

# 线上教育资源的缓存机制研究

吴伟华<sup>1</sup> 户 研<sup>2</sup> 刘润滋<sup>3</sup>

1. 西安电子科技大学·通信工程学院 陕西西安 710071

2. 河南开放大学·信息工程学院 河南郑州 450046

3. 西安建筑科技大学·信息与控制工程学院 陕西西安 710311

**摘要:** 本文提出了一个基于情景感知的代理缓存分配方案来高效地向代理缓存分发教学视频资源。设计了一个层次分析法来确定代理向学生传输教学视频的开销权重。随后建立了一个多播补丁传输方案来整合代理缓存和多播传输策略。紧接着将代理缓存分配描述成了一个关于传输收益和启动时延收益的最大化问题。最后提出了一个迭代算法来得到问题的最优解。仿真实验说明, 与传统策略相比所提方案具有小的启动时延和低的传输开销。

**关键词:** 教学视频分发; 情景感知; 内容分发网络

## Research on caching mechanism of online education resources

Weihua Wu<sup>1</sup>, Yan Hu<sup>2</sup>, Runzi Liu<sup>3</sup>

1. School of Communication Engineering, Xidian University, Xi 'an 710071, China

2. School of Information Engineering, Henan Open University, Zhengzhou 450046, China

3. School of Information and Control Engineering, Xi 'an University of Architecture and Technology, Xi 'an 710311, China

**Abstract:** This paper proposes a proxy cache allocation scheme based on situational awareness to efficiently distribute teaching video resources to proxy cache. An analytic hierarchy process (AHP) method is designed to determine the cost weight of teaching videos transmitted by agents to students. Then a multicast patch transport scheme is built to integrate proxy caching and multicast transport policies. Then the proxy cache allocation is described as a problem of maximizing the benefits of transmission and startup delay. Finally, an iterative algorithm is proposed to obtain the optimal solution of the problem. Simulation results show that the proposed scheme has smaller startup delay and lower transmission cost compared with the traditional strategy.

**Keywords:** Teaching video distribution; Situational perception; Content delivery network

### 作者简介:

1. 吴伟华、西安电子科技大学(通信工程学院)、邮编: 710071、男、河北省石家庄市、博士研究生、讲师、汉、1988.4.11、专业: 为无线资源分配和人工智能、基金: 国家自然科学基金-青年项目, 多时间尺度的未授权多制式无线网络资源管理(61801365)

2. 户研、河南开放大学(信息工程学院)、邮编: 450046、男、河南开封、硕士、讲师、汉、1989.1.28、专业: 智能天线, 图像处理、基金: 国家自然科学基金-青年项目, 多时间尺度的未授权多制式无线网络资源管理(61801365)

3. 刘润滋、西安建筑科技大学(信息与控制工程学院)、邮编: 710311、女、山东省潍坊市、博士研究生、副教授、汉、1988.9.19、专业: 无线网络、空间网络中的资源管理和性能分析、基金: 国家自然科学基金-青年项目, 多时间尺度的未授权多制式无线网络资源管理(61801365)

## 一、引言

由于教学视频的高清晰度和长播放时间的特性，目前的网络带宽难以支撑对其进行流畅的播放。内容分发网络（content distribution network, CDN）由源服务器和多个代理缓存组成，可以为线上的教学视频提供带宽高效的内容分发<sup>[1]</sup>。

现有CDN方案必须遍历从服务器到客户端的整个端到端路径，这可能会导致服务器过载，并导致较高的传输成本<sup>[2-4]</sup>。通过利用情景感知信息，可以在网络中有效地利用即时可用资源，以提高用户感知的服务质量和网络利用率<sup>[5, 6]</sup>。

本文提出了一个情景感知的代理缓存分配方案。为此，首先提出了一个代理缓存的权重分配方案，用于为正在听课的学生选择合适的代理缓存。随后，设计了一个基于补丁的多播传输方案来为代理与学生端之间的数据传输提供支持。最后，建立了一个代理缓存分配算法来确定每个缓存上需要存储的教学视频资源。

## 二、系统模型

CDN系统由一个源服务器， $N$ 个代理缓存和 $M$ 个正在向CDN网络请求教学视频资源学生班组成。假设源服务器为所有的学生班存储着 $M$ 个教学视频文件。使用 $S_n$ 来表示代理缓存 $n$ 的容量， $D_{un}$ 表示在不使用代理和多播传输方案的情况下传输的视频的启动时延。教学视频 $n$ 长度为 $L_n$ （秒）和播放带宽为 $b_n$ （bps）。用 $p$ 表示传输一个比特视频数据的成本。用 $v_{m,n}$ 表示代理 $n$ 缓存视频 $m$ 的 $v_{m,n}$ 秒前缀，则缓存矩阵 $\mathbf{v}=(v_{m,n})_{M \times N}$ 表示了整个CDN网络的缓存策略。缓存在代理 $n$ 中视频 $m$ 的传输成本 $h_{m,n}(v_{m,n})$ 包含从代理到学生的视频前缀的传输成本，以及由代理进行转发的视频后缀传输成本。假设不使用代理和多播传输方案的传输成本可以表示为 $h_{m,n}(0)$ 。则缓存视频前缀 $v_{m,n}$ 的传输收益为 $RC_{m,n}(v_{m,n})=h_{m,n}(0)-h_{m,n}(v_{m,n})$ 。

学生从代理 $n$ 接收到视频 $m$ 并进行播放的启动延迟由 $D_{m,n}$ 表示，其可以计算为 $D_{m,n}=D_{un}(L_{th}-v_{m,n})$ 。假设整个班级的同学对视频的请求过程为泊松到达且到达率为 $\lambda_m$ 。则节省的传输时延为 $RD_{m,n}(v_{m,n})=(\lambda_m v_{m,n})D_{un}$ 。同时考虑传输收益 $RC_{m,n}(v_{m,n})$ 和时延收益 $RD_{m,n}(v_{m,n})$ ，则代理 $n$ 缓存视频 $m$ 前缀的收益为 $U_{m,n}(v_{m,n})=\omega_c RC_{m,n}(v_{m,n})+\omega_d RD_{m,n}(v_{m,n})$ ，其中 $\omega_c$ 和 $\omega_d$ 分别是传输收益和时延收益的权重。当同时考虑 $M$ 个视频和 $N$ 个代理时，可以得到如下优化问题：

$$\max \sum_{m=1}^M \sum_{n=1}^N U_{m,n}(v_{m,n}) \quad (1)$$

$$s.t. \sum_{m=1}^M v_{m,n} \leq S_n, \forall n, \quad (2)$$

$$0 \leq v_{m,n} \leq L_m, \forall m, n \quad (3)$$

问题（1）中传输收益的具体表达式并不清楚，这就给确定最优的缓存策略带来了困难。为了解决问题（1），下文将借助层次分析法和多播补丁传输方案来明确传输成本。

## 三、权重代理传输方案

在对前缀进行分配之前，首先要确定学生选择不同代理缓存的权重。得到权重因子后，需要给出代理向学生进行视频传输方的具体方案。

### 1. 班级的代理权重计算

参考层次分析法（Analytic Hierarchy Process, AHP, AHP）的体系结构，将确定代理的权重作为AHP的目标层，将学生关心的网络性能作为准则层，学生能够选择的代理作为方案层。下面来具体介绍如何确定出学生选择不同网络代理的权重。

#### ① 学生对网络性能的忍受权重计算

需要确定出学生对不同网络性能的忍受权重。本文考虑以下六个的网络性能：1）代理的接口；2）最小化传输开销；3）最大化平均吞吐量；4）最小化时延；5）最小化抖动；6）最小化误帧率。

学生按照自身需求从低到高对这六个网络性能进行排序。根据学生的排序，系统会自动将1到6之间的优先级分数分配给不同的网络性能，其中6表示最在意的网络性能，1表示最不在意的性能。对于班级 $k$ ，任意两个网络性能之间的相对分数可以计算为 $A_{a,b}^k = u_a^k / u_b^k$ ，其中 $u_a^k$ 和 $u_b^k$ 是班级 $k$ 在性能 $a$ 和 $b$ 上的分数。矩阵 $\mathbf{A}_1^k = [u_a^k / u_b^k]_{6 \times 6}$ 是目标的成对比较矩阵。最后，班级 $k$ 对每个网络性能指标的权重由 $\mathbf{c}_1^k$ 表示，其是对应于最大特征值的特征向量。

#### ② 针对网络性能的代理权重计算

第二步是针对某个特定的网络性能，对代理进行评价。下面以传输开销为例来介绍一下代理权重的计算过程。服务器按照传输开销的大小对代理进行降序排列。随后，按照步骤1中所述的类似方式计算每个代理对网络性能的权重。与网络性能 $i$ 相对应的每个代理的权重由最大特征值对应的特征向量 $\mathbf{c}_{2,i}^k$ 表示。在计算出所有的网络性能所对应的特征向量之后，可以得到权重矩阵 $\mathbf{c}_2^k = [\mathbf{c}_{2,1}^k, \mathbf{c}_{2,2}^k, \dots, \mathbf{c}_{2,6}^k]$ 。

在得到  $\mathbf{c}_1^k$  和  $\mathbf{c}_2^k$  之后，可以计算出班级  $k$  选择代理的权重矩阵  $\mathbf{c}^k = \mathbf{c}_1^k \cdot \mathbf{c}_2^k$ 。随后可以得到整个网络的权重矩阵  $\mathbf{c} = [\mathbf{c}^1, \mathbf{c}^2, \dots, \mathbf{c}^M]$ ，其中  $c_{k,n}$  表示班级  $k$  选择代理  $n$  的权重。

### ③服务器的代理选择权重计算

当服务器和代理进行通信时，它们之间的距离是唯一需要考虑的因素。通常情况下，服务器和代理之间的距离越小越好。因此可以使用公式 (6) 来计算服务器选择不同代理的分数。可以得到成对矩阵  $\mathbf{A}$ ，矩阵  $\mathbf{A}$  的最大特征值所对应的特征向量  $\mathbf{w}$  就是服务器选择不同代理的权重。

根据得到的班级代理权重  $\mathbf{c}$  和服务器选择代理的权重  $\mathbf{w}$ ，可以得到代理  $n$  到学生班级  $k$  和服务器到代理  $n$  和的传输开销  $p^{-c_{k,n}}$  和  $p^{-w_n}$ 。

### 2. 基于前缀缓存的多播补丁传输策略

假设代理  $n$  上面的教学视频  $m$  的第一个请求在时刻 0 到达。在时间  $v_{m,n}$ ，教学视频  $m$  的前缀被传输完毕，代理开始向服务器请求视频的后缀，并将其传输给学生。对视频  $m$  后到来的请求可以开始一个新的多播，或者加入正在进行的多播流并且另开辟一个单播来从代理获取之前的前缀数据。令  $G_{m,n}$  表示代理  $n$  要不要为视频  $m$  开始一个多播流的门限。如果第二个请求是在  $G_{m,n}$  之前到达的，则第二个请求加入正在进行的多播流；反之，代理需要为其开辟一个新的多播。如果一个请求在多播已经开始后的  $t_1$  时刻到达，视频的传输可以根据  $v_{m,n}$  和  $G_{m,n}$  之间的关系分为以下两种模式<sup>[9]</sup>。

模式 1: 当  $G_{m,n} < v_{m,n} < L_m$  时，传输开销函数表示为  $g_1(v_{m,n}, G_{m,n})$ 。

模式 2: 当  $0 < v_{m,n} < G_{m,n}$  时，传输开销函数表示为  $g_2(v_{m,n}, G_{m,n})$ 。

$g_1$  和  $g_2$  的详细表达式可参考文献[9]。令  $f_z(v_{m,n})$  表示模式  $z$  下的最小传输开销， $z=1, 2$  则  $f_z(v_{m,n}) = \min_{G_{m,n}} \{g_z(v_{m,n}, G_{m,n})\}$ 。则对于给定的前缀  $v_{m,n}$  来说，可能的最大传输开销为  $h_{m,n}(v_{m,n}) = \max \{f_1(v_{m,n}), f_2(v_{m,n})\}$ 。当代理上面没有视频缓存，并且传输方式采用单播的形式，则传输开销为  $h_{m,n}(0) = (p^{-w_n} + p^{-c_{k,n}})L_m \lambda_m b_m$ 。最终，不难证明  $RC_{m,n}(v_{m,n})$  是关于  $v_{m,n}$  的凹函数，因此优化问题 (1) 的目标函数  $U_{m,n}(v_{m,n})$  也是一个凹函数。最终可以确定优化问题 (1) 是一个凹优化问题，因此可以通过以下迭代算法来找到问题的最优值

算法 1: 如下迭代策略能够找到问题的最优解

$$\begin{aligned} v_{m,n}(t+1) &= \left[ v_{m,n}(t) + \alpha \left( \dot{U}_{v_{m,n}} - \gamma_n \right) \right]_{v_{m,n}}^{[0, L_m]} \\ \gamma_n(t+1) &= \left[ \gamma_n(t) + \beta \left( \sum_m v_{m,n} - S_n \right) \right]_{\gamma_n}^{[0, +\infty)} \end{aligned} \quad (4)$$

其中  $t$  是迭代索引， $\dot{U}_{v_{m,n}}$  是函数  $U_{m,n}(v_{m,n})$  对  $v_{m,n}$  的偏导数， $\alpha$  和  $\beta$  是充分小的步长因子。

### 四、仿真分析

在仿真中假设有 5 个班级来向服务器请求 5 个对应的教学视频。代理缓存的数目为  $N=4$ ，并且它们的容量为  $S=[29, 23, 27, 25]$ 。算法的步长参数设置为  $\alpha=2.5$ ， $\beta=0.005$ 。传输收益和时延收益的权重设置为 0.6 和 0.4。图 1 展示了 5 个视频的启动时延。图 1 中所有视频的参数相同，从而验证网络情景信息对缓存算法的影响。

从图中可以观察到，缓存分配算法能够在 120 次迭代之后收敛。在这个试验中，假设不使用前缀缓存方案的策略的启动时延为 10 秒。因此，本文所提出的算法能够非常有效地降低学生观看视频的启动时延，这对改善学生的观看体验和提升线上教学的教学质量来说是非常重要的。

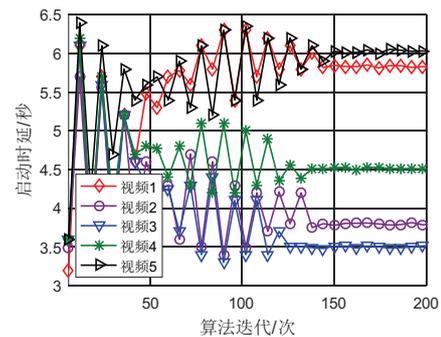


图 1 不同视频的启动时延

图 2 介绍了内容情景信息对视频启动时延的影响，其所有网络的性能参数相同，从而验证不同视频参数对缓存算法的影响。从图中可以看出，所有的视频都具有较小的启动时延。

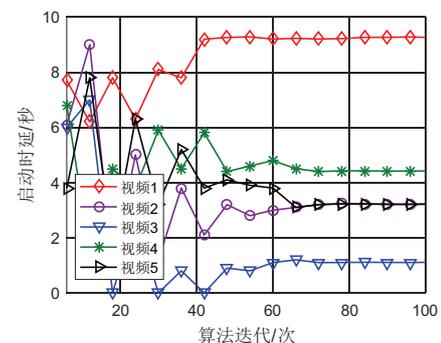


图 2 不同视频的启动时延

## 五、结论

本文提出了一个情景感知的代理缓存分配方案来应对线上教学视频分发。利用层次分析法确定了代理分发视频的权重，并提出了一个多播补丁传输方案来整合代理缓存和多播传输策略。随后提出了一个分布式缓存分配算法来最大化传输和启动时延的收益。最后用李雅普诺夫稳定性策略证明的算法的稳定性。

### 参考文献:

[1]李乔, 何慧, 张宏莉.内容分发网络研究[J].电子学报, 2013, 41 (8): 9..

[2]ZHANG H , CHEN M , PAREKH A , et al. An Adaptive Multi-channel P2P Video-on-Demand System using Plug-and-Play Helpers[J]. Computer Science, 2010.

[3]ALMEIDA J M, EAGER D L, FERRIS M, et al. Provisioning Content Distribution Networks for Streaming

Media[C]// INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, 2002.

[4]BING W , SEN S , ADLER M , et al. Proxy-based Distribution of Streaming Video over Unicast/Multicast Connections[J]. Proc IEEE Infocom, 2001.

[5]ZHAO X, ANMA F, NINOMIYA T, et al. Personalized Adaptive Content System for Context-Aware Mobile Learning[J]. international journal of computer science & network security, 2008.

[6]SIMSEK M, BENNIS M, I GÜVENC. Context-Aware Mobility Management in HetNets: A Reinforcement Learning Approach[C]// 2015 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2015.