structure of mapping framework. The video service is considered as the main research object, which is assumed approximating Poisson distribution, and service time obeys exponential distribution. A resource node in the evolution of the state of the service can be described in Markov process. Accordingly, all the N resource nodes of distributed collaborative admission control can be elaborated by a Markov switching space control process. According to the proposed Markov switching system, the performance of the system is optimized from the viewpoint of sensitivity. A stochastic approximation algorithm is put forward based on the strategy of the gradient. Meanwhile, the convergence of the algorithm is proved, and an example is given to verify the effectiveness of the algorithm.

Key words: Markov decision process; Markov switching system; performance potentials; policy iteration

随着网络和计算机技术的迅猛发展, 基于互联网的多媒体服务作为一种新型的网络服务模式, 带给广大互联网用户全新的体验和感官上的满足, 成为人们生活中不可缺少的一部分。其中, 视频点播服务(Video on Demand, VOD)是目前互联网多媒体服务的主要代表, 参见文献[1-6]。它综合了计算机技术、通信技术和电视技术的特点, 可以为用户提供实时的按需点播服务。另外, 作为一种跨媒体合作的新模式, 也成为了中国传统电信网、有线电视网和互联网进行“三网合一”的新的突破口。在这种背景下, 一种新型媒体分发模式P2P-MDN自然诞生并迅速发展起来。P2P-MDN采用了内容分发网络(Content Distribution Network, CDN)^[7]和对等网络(Peer-to-Peer, P2P)^[8-10]相融合的网络构架, 将CDN的存储设备以P2P的方式组织, 形成一个分布式的海量存储系统, 同时利用P2P的目录服务和多点传输能力, 实现CDN存储设备之间的内容交换, 提升CDN的内容分发效率。然而, 接入控制是媒体分发网络(Media Delivery Network, MDN)^[11]的重要组成部分, 它直接影响着系统的服务强度、系统运

收稿日期: 2012-11-16; 修回日期: 2013-01-22

基金项目: 国家重点基金资助项目(61233003); 中国博士研究生教育基金资助项目(20093402110019)

营的稳定性和已接入系统用户的服务质量等。一个好的接入控制策略在良好的网络资源部署环境以及任务调度策略的基础上,不仅可以使系统在最大限度上利用其现有资源,还可以显著减少响应延迟,使系统拥塞的可能性降到最低,让更多的用户享受到高质量的服务;同时,它可以减小运营成本,实现系统的良性运营。反过来从另一个方面看,一个较差的接入控制策略可能会降低服务质量,从而打击用户服务请求的积极性,甚至使系统无法适应竞争日益激烈的媒体服务市场。

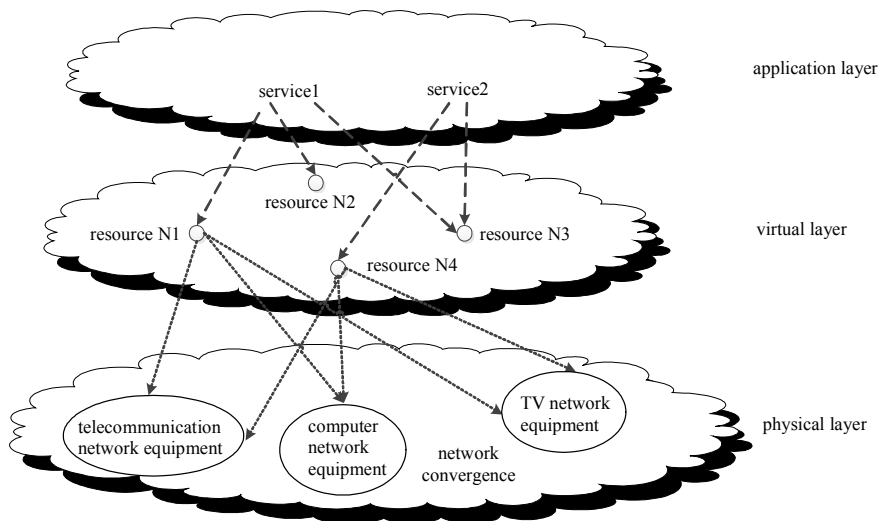


Fig. 1 Three-layer mapping framework

图 1 基于 3 层架构的映射框架

如上所述,要实现大规模视频服务系统,高性能、可扩展的视频服务器是必不可少的。视频服务系统的体系结构有多种,有单服务器结构的,也有分布式结构的。在视频服务系统中,视频数据作为连续媒体,有一些显著特点。首先,视频对象(如影片)的数据量较大,在存储和传输过程中需要较大的磁盘输入/输出(Input/Output, IO)带宽与网络带宽。其次,视频对象的播放具有连续性和实时性要求。由于服务器端的资源是有限的,因此,研究如何通过有效的接入控制策略来提升服务器的吞吐率,提高资源利用率,而且接入尽可能多的用户,又能较好地满足用户的服务质量(Quality of Service, QoS)要求,就显得非常必要。在基于单服务器的视频服务系统中,对接入控制问题已有很多研究,如文献[12-13]。但是,对于分布式视频服务器系统而言,这方面的研究和实现则相对较少。本文以视频业务为研究对象,对分布式视频服务系统中的接入控制问题进行了研究,在“业务层-虚拟层-物理层”的 3 层映射构架^[14]下,构建了 Markov 切换空间系统,基于系统灵敏度的观点对系统性能进行优化,以达到既保证用户的服务质量,又能实现多服务器间的负载平衡,改善整个系统的资源利用率的目标。

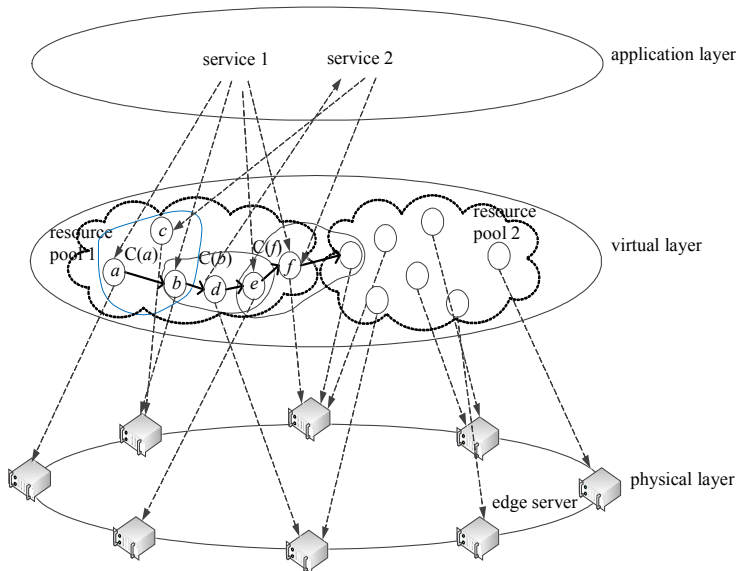


Fig. 2 Resource pool division on virtual layer

图 2 资源池划分示意图

1 问题描述

本文拟在“业务层-虚拟层-物理层”3 层映射的架构下,如图 1 所示,来研究分布式接入系统的建模和调控机理,并且注意到视频业务已成为引起接入网性能瓶颈的主要因素,因此本文以视频业务为主要研究对象。

将底层跨网部署的边缘服务器或服务器集群虚拟化为一个资源节点,各节点间的互联互通赋予了它们对网络环境具有智能感知的功能,并可以根据业务资源部署的协作准则,来组成一个协同提供服务的资源池,每个节点维护一个资源池信息表,通过节点之间的协作来实现业务资源的跨网协同。在虚拟层,假设有 N 个资源节点和 M 项业务,构成一个 $N \times M$ 的资源池信息表,如下所示:

$$\mathbf{S} = \begin{matrix} & \begin{matrix} C^1 & C^2 & \cdots & C^j & \cdots & C^M \end{matrix} \\ \begin{matrix} C_1 \\ C_2 \\ \vdots \\ C_i \\ \vdots \\ C_N \end{matrix} & \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1j} & \cdots & S_{1M} \\ S_{21} & S_{22} & \cdots & S_{2j} & \cdots & S_{2M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{i1} & S_{i2} & \cdots & S_{ij} & \cdots & S_{iM} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{Nj} & \cdots & S_{NM} \end{bmatrix} \end{matrix}$$

式中, S_{ij} , $i=1,2,\dots,N$, $j=1,2,\dots,M$ 表示第 i 个节点部署了第 j 项业务资源, $S_{ij}=0$ 表示第 i 个节点未部署支撑第 j 项业务的资源; $C_i=(S_{i1},S_{i2},\dots,S_{iM})$, $i=1,2,\dots,N$ 表示第 i 个资源节点上部署的业务资源向量; $C^j=(S_{1j},S_{2j},\dots,S_{Nj})$, $j=1,2,\dots,M$ 表示支撑第 j 项业务的资源向量。本文是在建立如图 2 所示的资源池划分以后,来构建 Markov 切换系统模型。

然后假设视频业务需求到达第 i 个资源节点近似服从泊松分布,而服务时间服从指数分布,向量 $n_i=(n_{i1},n_{i2},\dots,n_{iM})$ 表示 M 项业务的接入情况,其中 n_{ij} 表示第 i 个节点接入第 j 项业务的数量,则该业务接入状态的演化可以用有限状态 Markov 过程来描述,记 X_i^t , 其状态空间为 $\Phi_i=\{n_i=(n_{i1},n_{i2},\dots,n_{im}), n_{ij} \leq \Pi_{ij}, j=1,2,\dots,M\}$ 记 A_i 为状态转移速率矩阵, r_i 为性能函数,则这个有限状态 Markov 控制过程可以记为 $MPC_i=(\Phi_i, A_i, r_i)$, 该嵌入 Markov 链的一步转移概率矩阵的第 (n_i, n_i') 元素为

$$p_{n_i n_i'} = \begin{cases} \lambda_{ij} / \mu(n_i) & n_i' \in N^+(n_i) \\ n_{ij} u_j / \mu(n_i) & n_i' \in N^-(n_i) \\ 0 & \text{else} \end{cases} \quad (1)$$

式中 $N^+(n_i)=\{n_i': n_i'=(n_{i1}, n_{i2}, \dots, n_{ij}+1, \dots, n_{iM}), j=1,2,\dots,M\}$ 表示第 j 项业务增加了一个连接, $N^-(n_i)=\{n_i': n_i'=(n_{i1}, n_{i2}, \dots, n_{ij}-1, \dots, n_{iM}), j=1,2,\dots,M\}$ 表示第 j 项业务减少了一个连接。 λ_{ij} 是第 i 个节点上第 j 项业务的到达率, u_j 是第 i 个节点第 j 项业务的服务率。而 $\mu(n_i)=\sum_{j=1}^M (\lambda_{ij} + n_{ij} u_j)$ 表示状态 n_i 的转移速率, 则该 Markov 的无穷小矩阵为 $A_i=[a_{n_i n_i'}]$, 其中

$$a_{n_i n_i'} = \begin{cases} -\mu(n_i) & n_i' = n_i \\ \lambda_{ij} & n_i' \in N^+(n_i) \\ n_{ij} u_j & n_i' \in N^-(n_i) \\ 0 & \text{else} \end{cases} \quad (2)$$

本文将阐述这 N 个资源节点的分布式协同接入控制可以由一个 Markov 切换空间控制过程来描述,并对此 Markov 切换系统进行灵敏度分析,提出基于策略梯度的随机逼近算法。

2 主要结果

2.1 Markov 切换空间控制系统描述

考虑一个两层级的 Markov 切换空间控制过程,上层为 N 种运行模式,在一系列随机到达事件的驱动下进行切换,下层为对应于上层运行模式的一系列状态转移,其状态演化规律由上层的运行模式 $k \in \{1,2,\dots,N\}$ 确定,并可以用一个 Markov 过程来描述, N 种运行模式对应于 N 个不同演化规律的子 Markov 过程 $MPC_k: \{\phi_k, P_k, A_k\}$, $k=1,2,\dots,N$, 其中 ϕ_k 是第 k 个子 Markov 过程的状态空间,称之为子状态空间, P_k, A_k 分别是第 k 个子 Markov 过程嵌入链的一步转移概率矩阵和无穷小矩阵。

定义 $k=1,2,\dots,N$ 为系统的运行状态, $\phi_0=\{1,2,\dots,N\}$ 为运行状态空间,则由 N 个子 Markov 过程组成的大系统

状态 $X_i \in \Phi = \bigcup_{k \in \phi_0} \phi_k$ 。事件定义为满足某些公共特性的一类状态转移集合, 以 $\langle i, j \rangle$, $i, j \in \phi_k, k \in \phi_0$ 表示从内部状态 i 到 j 的转移, 激发这一状态转移的事件记为基本事件 e_{ij} , 事件 $e_n \subseteq \{\emptyset, e_{ij} : i, j \in \phi_k, k \in \phi_0\}$, 事件集合 $E = \{e_1, \dots, e_n, \dots, e_N\}$, 即基本事件 e_{ij} 发生则事件 e_n 发生; 切换控制行动 d_{kl} , $k, l \in \phi_0$ 表示从第 k 个运行状态至第 l 个运行状态的切换, 行动集合 $D = \{d_{kl}, k, l \in \phi_0\}$ 。切换控制行动由事件触发, 随机平稳 Markov 型切换控制策略 $V: E \rightarrow [0, 1]^{|\phi|}$ 定义了事件空间 E 到行动空间 D 上的概率分布的映射, $V_{d_{kl}}(e_n)$ 表示在策略 V 下, 事件 e_n 发生采取切换行动 d_{kl} 的概率。令 j^k 表示 $j \in \phi_k$, 即系统状态处于子状态空间 ϕ_k 的内部状态 j , 当系统处于状态 $i^k \in \phi_k$, 事件 e_n 发生, 选择切换行动 d_{kl} 将运行状态切换至 l , 则系统状态转移至 $j^l \in \phi_l$, 这里的下一个状态 $j^l \in \phi_l$ 由事件 e_n 和行动 d_{kl} 确定。记事件 e_n 和切换行动 d_{kl} 的状态转移为 $\langle i^k, j^l \rangle \in e_n^{kl}$, 记集合 $E_D = \{e_n^{kl}, e_n \in E, d_{kl} \in D\}$ 。考虑子空间 $\phi_k, k=1, 2, \dots, N$ 的内部状态有限, 本文首先研究子空间有交集的情形, 可以表示为: $\exists 1 \leq k_1, k_2, \dots, k_N \leq N$, $s.t. j^{k_1} = j^{k_2} = \dots = j^{k_N} \in \phi_{k_1} \cap \phi_{k_2} \cap \dots \cap \phi_{k_N}$, 这样当系统处于状态 $i^k \in \phi_k$ 且事件 e_n 发生, 选择切换行动 d_{kl} , 则系统状态转移至 $j^l = j^k \in \phi_l \cap \phi_k$ 。如下举例说明 3 个 Markov 切换空间控制过程的运行。

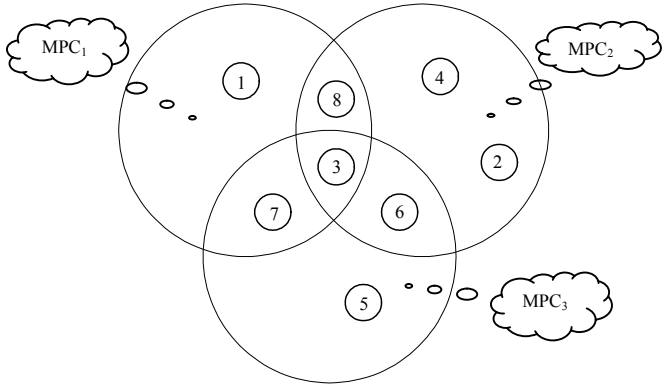


Fig.3 Switching system with three intersectant subspaces
图 3 三个子空间有交集的 Markov 切换过程示意图

如图 3 所示, $MPC_i: \{\phi_i, P_i, A_i\}, i=1, 2, 3$ 是 3 个子 Markov 过程, 即 $\phi_0 = \{1, 2, 3\}$, 其中子状态空间 $\phi_1 = \{1, 3, 7, 8\}$, $\phi_2 = \{2, 3, 4, 6, 8\}$, $\phi_3 = \{3, 5, 6, 7\}$, 则 3 个子空间构成的大系统状态空间 $\Phi = \bigcup_{k \in \phi_0} \phi_k = \{1, 2, 3, 4, 5, 6, 7, 8\}$ 且 $\phi_1 \cap \phi_2 = \{3, 8\}$, $\phi_1 \cap \phi_3 = \{3, 7\}$, $\phi_2 \cap \phi_3 = \{3, 6\}$, $\phi_1 \cap \phi_2 \cap \phi_3 = \{3\}$ 。当系统处于状态 $1 \in \phi_1$, 过程会按照 MPC_1 的无穷小矩阵 A_1 在状态 1 逗留一段时间, 这段时间服从指数分布, 然后跳转到状态 j^1 , 与此同时 $j^1 = j^l \in \phi_l \cap \phi_1, l=2$ or 3 , 过程将可能以行动 d_{1l} 切换到第 l 个子 Markov 过程运行。假设跳转到 $8 \in \phi_1 \cap \phi_2$, 则过程将可能以行动 d_{12} 切换到 MPC_2 进行状态演化; 同理跳转到 $7 \in \phi_1 \cap \phi_3$, 则过程将可能以行动 d_{13} 切换到 MPC_3 进行状态演化; 特别的跳转到 $3 \in \phi_1 \cap \phi_2 \cap \phi_3$, 则过程可能采取的切换行动 d_{12} 或 d_{13} 即切换到 MPC_2 或 MPC_3 都有可能。显然上述过程 $X_i \in \Phi = \{1, 2, 3, 4, 5, 6, 7, 8\}$, 其实是一个 Markov 过程。结果如下:

定理 1 令 $\theta_{ij}^{kl} = V_{d_{kl}}(e_n)$, 子状态空间有交集的 Markov 切换过程在策略 V 下相应的策略空间参数化为 $\Theta = \{\theta = (\theta_{ij}^{kl}, e_{ij} \in e_n \in E, d_{kl} \in D) | 0 \leq \theta_{ij}^{kl} \leq 1, \sum_{\phi_k \cap \phi_l \neq \emptyset} \theta_{ij}^{kl} = 1\}$,

Markov 切换控制过程 $\{X_i, D_i\}, X_i \in \Phi = \bigcup_{k \in \phi_0} \phi_k, D_i \in D$ 仍然是一个 Markov 过程, 且在策略 $v \in V$ 诱导下的无穷小矩阵为

$$A^v_{i^k j^l}(\theta) = \begin{cases} a_k(i, i) & e_{i^k j^k} \notin e_n, k=l \\ 0 & e_{i^k j^k} \notin e_n, k \neq l \\ a_k(i, j)\theta_{ij}^{kl} & e_{i^k j^k} \in e_n, e_{i^k j^l} \in e_n^{kl} \end{cases} \quad (3)$$

证明: 因为在子状态空间有交集的情况下, Markov 切换过程 $X_i \in \Phi = \bigcup_{k \in \phi_0} \phi_k$ 的每次状态改变都按照子 Markov 过程进行, 只是可能每步状态改变的时候依据不同的子 Markov 过程演化规律。这样, 系统下一步所处的状态只与当前所处状态有关, Markov 切换过程 X_i 实际上仍是一个时变的 Markov 过程。由无穷小矩阵和嵌入链的转移概率矩阵的关系: $A^v = \Lambda(P^v - I)$, 其中 I 是相应维数的单位矩阵, P^v 是在策略 v 诱导下的嵌入链的一步转移概率矩阵, Λ 是转移率所构成的对角矩阵, 经

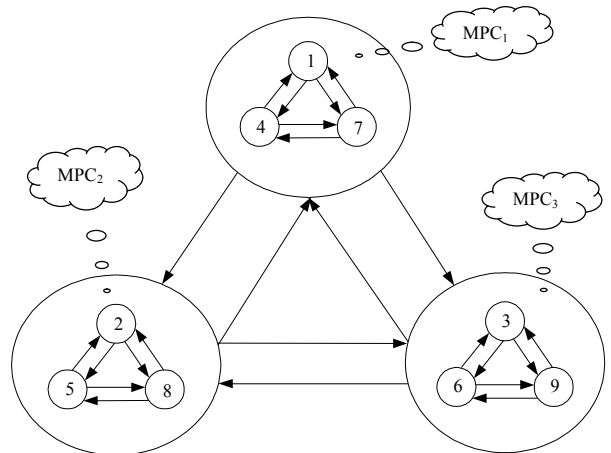


Fig.4 Switching system with three non-intersect subspaces
图 4 三个子空间无交集的 Markov 切换过程示意图

计算即可求出策略 $v \in V$ 诱导下的无穷小矩阵 $A_{i^k j^l}^v(\theta)$ 。

当子状态空间两两之间无交集, 即 $\forall k, l \in \{1, 2, \dots, N\}, \phi_k \cap \phi_l = \emptyset$ 。这样当系统处于状态 $i^k \in \phi_k$ 且事件 e_n 发生, 选择切换行动 d_{kl} , 则系统状态转移至 $j^l \in \phi_l$ 。如图 4 所示, $\text{MPC}_i: \{\phi_i, P_i, A_i\}, i=1, 2, 3$ 是 3 个子 Markov 过程且 $\phi_0 = \{1, 2, 3\}$, 其中子状态空间 $\phi_1 = \{1, 4, 7\}, \phi_2 = \{2, 5, 8\}, \phi_3 = \{3, 6, 9\}$, 则 3 个 Markov 空间构成的大系统状态空间 $\Phi = \bigcup_{k \in \phi_0} \phi_k = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ 。当系统处于状态 $1 \in \phi_1$ 时, 过程会按照 MPC_1 的演化规律进行状态跳变, 同时也可能以行动 d_{12} 或 d_{13} 切换到 MPC_2 或 MPC_3 的任何一个内部状态, 然后按照各自所处的子 Markov 过程的演化规律进行状态跳变。

定理 2 令 $\theta_{ij}^{kl} = V_{d_{kl}}(e_n)$, 子状态空间无交集的 Markov 空间切换过程在策略 V 下相应的策略空间参数化为: $\Theta = \{\theta = (\theta_{ij}^{kl}, e_{ij} \in e_n \in E, d_{kl} \in D) | 0 \leq \theta_{ij}^{kl} \leq 1, \sum \theta_{ij}^{kl} = 1\}$ 。Markov 切换控制过程 $\{X_t, D_t\}, X_t \in \Phi = \bigcup_{k \in \phi_0} \phi_k, D_t \in D$ 仍然是一个 Markov 过程, 并且在策略 $v \in V$ 诱导下, 嵌入链的一步转移概率矩阵满足关系

$$P_{i^k j^l}^v(\theta) = \begin{cases} p_k(i, j) & k = l, e_{i^k j^k} \in e_n, e_{i^k j^l} \notin e_n^{kl} \\ 0 & k \neq l, e_{i^k j^k} \notin e_n, e_{i^k j^l} \notin e_n^{kl} \\ [1 - p_k(i, i)]\theta_{ij}^{kl} & k \neq l, e_{i^k j^k} \in e_n, e_{i^k j^l} \in e_n^{kl} \end{cases} \quad (4)$$

证明 类似定理 1 的证明, 故略去。

2.2 性能灵敏度分析

综上所述, 本文将 N 个资源节点的分布式协同接入控制构建为一个 Markov 切换空间控制模型: $\{X_t, t \geq 0, (\text{MPC}_i, i=1, 2, \dots, N), D, f, \eta^v\}$ 。其中 $X_t, (\text{MPC}_i, i=1, 2, \dots, N), D$ 的定义如前所述; $f=(f_r, f_c)$, $f_r: \Phi \rightarrow \mathbf{R}$ 是性能报酬函数, $f_c: D \rightarrow \mathbf{R}$ 是切换行动的瞬时代价函数; η^v 是 Markov 切换系统的平均性能测度。

定义 平均性能测度如下:

$$\eta^v(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} E \left[\int_0^T (f_r(X_t) - f_c(D_t)) dt \right] = \sum_{i^k \in \Phi} \pi_{i^k}(\theta) f_r(i^k) - \sum_{i^k \in \Phi} \pi_{i^k}(\theta) \sum_{j^l \in \Phi} a_{i^k j^l}^v(\theta) f_c(d_{kl}) = \sum_{i^k \in \Phi} \pi_{i^k}(\theta) f(i^k, \theta) \quad (5)$$

式中 $f(i^k, \theta) = f_r(i^k) - \sum_{j^l \in \Phi} a_{i^k j^l}^v(\theta) f_c(d_{kl})$ 。状态 i^k 的性能势如下:

$$g_{i^k}^v(\theta) = E \left[\int_0^\infty (f_r(X_t) - f_c(D_t) - \eta^v(\theta)) dt | X_0 = i^k \right] \quad (6)$$

定义 状态 i^k 的平均逗留时间 $m_{i^k}^v(\theta) = \sum_{j^l \in \Phi} \int_0^\infty t p_{i^k j^l}^v(\theta) dt$, 且令 $\lambda_{i^k}^v(\theta) = \frac{1}{m_{i^k}^v(\theta)}$, $I(i^k, j^l) = \begin{cases} 1 & i^k = j^l \\ 0 & i^k \neq j^l \end{cases}$, 则有如下性能梯度公式:

$$\nabla \eta^v(\theta) = \pi(\theta) (\nabla A^v(\theta) g^v(\theta) + \nabla f(i^k, \theta)) \quad (7)$$

设 $\Phi_k = \{j^l \in \Phi, a_{i^k j^l}^v(\theta) > 0\}$, 考虑到 $a_{i^k i^k}^v(\theta) + \sum_{j^l \in \Phi, j^l \neq i^k} a_{i^k j^l}^v(\theta) = 0$, 并且注意到 $f_c(d_{kk}) = 0, k \in \phi_0$, 记 $\chi_{i^k j^l}(\theta) = \frac{\nabla a_{i^k j^l}^v(\theta)}{a_{i^k j^l}^v(\theta)}$, 则

上述性能梯度公式(7)变为如下形式:

$$\nabla \eta^v(\theta) = \sum_{i^k \in \Phi} \pi_{i^k}(\theta) \sum_{j^l \in \Phi} a_{i^k j^l}^v(\theta) (g_{j^l}^v(\theta) - f_c(d_{kl})) \chi_{i^k j^l}(\theta) = \sum_{i^k \in \Phi} \pi_{i^k}(\theta) \sum_{j^l \in \Phi_k} a_{i^k j^l}^v(\theta) (g_{j^l}^v(\theta) - g_{i^k}^v(\theta) - f_c(d_{kl})) \chi_{i^k j^l}(\theta) \quad (8)$$

式中 $\chi_{i^k j^l}(\theta) = \begin{cases} (0, \dots, 0, \frac{1}{\theta_n^{kl}}, 0, \dots, 0) & e_{ij} \in e_n^{kl} \\ 0 & e_{ij} \notin e_n^{kl} \end{cases}$ 。

设 $\{X_t, t \geq 0\}$ 是在策略参数 θ 驱动下, Markov 切换控制过程的一条样本轨道, $s^* \in \Phi$ 是一个正常返状态, t_m 是第 m 次抵达 s^* 的时间。如果 $\{X_t, t \geq 0\}$ 在每一时刻, $t_m, m=1, 2, \dots$, 在统计意义上以概率 1 等价, 则称 t_m 为再生时刻, s^* 为再生状态, $\{X_t, t \geq 0\}$ 为一个再生过程, $\{X_t, t_m \leq t < t_{m+1}\}$ 为第 m 个再生周期, 其中第 m 个再生周期的长度为 $T_m^n = t_m^{n+1} - t_m^n$ 。用 t_m^n 表示在此周期中第 n 次状态转移发生的时刻, 在每个状态转移时期 $[t_m^n, t_m^{n+1})$, 马尔可夫切换过程都是右连续的, 且这个时期的状态记为 $S_{t_m^n}^n$, 2 次状态转移的时间间隔为 $T_m^n = t_m^{n+1} - t_m^n$ 。最后用 n_m 表示第 m 个再生周期中发生的状态转移次数。根据上述定义, 对于固定的策略参数 $\theta, \{X_t, t \geq 0\}$ 在每一再生周期中独立

同分布, 即随机变量 T_m 独立同分布, 具有有限均值。

基于上述样本轨道 $\{X_i, t \geq 0\}$, 根据文献[15], 给出性能势的估计如下:

$$\begin{cases} \hat{g}_{S_{t_m}^n}^v(\theta) = \sum_{k=n}^{n_m} \left\{ (f_r(S_{t_m}^k) - \hat{\eta}^v(\theta)) T_m^k - f_c(d_{S_{t_m}^k, S_{t_m}^{k+1}}) \right\} \\ \hat{\eta}^v(\theta) = \frac{1}{T_m} \sum_{n=1}^{n_m} \left\{ f_r(S_{t_m}^n) T_m^n - f_c(d_{S_{t_m}^n, S_{t_m}^{n+1}}) \right\} \end{cases} \quad (9)$$

由式(8), 在第 m 个再生周期给出 $\nabla \eta^v(\theta)$ 的一个估计如下:

$$\begin{aligned} \hat{\nabla} \eta_m^v(\theta) &= \frac{1}{T_m} \sum_{n=1}^{n_m} \left(\hat{g}_{S_{t_m}^{n+1}}^v(\theta) - \hat{g}_{S_{t_m}^n}^v(\theta) - f_c(d_{S_{t_m}^n, S_{t_m}^{n+1}}) \right) \chi_{S_{t_m}^n, S_{t_m}^{n+1}}(\theta) = \\ &= \frac{1}{T_m} \sum_{n=1}^{n_m} \left(\left(\hat{\eta}^v(\theta) - f_r(S_{t_m}^n) \right) T_m^n - f_c(d_{S_{t_m}^n, S_{t_m}^{n+1}}) \right) \chi_{S_{t_m}^n, S_{t_m}^{n+1}}(\theta) \end{aligned} \quad (10)$$

基于这一样本轨道的第 m 个再生周期, 本文给出了性能测度 $\eta^v(\theta)$ 关于 θ 的梯度估计 $\hat{\nabla} \eta_m^v(\theta)$, 进而结合 RM(Robbins-Monro)随机逼近算法, 即可在下一个再生周期的开始时刻进行策略改进。本文采用 RM 算法:

$$\theta_{m+1} = \theta_m + \gamma_m \mathbf{H}_m \hat{\nabla} \eta_m^v(\theta_m) \quad (11)$$

其中 γ_m 是一个正的步长序列, 且满足 $\gamma_m > 0$, $\sum \gamma_m = \infty$, $\sum_{m=1}^{\infty} \gamma_m^2 < \infty$, $\mathbf{H}_m$ 为适当维数的正定矩阵, 选取 $\mathbf{H}_m$ 满足

$\|\mathbf{H}_m\| < \infty$, $\theta_m + \gamma_m \mathbf{H}_m \hat{\nabla} \eta_m^v(\theta_m) \in \Theta$ 。至此, 本文给出如下的基于策略梯度估计的 RM 随机逼近算法:

Step 1: (初始化)选择初始策略 θ_1 , 选取再生状态 s^* , 令 $m=1$, $n=1$, 初始状态 $S_m = S_0$, $t_m^n = 0$;

Step 2: 观测系统运行, 记录系统下一次状态转移发生时刻 t_m^{n+1} 和转移到的状态 $S_{t_m^{n+1}}^n$, 计算在状态 $S_{t_m^n}^n$ 的逗留时间 $T_m^n = t_m^{n+1} - t_m^n$;

Step 3: 若 $S_{t_m^{n+1}}^n = s^*$, 由式(9)和式(10), 计算 $\hat{\nabla} \eta_m^v(\theta)$; 否则, 令 $n=n+1$, 重复 Step 2;

Step 4: 取 $\gamma_m = 1/m$, 根据式(11)计算改进的策略 θ_{m+1} ;

Step 5: 更新策略, 置 $S_{t_{m+1}}^1 = s^*$, $t_{m+1}^1 = t_m^{n+1}$, 然后令 $n=1$, $m=m+1$, 转 Step 2。

推论 1 $E[\hat{\nabla} \eta_m^v(\theta)] = \nabla \eta^v(\theta)$, $\varepsilon = |\hat{\nabla} \eta_m^v(\theta) - \nabla \eta^v(\theta)| \leq c < \infty$, 即 $\hat{\nabla} \eta_m^v(\theta)$ 是 $\nabla \eta^v(\theta)$ 的一个误差有界的无偏估计。

证明 依据 $a_{i^k j^l}^v(\theta)$ 的有界连续且二次可导性质, 由式(8)可知 $|\nabla \eta^v(\theta)| \leq c_1 < \infty$, $|\hat{\nabla} \eta_m^v(\theta)| \leq c_2 < \infty$, 故有 $\varepsilon = |\hat{\nabla} \eta_m^v(\theta) - \nabla \eta^v(\theta)| \leq |\hat{\nabla} \eta_m^v(\theta)| + |\nabla \eta^v(\theta)| \leq c < \infty$ 。记 $F_t = \sigma(X_s, 0 \leq t \leq t')$, 则考虑如下条件期望:

$$E[\hat{\eta}^v(\theta) | F_m] = E\left[\frac{1}{T_m} \sum_{n=1}^{n_m} \left\{ f_r(S_{t_m}^n) T_m^n - f_c(d_{S_{t_m}^n, S_{t_m}^{n+1}}) \right\} \middle| F_m\right] = \eta^v(\theta) \quad (12)$$

$$E\left[\hat{g}_{S_{t_m}^n}^v(\theta) \middle| F_m\right] = E\left[\sum_{k=n}^{n_m} \left\{ (f_r(S_{t_m}^k) - \hat{\eta}^v(\theta)) T_m^k - f_c(d_{S_{t_m}^k, S_{t_m}^{k+1}}) \right\} \middle| F_m\right] = g_{S_{t_m}^n}^v(\theta) \quad (13)$$

由式(9)和式(10), 推出 $\hat{\nabla} \eta_m^v(\theta)$ 是 $\nabla \eta^v(\theta)$ 的一个误差有界的无偏估计。

进一步, 基于策略梯度的 RM 随机逼近算法产生的向量序列 $\{\theta_m, m \geq 0\}$ 收敛到最优策略 θ^* 。即:

$\theta^* = \arg \max_{\theta \in \Theta} \eta(\theta)$, 给定初始策略 $\theta_0 \in \Theta$, 由算法产生的向量序列 $\{\theta_m, m \geq 0\}$ 满足 $\lim_{m \rightarrow \infty} \theta_m = \theta^*$ 。

3 仿真实例

本文考虑一个接入控制问题的实际算例, 该系统中媒体服务器数量为 6, 每台服务器的可用带宽为 2, 单位时间的能耗即切换代价也为 2, 而且假定系统中存在 3 种视频, 每种视频的带宽消耗分别为 $b_1=1$, $b_2=2$, $b_3=2$, 系统接受每种视频的一个请求可以获得的收益或报酬分别为 $r(1)=8$, $r(2)=20$, $r(3)=9$ 。访问每种视频的业务达到率分别服从独立的泊松分布, 而且达到率分别为 $\lambda_1=3.0$, $\lambda_2=2.0$, $\lambda_3=2.5$ 。每种视频的持续访问时间服从参数为

$\frac{1}{l_n/\alpha}$ 的指数分布, 并且 $l_1/\alpha=0.5$, $l_2/\alpha=0.8$, $l_3/\alpha=0.9$, 如表 1 所示。

给定上述参数后,利用上述 RM 随机逼近算法来求解切换系统的最优策略。另外,为了说明本文提出策略的有效性,使用 2 种策略,即分别称为策略 1(policy 1)和策略 2(policy 2)与本文所给策略进行比较。其中策略 1 将所有的服务器全部开启,当系统带宽足够时接受所有的视频请求;而策略 2 将所有的服务器开启,但是根据系统状态和收益来接受请求,即接受收益更大的视频种类。所得长期平均性能的比较如表 2 所示。

表 1 切换系统的基本参数

Table 1 Basic parameters of switching system				
video types	bandwidth	arrival rate	leave rate	reward
1	1	3.0	0.5	8
2	2	2.0	0.8	20
3	2	2.5	0.9	9

表 2 切换系统在不同策略下的长期平均性能比较

Table 2 Long-term average performance comparison of switching system with different policies	
policy	long term average performance
policy 1	4.19
policy 2	4.35
policy in this paper	4.52

可以看出,本文给出的策略所得到的长期平均性能是最大的,即在如此的分布式系统中利用本文提出的 RM 随机逼近算法得到的策略在某种意义上是最优的,从而验证了本文的正确性。

4 结论

本文研究了分布式系统协同接入控制的建模和优化,以分布式视频服务为主要研究对象,将一个资源节点上的业务接入状态演化用 Markov 过程来描述,并且证明了所有的 N 个资源节点的分布式协同接入控制可以由一个 Markov 切换空间控制过程来描述。针对上述构建的 Markov 切换系统,基于系统灵敏度的观点对系统性能进行优化,提出了基于策略梯度的随机逼近算法并给出了算法的收敛性。最后通过仿真实例验证了本文策略的正确性。

参考文献:

- [1] 东庭,孙学康. 视频点播(VOD)及其在 Internet 上的实现[J]. 计算机系统应用, 2006(1):21-26. (DONG Ting,SUN Xuekang. Video on demand(VOD) and its realization on the Internet[J]. Computer Systems and Applications, 2006(1):21-26.)
- [2] 沈建斌,耿兆丰,邹轩. 视频点播系统的设计与实现[J]. 计算机工程, 2006,32(2):209-211. (SHEN Jianbin,GENG Zhaofeng,ZOU Xuan. Video on demand system design and realization[J]. Computer Engineering, 2006,32(2):209-211.)
- [3] Gemmell D J,Vin H M. Multimedia storage servers: a tutorial[J]. IEEE Multimedia, 1995:40-49.
- [4] Debasish G,Hyoung J K. Scheduling video streams in video-on-demand systems:a survey[J]. Multimedia Tools and Applications, 2000:167-195.
- [5] Cleary K. Video on demand—competing technologies and services[J]. International Broadcasting Convention, 1995:432-437.
- [6] Barnett S A,Anido G J. A cost comparison of distributed and centralized approaches to video-on-demand[J]. IEEE Journal on Selected Areas in Communications, 1996,14(6):1173-1183.
- [7] Zhang L Y,Zhang X. The research of VoD system based on CDN and P2P technologies[C] //The 2nd International Conference on Computer and Automation Engineering. Singapore:[s.n.], 2010:385-388.
- [8] 郑维玮,马卫东,刘作臣,等. P2P 流快速识别技术[J]. 太赫兹科学与电子信息学报, 2011,9(1):117-120. (ZHENG Weiwei, MA Weidong,LIU Zuochen,et al. Technology of rapid P2P stream identification[J]. Journal of Terahertz Science and Electronic Information Technology, 2011,9(1):117-120.)
- [9] Gaeta R,Gribaudo M,Manini D,et al. Analysis of resource transfers in peer-to-peer file sharing applications Using fluid models[J]. Performance Evaluation, 2006,63(3):149-174.
- [10] 蒋瑜,刘嘉勇,李波. 基于信誉的 P2P 网络信任模型研究[J]. 太赫兹科学与电子信息学报, 2007,5(6):452-456. (JIANG Yu,LIU Jiayong,LI Bo. Recommendation-based Trust Model in P2P Network[J]. Journal of Terahertz Science and Electronic Information Technology, 2007,5(6):452-456.)
- [11] Conklin G J,Greenbaum G S,Lillevoild K O,et al. Video coding for streaming media delivery on the Internet[J]. IEEE Transactions on Circuits and Systems, 2001,11(3):269-280.
- [12] Jin B K,Heon Y Y. An admission control scheme for continuous media servers using catching[C] //Proceedings of 19th IEEE International Performance,Computing and Communications Conference. Phoenix,USA:[s.n.], 2000:456-462.

(下转第 979 页)