

云计算环境下的移动视频传输速率控制机制

邹露宁, 白光伟, 沈 航, 纪传伟

(南京工业大学 计算机科学与技术系, 江苏 南京 211816)

摘 要: 提出云计算环境下的移动视频传输速率控制机制, 利用云强大的计算和存储能力, 扩展手机性能. 首先通过隐马尔科夫链预测无线带宽, 利用屏幕分辨率和码率的关系, 自适应地调节视频传输速率. 在此基础上, 设计缓存管理算法, 进一步提高用户视频访问命中率. 仿真实验表明, 该机制可以显著提高用户体验和网络带宽的利用率.

关键词: 移动视频; 云计算; 带宽预测; 隐马尔科夫链; 速率控制

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2016)04—0102—06

Transmission Rate Control Mechanism for Mobile Video in Cloud Computing Environment

ZOU Lu-ning, BAI Guang-wei, SHEN Hang, JI Chuan-wei

(Department of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China)

Abstract: This paper presents a transmission rate control mechanism in cloud computing environment for mobile video. The combination of mobile and cloud computing, which has a powerful computing and storage capacity, improves the performance of the mobile device. First, the proposed mechanism algorithm can forecast bandwidth by using hidden markov chain. Then using the relationship between hardware and bit rate, the mechanism adjusts bit rate adaptively. On this basis, we design a buffer management algorithm to improve accessing hit rate. Our simulation results demonstrate that the proposed mechanism achieves significant improvement in terms of QoE and network resource utilization.

Key words: mobile video; cloud computing; bandwidth prediction; hidden markov chain; rate control

1 引言

移动云计算把无线通信技术和云计算结合, 在提供大量存储空间的同时, 也提供大量的计算资源, 把手机计算量大的任务迁移到远端云, 提高手机性能, 节省手机能量, 为解决移动多媒体业务存在的问题提供了可能.

现有的自适应视频传输机制无法准确预测无线带宽, 没有考虑移动设备的硬件性能, 导致无线通信资源的浪费, 降低网络服务能力和用户体验.

文献[1]提出基于 QoE 的速率选择机制. 文献

[2]采用私人代理和可扩展的网络编码技术, 基于反馈的链路质量, 调整视频传输速率. 文献[3]考虑手机设备的格式和屏幕分辨率. 云转码器根据用户的请求参数, 通过转码技术, 把符合硬件参数的视频码率发送给用户; 文献同时提出视频请求预测和缓存管理策略, 但并未给出码率和屏幕分辨率的具体关系.

本文提出自适应的移动视频速率控制机制, 能够准确预测网络带宽, 考虑用户的硬件设备, 动态调整视频传输速率, 利用缓存管理策略, 提高用户访问命中率.

收稿日期: 2015—07—15; **修回日期:** 2015—08—10

基金项目: 国家自然科学基金(60673185, 61073197); 江苏省自然科学基金(BK2010548); 江苏省科技支撑计划(工业)(BE2011186)

2 系统模型

如图 1 所示,提出的系统模型包括 3 个模块:多媒体云数据中心、移动终端以及多媒体雾. 数据中心存储大量原始视频,同时将视频转码并传送给用户. 移动终端作为视频请求和接收方,除了上传手机硬件信息,也要反馈视频观看情况. 多媒体雾的目的是提高视频请求命中率,它预测无线带宽,分析环境,缓存流行度高的视频.

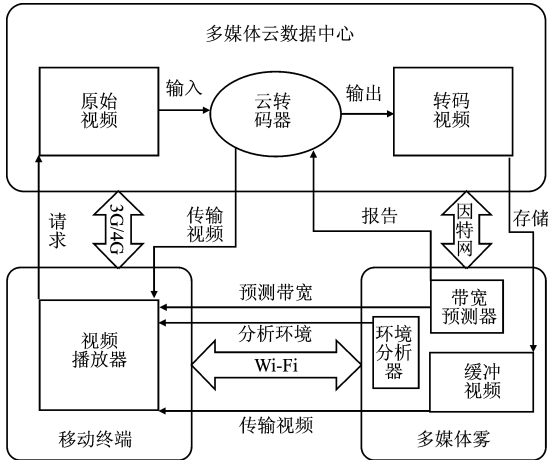


图 1 移动视频传输速率系统模型

2.1 多媒体云数据中心

图 2 是自适应视频流机制流程图. 计算和存储是多媒体云数据中心的主要任务,在其中存储着大量原始视频. 用户发出观看视频请求时,上传硬件信息(包括屏幕分辨率,CPU 使用情况和内存大小),云数据中心计算初始码率. 如果请求视频不在本地缓存中,或者在本地缓存中,但码率与用户硬件需求相差太大,多媒体云数据中心会根据预测带宽,动态调整码率,传输给用户. 同时,数据中心把转码之后的视频存储在临时缓存中,并把流行度高以及转码格式次数最多的视频,存储在本地雾中.

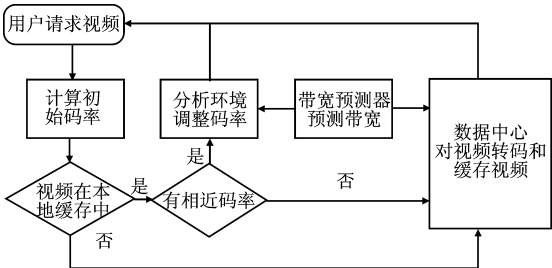


图 2 速率选择机制流程图

2.2 移动终端

用户在发出视频请求的同时,上传手机硬件信

息,并接收从数据中心或者多媒体雾发送的视频. 在播放视频时,移动终端上传视频放映情况,供多媒体雾分析.

2.3 多媒体雾

多媒体雾是远端云在本地的一个镜像^[4]. 相对于云,雾的计算和存储能力都较弱,但可以更快速地给手机用户提供服务. 一方面,雾可以预测带宽,并把预测得到的数据传送给远端云,以便远端云实时调整速率. 另一方面,多媒体雾中缓存很多流行度高的视频,当用户请求的视频在本地雾中并且码率相近时,可以从本地雾中直接发送给用户,减少传播时延,降低网络负担. 最后,雾中的环境分析器可以分析网络环境. 当收到用户的反馈信息时,分析器结合预测带宽,对速率做轻微调整,提高用户的 QoE.

3 移动视频传输速率控制机制

准确的带宽预测是自适应传输技术的基础. 由于用户的移动设备千差万别,在屏幕分辨率、内存和 CPU 等性能上存在差异,在预测网络带宽后,还要考虑用户的硬件条件,使传输速率满足用户的设备需求. 为提高移动视频自适应机制的命中效率,减少请求响应时间,缓解网络负载状况,本文提出缓存管理算法. 为了便于描述机制,将本文涉及的部分符号标记归纳为表 1.

表 1 本文使用的标记符号

参数符号	参数意义
U	用户的集合
V	视频集合
B_t	t 时刻的带宽
B'_t	t 时刻预测带宽
$M(t)$	t 时刻向用户传输的码率
$P(V_i)$	视频 V_i 被观看次数
$U(t)$	t 时刻有多少用户观看视频
P	移动设备分辨率

3.1 带宽预测器

带宽具有时变性,且带宽状态是不能直接观察到的. 如果向网络中发送数据包来预测带宽,会增加网络负担. 但是网络带宽与延迟、丢包和抖动等因素密切相关,并且这些因素是可以直接观测得到的,故本地雾端的带宽预测器利用隐马尔科夫模型来预测带宽,其中延迟、丢包和抖动等任何一个都可作为观

测对象.

假设 1:在较短的时间片 $[T_i, T_{i+1}]$ 内,无线信道带宽 S_i 处于相对稳定的状态.

假设 2:无线信道带宽为慢衰弱,带宽只在相邻状态转变.

隐马尔科夫模型(HMM)是一种双重随机过程,其一为马尔科夫链,即带宽状态序列;其二为观察序列,可以被直接观测到,与需要的状态序列关联很大.如表 2 所示, $\lambda=(N, M, \pi, A, B)$ 为有限带宽状态的隐马尔科夫模型参数.

表 2 隐马尔科夫模型参数表

参数	意义
N	带宽状态数目
M	每个带宽状态可能的观察值数目
A	带宽状态之间的转移概率矩阵
B	每个带宽状态下所有观测值对应的概率矩阵
π	每个带宽状态的概率分布量

HMM 将带宽的变化范围划分为 5 状态,且状态集合 $S=\{S_1, S_2, S_3, S_4, S_5\}$.那么 n 时刻信道 X_n 的概率分布量 $\pi(n)$ 为:

$$\pi_i(n)(X_n = S_n) \tag{1}$$

定义 1 $\forall S_i, S_j \in S$,称 $p_{ij} = P(X_{t+1} = S_j | X_t = S_i)$ 为 t 时刻的一步转移概率.记 $A = (p_{ij})$ 为 $\{X_t, t \geq 0\}$ 的一步转移概率矩阵,其中 $1 \leq i, j \leq 5$.

由于带宽只会保持不变或者向相邻的状态改变,所以 $p_{ij} = P(X_{t+1} = S_j | X_t = S_i) = 0, |i - j| \geq 2$,那么矩阵 A 可化为:

$$A = \begin{bmatrix} p_{11} & p_{12} & 0 & 0 & 0 \\ p_{21} & p_{22} & p_{23} & 0 & 0 \\ 0 & p_{32} & p_{33} & p_{34} & 0 \\ 0 & 0 & p_{43} & p_{44} & p_{45} \\ 0 & 0 & 0 & p_{54} & p_{55} \end{bmatrix} \tag{2}$$

状态图如图 3 所示.

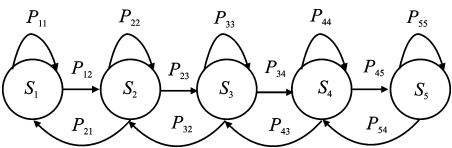


图 3 状态之间的转换图

设观测值状态集合 $M = \{m_1, m_2, m_3, m_4, m_5\}$,可以通过在 t 时刻的观察值 ξ_t 来估计带宽的状

态,即每一个 ξ_t 值,都对应一个状态 X_t ,如图 4 所示,其中 $\xi_t \in M$.

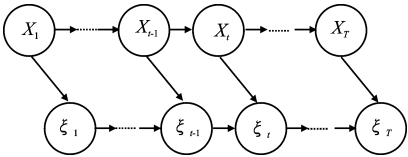


图 4 状态和观测值的映射关系

定义 2 $b_{ij} = P\{M_t = m_i | X_t = S_j\}, 1 \leq i, j \leq 5$,且 $\sum m_{j=1} P\{X_t = S_j | \xi_t = m_i\} = 1$,这样,就能得到观察值与状态的关系矩阵 B :

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{15} \\ b_{21} & b_{22} & \cdots & b_{25} \\ \vdots & \vdots & \vdots & \vdots \\ b_{51} & b_{52} & \cdots & b_{55} \end{bmatrix} \tag{3}$$

在得到模型 $\lambda=(N, M, \pi, A, B)$ 后,需要对参数进行优化:给定 T 时间内的观察序列 M ,估算出最有可能产生此序列的模型 λ .对于解决隐马尔科夫模型的参数选择和优化问题,使用最广泛的是 Baum-Welch 算法.该算法是一种迭代公式,初始时刻由用户给出各参数的经验估计值,通过不断迭代,使各个参数逐渐趋向更为合理的最优值.

定义一些函数: $\gamma_t = \gamma_1(i)$ 是 $t=1$ 信道时处于 S_i 的期望值, $\zeta_t(i, j)$ 表示 t 时状态为 S_i 以及 $t+1$ 时状态为 S_j 的概率:

$$\begin{aligned} \zeta_t(i, j) &= \frac{P(X_t = S_i, X_{t+1} = S_j, M | \lambda)}{P(M | \lambda)} \\ &= \frac{\alpha_t(i) p_{ij} b_j(m_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^n \sum_{j=1}^n \alpha_t(i) p_{ij} b_j(m_{t+1}) \beta_{t+1}(j)} \end{aligned} \tag{4}$$

$$\gamma_t(i) \sum_{j=1}^N \zeta_t(i, j) \tag{5}$$

式中, t 时刻处于状态 S_i 的概率, $\sum_{t=1}^{T-1} \gamma_t(i)$ 是整个过程中从状态 S_i 转出次数的预期, $\sum_{t=1}^{T-1} \zeta_t(i, j)$ 是从 S_i 跳转到 S_j 次数的期望,终止条件为式(6),其中 ϵ 是预先定义的阈值,这样得到概率转移矩阵 A (式 7) 和混合矩阵 B (式 8):

$$| \log P(M | \lambda) - \log P(M | \lambda_0) | < \epsilon \tag{6}$$

$$\tilde{p}_{ij} = \sum_{t=1}^{T-1} \zeta_t(i, j) / \sum_{t=1}^{T-1} \gamma_t(i) \tag{7}$$

$$\tilde{b}_j(k) = \sum_T \gamma_t(j) / \sum_{t=1}^T \gamma_t(j) \tag{8}$$

对于一个给定的观察序列 M 以及更新的 λ ,下一个问题是怎样选取一个状态序列 S ,使得 S 最有

可能产生这个观察序列. 这个问题就是解码问题, 采用 Viterbi 算法, 用动态规划解决. 也即对于给定的观察序列 $M = M_1 M_2 \cdots M_T$ 和模型 $\lambda = (N, M, \pi, A, B)$, 求最优相关状态序列 $X_1 X_2 \cdots X_{t-1}$.

$$\delta_t(i) = \max_{X_1 X_2 \cdots X_{t-1}} P[X_1 X_2 \cdots X_{t-1}, X_t = S_i, M_1 M_2 \cdots M_t | \lambda] \quad (9)$$

表示 T 时刻最大的 $\delta_T(i)$ 所代表的状态.

当 $t = 1$ 时, 对于某个 $1 \leq i \leq N$, 此时只有一个状态, 一个观测值, 因此最大值为其本身, 即:

$$\delta_1(i) = \pi_i b_i(M_1) \quad (10)$$

此时 S_i 为初始状态, 没有状态转移而来, 记做:

$$\varphi_1(i) = 0 \quad (11)$$

然后进行递归运算, 当 $2 \leq t \leq T, 1 \leq j \leq N$ 时, $\delta_t(i)$ 是在 $\delta_{t-1}(i)$ 的基础上, 向右移动一步产生, 每移动一步, 都有 N 个结果, 然后取最大值:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) p_{ij}] b_j(M_t) \quad (12)$$

此时记录状态 $X_t = S_i$ 由哪个状态转移而来:

$$\varphi_t(j) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) p_{ij}] \quad (13)$$

然后得到最优总概率, 并记录最后时刻 n 的状态:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] X_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)] \quad (14)$$

由于 $0, \varphi_2(j), \cdots, \varphi_t(j), \varphi_n(j)$ 全部记录了使得概率 $P(O | \lambda)$ 最大的状态, 采用回溯, 可得最优序列 S :

$$X_t^* = \varphi_{t+1}(X_{t+1}^*) \quad (15)$$

其中, $t = T - 1, T - 2, \cdots, 1$.

3.2 云转码器

表 3 给出了 H. 264 视频在手机屏幕大小为 4 英寸时, 4 种分辨率下建议的码率^[5].

表 3 屏幕分辨率与建议码率

屏幕分辨率/P	1 080	720	480	320
码率/(kB/s)	8 500	3 500	1 800	1 000

屏幕分辨率与码率的关系, 也即屏幕像素与码率的关系. 分辨率包含水平和竖直两个方向, 将两个方向像素数值相乘, 通过拉格朗日插值法, 得出屏幕像素和码率的关系, 如图 5 所示. 其中, 横坐标是屏幕像素数, 纵坐标表示码率.

在得到屏幕像素和码率的关系 $M(P)$ 后, 算法 1 根据隐马尔科夫链预测的带宽, 动态的向用户传输视频, 此时视频码率更符合移动设备的硬件处理能力.

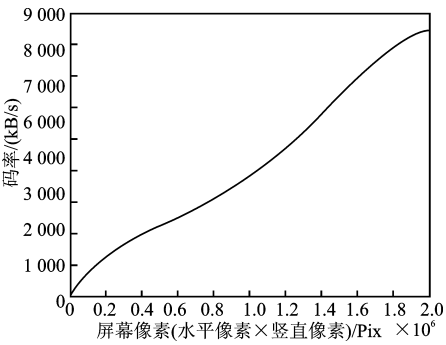


图 5 屏幕像素和码率关系图

算法 1 自适应视频流速率调整算法

```
1: Input  $B_0, P$ 
2: 初始化  $M(t)$ 
3:  $M(0) \leftarrow M(P)$ 
4: while 视频未传输完 do
5:   以  $M(t)$  速率向用户发送,  $t++$ 
6:   预测下一阶段带宽  $B'_{t+1}$ 
7:   if  $B'_{t+1} > M(P)$ 
8:      $M(t+1) = M(P)$ 
9:   end if
10:  else  $M(t+1) = B'_{t+1}$ 
11:    跳转到第四步
12:  end else if
13: end while
```

3.3 视频缓存器

由于多媒体业务所占用的带宽较大, 如果所有的请求都交给数据中心来处理, 会给系统带来很大的负担, 甚至严重时, 造成网络拥塞, 不能响应用户的请求. 为解决这个问题, 本文在靠近用户的网络边缘设置代理缓存. 设置缓存区可以减小网络延迟, 降低网络的负载, 提高用户的响应时间.

通常的缓存替代算法是 LRU 算法, 虽然算法比较稳定, 但是视频大量的进出会占用带宽资源, 造成网络性能下降.

现有的多媒体缓存策略主要基于访问次数, 也即流行度. 根据帕累托定律, 80% 的用户会访问 20% 的视频, 所以在本地缓存流行度高的视频, 快速响应用户请求. 本文的缓存管理策略是给每个视频定义流行值:

$$\text{value}(v_i) = S(v_i) \frac{P(v_i)}{T_{\text{current}} - T_{\text{last}}} \quad (16)$$

式中, T_{current} 是最近访问的时间, T_{last} 是上次访问的时间. $S(v_i)$ 和 $P(v_i)$ 分别是视频 v_i 的大小和访问次数.

QoE 的定义如式(17) 所示, 参考文献[6], 在满

足带宽的基础上, $\text{Max}(\text{QoE})$.

$$\text{QoE}=\text{Max} \frac{1}{(\text{Delay}+\text{k. Jitter})e^{\text{Packetloss}}}$$

(17)

s. t $M(t) \leq B(t)$

式中, Delay 是延迟, Jitter 是抖动, packetloss 是丢包. 根据式 (17) 的定义, QoE 最大, 即 $[(\text{Delay}+\text{K. Jitter})e^{\text{packetloss}}]$ 最小. 所以在评价用户体验时, 只需要考虑丢包、延迟、抖动 3 个参数, 这也是与视频流关系最密切的参数^[7], 如算法 2 所示.

算法 2 缓存管理算法

```
1: for 所有  $V_i$ 
2:   初始化  $P(V_i)$ 
3:   if  $V_i$  被请求观看
4:      $P(V_i)++$ 
5:   end if
6:    $\text{Value}(V_i)$ 
7:   把  $\text{value}(V_i)$  从大到小排序
8:   if  $V_i$  不在缓存中
9:     if 缓存足够
10:      把  $V_i$  放入缓存中
11:    else  $V_j$  在缓存中 and  $\text{value}(V_i) > \text{value}(V_j)$ 
12:       $V_i$  替换  $V_j$ 
13:    end if
14:  end if
```

当用户数量 $U(t)$ 小于设定阈值时, 服务器把流行度高的视频放入本地雾中. 本文在 3G/4G 环境下, 根据当下普及度较高的几款手机, 预先设定几种转码格式, 节约网络资源, 提高命中率^[8].

4 性能分析与评价

在网络模拟器 WANem 中对提出的云计算环境下的移动视频速率控制机制进行仿真研究和性能评价.

4.1 实验设计

设置网络环境变量如表 4 所示. 带宽预测实验时间为 60 s, 时间片为 5 s, 分别取模型状态 $N=5$ 和 $N=10$, 并对比连续预测 3 步和 1 步时的带宽预测值. 设置缓存区为 20 块, 每个块大小为 100 MB, 视频大小为 20~50 MB 不等. 由于 delay 和 Jitter 在数值上比较接近, 假设两个参数对网络的影响相似, 令 $K=1$.

表 4 实验网络参数设置

实验参数	带宽/(kB/s)	延迟/ms	抖动/ms	丢包
参数值	180	0.77	1	0.12

4.2 结果分析

比较分别是 $N=5$ 和 $N=10$, 步长为 3 时预测带宽与真实带宽的关系图. 当 $N=5$ 时的拟合程度优于 $N=10$. 如果状态分的过密, 则每个状态的值就很相近, 当带宽变化较大时, 误差就会增大. 而且, 不论是 $N=5$ 时还是 $N=10$ 时, 前几步的预测都非常精确, 随着预测步数的增多, 预测有了较大的偏差.

若把预测步长从 3 步减少到 1 步时, 拟合度会大幅提升. 因为带宽预测器连续预测 3 步, 后两步依据前 1 步的预测值, 如果前 1 步预测有误差, 将会影响后 2 步预测. 而预测 1 步时, 会基于当前的真实带宽重新预测, 更加精确.

基于流行度(PoP)算法, 如图 6 所示, 命中率高于 LRU 算法. 用流行度高的视频替换流行度低的视频, 减少了视频交换的次数, 节省大量带宽资源, 提高了 QoE.

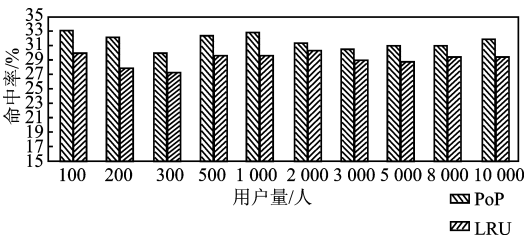


图 6 缓存命中率对比图

实验表明随着命中率的提高, QoE 也提升. 因为, 提高命中率, 意味着更多的用户请求可以在本地得到满足, 降低网络负担, 提高用户响应速度, 减少丢包和延迟, 根据式 (17), QoE 值得到了提高.

5 结束语

本文提出了云计算环境下的移动视频速率控制机制, 通过隐马尔科夫模型有效预测带宽, 根据用户硬件设备, 利用云转码器, 动态调整传输速率. 在此基础上, 提出缓存管理算法, 有效地提高用户请求视频的命中率. 仿真实验表明, 该机制能够有效地预测带宽, 提高用户 QoE, 节省资源. 本文并未考虑机制对移动设备能量的影响, 下一步将对此类问题展开研究.

参考文献:

[1] Thakolsri S, Kellerer W, Steinbach E. QoE-based rate adaptation scheme selection for resource-constrained wireless video transmission [C]// Proceedings of the

- ACM international conference on Multimedia. New York, NY, USA: ACM, 2010: 783-786.
- [2] Min Chen. AMVSC: A framework of adaptive mobile video streaming in the cloud[J]. IEEE Transactions on Multimedia, 2013, 15(4): 811-820.
- [3] Zhenhua Li, Yan Huang, Gang Liu, et al. Cloud-transcoder: bridging the format and resolution gap between internet videos and mobile devices[C]//Proceedings of the 22nd ACM international workshop on Network and Operating System Support for Digital Audio and Video. New York, NY, USA: ACM, 2012: 33-38.
- [4] Vaquero L M, Rodero-Merino L. Finding your way in the fog: towards a comprehensive definition of fog computing[J]. ACM SIGCOMM Computer Communication Review, 2014, 44(5): 27-32.
- [5] Shen H, Sun X, Wu F, et al. A fast downsizing video transcoder for H. 264/AVC with rate-distortion optimal mode decision[C]// Proceedings of IEEE International Conference on Multimedia and Expo. [s. l.]: IEEE, 2006: 2017-2020.
- [6] 林闯,胡杰,孔祥震. 用户体验质量(QoE)的模型与评价方法综述[J]. 计算机学报, 2012, 35(1): 1-15.
- [7] Zhou Y, Fu T Z J, Chiu D M, et al. An adaptive cloud downloading service[J]. IEEE Transactions on Multimedia, 2013, 15(4): 802-810.
- [8] Jiwei Li, Kai Bu, Xuan Liu, et al. ENDA: Embracing network inconsistency for dynamic application offloading in mobile cloud computing[C]//Proceedings of the second ACM SIGCOMM workshop on Mobile cloud computing. New York, NY, USA: ACM, 2013: 39-44.

作者简介:

邹露宁 男, (1989-), 硕士研究生. 研究方向为移动云计算、移动互联网、移动多媒体. E-mail: zouluning@163.com.

白光伟 男, (1961-), 博士, 教授, 博士生导师. 研究方向为云计算、移动互联网、无线多媒体网络、多媒体网络服务质量等.

沈航 男, (1984-), 博士, 讲师. 研究方向为云计算、无线网络编码、无线多媒体通信协议等.

纪传伟 男, (1989-), 硕士研究生. 研究方向为云计算、视频通话.

(上接第 101 页)

参考文献:

- [1] 邹长华, 谭世恒, 林土胜. 基于混沌置乱和混沌加密的 DCT 域数字水印算法[J]. 微电子学与计算机, 2011, 28(5): 58-62.
- [2] 熊祥光, 王力, 瞿云云. 基于整数小波变换的彩色视频水印算法[J]. 微电子学与计算机, 2013, 30(4): 41-46.
- [3] 温泉, 孙钺锋, 王树勋. 零水印的概念与应用[J]. 电子学报, 2003, 31(2): 214-216.
- [4] 吴胜兵, 张定会, 霍瑶, 等. DWT 系数分块的零水印方案[J]. 数据通信, 2014, 34(3): 22-25.
- [5] Gao G Y, Jiang G P. Zero-bit watermarking resisting geometric attacks based on composite chaos optimized SVR model [J]. The Journal of China Universities of Posts and Telecommunications, 2011, 18(2): 94-101.
- [6] Vellaisamy S, Ramesh V. Inversion attack resilient zero-watermarking scheme for medical image authentication [J]. IET Image Processing, 2014, 8(12): 718-727.
- [7] 叶天语. 基于方差的奇异值分解域鲁棒零水印算法[J]. 光子学报, 2011, 40(6): 961-966.
- [8] 潘黄琼, 陈光喜, 丁勇, 等. 彩色图像 DWT 变换下的块 SVD 零水印[J]. 微电子学与计算机, 2012, 29(5): 50-53.

作者简介:

曾文权 男, (1979-), 硕士, 副教授. 研究方向为数字水印、图像处理和分析.

熊祥光(通信作者) 男, (1984-), 硕士, 讲师. 研究方向为图像处理和数字水印技术. E-mail: xxg0851@163.com