

## 感知应用特征与网络带宽的虚拟机在线迁移优化策略

李湘<sup>1</sup>, 陈宁江<sup>1,2</sup>, 杨尚林<sup>1</sup>, 李华<sup>1</sup>

(1. 广西大学计算机与电子信息学院, 广西 南宁 530004; 2. 广西大学广西多媒体通信与网络技术重点实验室, 广西 南宁 530004)

**摘 要:** 首先对虚拟机在迁移过程中, 内存脏页面产生数量与不同应用特征的密切关系进行了实验验证。在感知不同的虚拟机应用特征环境的基础上, 使用  $GM(1,N)$  灰色预测模型对每次迭代周期产生的内存脏页面数进行预测。同时使用残差修正进行误差调整, 使预测结果更可靠, 结合预测的内存脏页面数量进行网络带宽动态预留调整。实验表明: 与传统的 pre-copy 策略相比, 本方法在含网络密集型应用或内存密集应用的虚拟机迁移时, 能够有效提高网络性能, 降低迁移时间。

**关键词:** 虚拟机迁移; 灰色预测模型; 网络带宽; 应用特征

**中图分类号:** TP393

**文献标识码:** A

## Optimization strategy of virtual machine online migration with awareness of application characteristics and network bandwidth migration

LI Xiang<sup>1</sup>, CHEN Ning-jiang<sup>1,2</sup>, YANG Shang-lin<sup>1</sup>, LI Hua<sup>1</sup>

(1. School of Computer and Electronic Information, Guangxi University, Nanning 530004, China;

2. Guangxi Key Laboratory of Multimedia Communications and Network Technology, Guangxi University, Nanning 530004, China)

**Abstract:** Firstly the experiments to verify the relationship between the number of dirty memory pages and application characteristics which exist in virtual machine migration was conducted. Then, different virtual machine application characteristics were perceived, with which the number of dirty memory pages produced during the migrations was predicted by the use of  $GM(1,N)$  grey prediction model. At the same time, using residual correction to adjust error makes results more reliable. According to the prediction of memory dirty pages, network bandwidth was adjusted and reserved. Compared with the traditional pre-copy strategy, the given experiments show that the optimized strategy proposed can improve the performance of network and reduce migratory cost for the memory-intensive and network-intensive applications.

**Key words:** virtual machine migration, grey prediction model, network bandwidth, application characteristic

### 1 引言

在线迁移是一种网络密集型活动, 它要求传输几 GB 甚至几十 GB 的虚拟机内存状态从源宿主机到目的宿主机。目前, 主流的虚拟化平台(KVM、XEN 和 VMware 等)都支持 pre-copy 算法进行在线迁移。pre-copy 方法能够有效缩短迁移时间和提高迁移性

能, 但在实际应用中, 受迭代收敛性、虚拟机不同应用特征和资源限制的影响, 该方法迁移网络密集型和内存密集型虚拟机时迁移性能并不理想。由于 pre-copy 的广泛性和可研究性, 所以本文主要对 pre-copy 在线迁移方法进行优化。目前, 在预测虚拟机内存脏页面方法中都没有考虑到虚拟机应用特征环境对虚拟机内存脏页面的影响。虚拟机的应用特

收稿日期: 2017-09-14

通讯作者: 陈宁江, chnj@gxu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61363003); 国家科技支撑计划课题基金资助项目 (No.2015BAH55F02); 广西自然科学基金资助项目 (No.2017GXNSFAA198141)

**Foundation Items:** The Natural Science Foundation of China (No.61363003), The National Key Technology R&D Program of China (No.2015BAH55F02), The Natural Science Foundation of Guangxi Province (No.2017GXNSFAA198141)

征指的是虚拟机应用的资源使用情况，包括 CPU 资源使用率、内存资源使用率和 I/O 资源使用率。感知虚拟机应用特征可以有助于研究不同虚拟机应用特征产生内存脏页面的规律。目前，pre-copy 在线迁移改进方法往往通过删除重复内存页面数据和压缩内存页面、抑制相似内存页面产生等方法减少内存脏页面的数据传输，以降低迁移时间和宕机时间。然而上述方法往往会增加大量额外的 CPU 消耗，同时未考虑到网络带宽环境对传输内存脏页面的影响。

针对上述问题，本文首先感知应用特征环境，研究不同应用特征环境下产生内存脏页面的规律，并建立  $GM(1, N)$  灰色预测模型对含不同应用特征的虚拟机进行内存脏页面产生数量的预测；接着，为优化迁移性能，本文提出感知应用特征网络带宽预留调整算法：根据内存脏页面数量和应用对外负载量计算出所需网络带宽，并根据实际网络环境对迁移过程中的网络带宽进行调整或者预留。这样能够有效地减少迭代的总时间和产生内存脏页面的数量，从而提高虚拟机迁移的性能。

## 2 虚拟机在线迁移优化相关研究

### 1) 内存脏页面率预测相关工作

在迁移过程中，传输的数据量影响着迁移的性能。而每次迭代传输的数据量是由上一次迭代时间内所产生的内存脏页面数量所决定的。因此，对每个迭代周期中产生的内存脏页面数量进行有效预测对提升虚拟机迁移有着重要意义。目前，在内存脏页面率预测有一定的研究。获取内存脏页面采用 Hypervisor 提供的脏页面位图。常用的预测虚拟机内存页面率的方法有：采用平均法<sup>[1]</sup>、加权平均数<sup>[2,3]</sup>、概率法<sup>[4]</sup>。以上预测方法都忽略了虚拟机的应用特征对内存脏页面产生率的影响。因此，感知虚拟机的应用特征，本文提出基于灰色预测的内存脏页面预测方法。选择灰色预测模型的原因是它能鉴别系统因素之间发展趋势的相异程度，并能通过少量的

原始数据预测事物未来发展趋势的状况。

### 2) 在线迁移网络性能优化相关工作

在 pre-copy 迁移过程中，会有许多相同的页面被重复复制到目的虚拟机，这样造成了复制时间的变长。目前在迁移过程中优化网络性能已经取得了一定的研究成果。在优化网络性能方面，主要是基于 pre-copy、post-copy 和 hybrid 3 种典型在线迁移方法进行改进的。主要归为以下几类：数据删除技术<sup>[5,6]</sup>、内存页面压缩技术<sup>[7-10]</sup>、抑制空闲内存页面<sup>[11]</sup>和调整网络带宽方法<sup>[12-14]</sup>。

从表 1 可以看出，以上改进方法都是旨在减少迁移的内存脏页面数量，已达到降低迁移时间的目的，而大部分研究未考虑虚拟机的应用特征环境对内存脏页面的影响。另外，以上大部分方法都需要消耗大量额外资源消耗，比如 CPU。因此，针对上述的问题，本文首先对内存脏页面与虚拟机的应用特征是否有关系进行验证；然后感知虚拟机的应用特征环境，根据  $GM(1, N)$  对内存脏页面产生量进行预测，结合网络带宽环境进行网络带宽调整、预留。与别人工作不同之处：本文感知应用特征和网络环境的情况，能够优化网络密集型应用迁移和内存密集型应用迁移。

## 3 应用特征和网络带宽对虚拟机迁移影响分析

本节主要目的是通过实验验证应用特征和网络带宽对虚拟机迁移的影响。在虚拟机中运行不同的应用：内存密集型应用、CPU 密集型应用和网络密集型应用，然后对包含这些应用的虚拟机进行在线迁移，收集迁移过程中虚拟机应用特征、源和目的宿主机的资源使用情况以及虚拟机迁移性能特征（包含各迭代周期内存脏页面产生数量和迁移时间等）。通过实验分析含不同应用特征的虚拟机和网络带宽对产生内存脏页面和迁移总时间的影响。采用 KVM 虚拟化默认的在线迁移方法：pre-copy 在线迁移方法。

表 1 面向网络性能优化的在线迁移改进方法对比

模型	优化类型	优化指标	特点	不足
文献[6]的 IRLM	数据删除	迁移时间	减少迁移时间和网络流量	计算每个页面的散列值，应用 QoS 下降
文献[7]的 Live gang migration	数据删除	迁移时间网络带宽	考虑了部分内存页面内容相似性	未考虑不同负载和不同操作系统；额外消耗 CPU 和内存
文献[8]的 XBRL	数据压缩	宕机时间迁移时间应用 QoS	有效降低宕机和迁移时间	需要大量缓存，额外 CPU
文献[9]的 CBR	数据压缩	迁移时间网络带宽	减少迁移时间	影响应用性能，扩展性差
文献[10]的 HMDC	数据压缩	宕机时间迁移时间应用 QoS	消除了网络密集型页面容错	影响同一宿主机的应用性能，顽健性差

实验分组为 3 种场景：1) 无网络带宽限制；2) 网络带宽限制为 30 Mbit/s；3) 网络带宽限制为 50 Mbit/s。在 3 种场景下，依次对 VM1、VM2、VM3 和 VM4 进行在线迁移，收集虚拟机应用特征和迁移性能信息以及宿主机的资源使用情况，每种场景执行 5 次。

图 1 显示在不同应用特征环境和网络带宽环境下迁移过程所产生内存脏页面总数。从图 1 中可知，在无网络带宽限制和 50 Mbit/s 网络带宽限制的条件下，含内存密集型应用虚拟机所产生的内存脏页面数最多。而当网络限制在 30 Mbit/s 时，网络密集型应用的虚拟机产生内存脏页面最多，超过内存密集型应用的虚拟机。这是因为在网络竞争较小的情况下，内存密集型应用内存使用率高，容易产生更多的内存脏页面。而当网络带宽资源的竞争加大，网络传输速率受到严重影响，使传输内存脏页面的时间变长，这样在这个迭代周期内产生的内存脏页面就更多。同时，在相同网络带宽条件下，无负载的虚拟机与含有 CPU 密集型应用的虚拟机产生的内存脏页面数量大体相近。

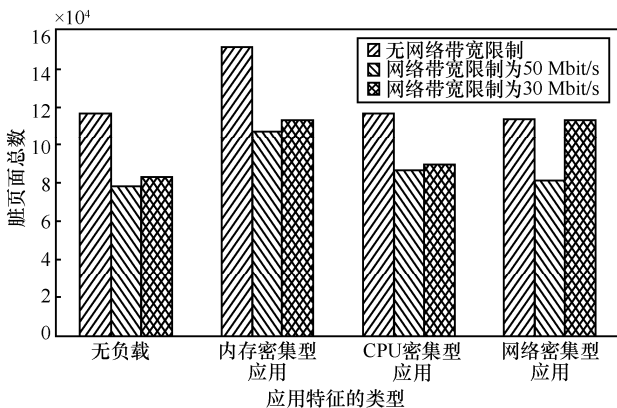


图 1 不同网络带宽和应用特征下内存脏页面的产生量

由此可知不同应用特征产生内存脏页面量有一定差别，虚拟机的应用特征影响着内存脏页面的产生，不同应用特征影响的程度不同。同时，在网络带宽可用资源较小情况下，含网络密集型应用的虚拟机会加大网络带宽的竞争，降低了传输效率，增加内存脏页面的产生量。

图 2 为不同网络带宽条件下，不同应用特征的虚拟机迁移时间对比。从图 2 可知，在无网络限制和网络限制在 50 Mbit/s 条件下，含内存密集型应用的虚拟机迁移所需时间最长，这是因为在网络带宽竞争较小的情况下，迁移时间主要受内存脏页面产

生量影响。内存密集型应用在迁移过程中会产生更多的内存脏页面，所需迁移时间更长。从总体上看，随着网络带宽的竞争压力增大，在网络带宽限制在 30 Mbit/s 时，含网络密集型应用的迁移时间超过含内存密集型应用的迁移时间。在同一应用特征中，无网络限制条件的迁移时间比在网络限制（50 Mbit/s 和 30 Mbit/s）条件下的迁移时间更短；而 30 Mbit/s 网络限制条件下所需迁移时间最长，这是因为在迁移过程中，虚拟机应用与迁移都需要网络带宽，造成了网络资源的竞争。尤其是网络带宽被限制在 30 Mbit/s 时，竞争最激烈。

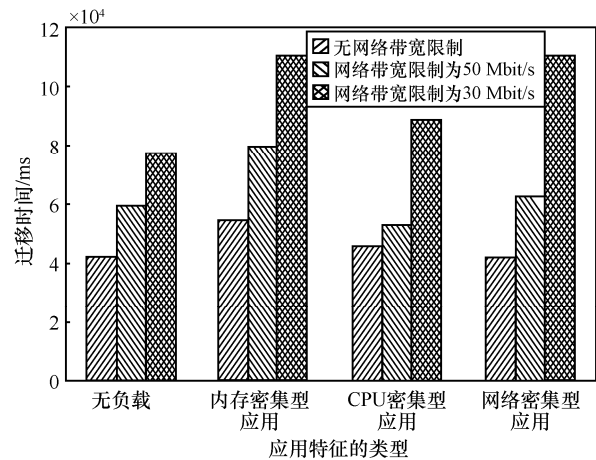


图 2 不同网络带宽和应用特征下虚拟机迁移时间

由此可知虚拟机迁移时间受虚拟机中的不同应用特征影响。当网络资源带宽竞争较小的情况下，含内存密集型应用的虚拟机影响迁移性能最大。在网络资源竞争较大的情况下，网络密集型应用对迁移时间影响加大，影响迁移性能。

综上所述，通过对比含不同应用特征的虚拟机产生内存脏页面数量和虚拟机迁移时间，得到不同应用特征影响内存脏页面的产生数量和迁移时间。含不同应用特征的虚拟机迁移，内存脏页面产生量有较大区别。网络带宽不同，内存脏页面的传输率也不同。内存脏页面产生的数量直接影响着每次迭代的时间，同时网络带宽影响着内存脏页面的传输效率，所以在第 4 节中，感知不同应用特征环境下，进行内存脏页面预测，并结合网络环境通过网络带宽预留调整对内存脏页面传输率进行调整。

#### 4 基于 GM(1, N)的内存脏页面数量预测模型

由于在线迁移的内存复制过程内存脏页面信

息采集比较难，采集的数量比较少，所以在迁移过程中得到的数据样本就比较少。针对样本数据较少的问题，灰色预测模型能较好地进行预测。因此本文采用灰色预测模型进行内存脏页面的预测。另外灰色预测模型还能够对内存脏页面与虚拟机的应用特征环境以及网络带宽环境的相关性进行分析。

灰色预测方法是一种对含有不确定因素的系统进行预测的方法。灰色预测通过鉴别系统因素之间发展趋势的相异程度，即进行关联分析，并对原始数据进行生成处理来寻找系统变动的规律，生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。灰色预测能更好解决“小样本”和“信息贫”的问题，其原理是通过原始数据进行处理：累加法、累减法 and 映射法，得到规律性较强的曲线，然后通过指数曲线拟合得到系统模型。灰色预测模型主要分为  $GM(1,1)$ 模型和  $GM(1,N)$ 模型。

从第 2 节得到，影响内存脏页面产生数量的因素主要是虚拟机的应用特征环境和网络带宽环境。而这些因素对内存脏页面产生量的耦合作用目前还没有人进行深入探讨。虚拟机内存使用率、CPU 使用率和虚拟机网络使用情况都是影响虚拟机产生内存脏页面的主要因素。 $GM(1,N)$ 状态模型可以对虚拟机内存使用率、虚拟机 CPU 使用率、虚拟机网络使用情况、迁移过程中整个网络使用情况和上一次迭代时间 5 个因素以及内存脏页面变量进行相关性分析，得到其变化发展规律，从而预测内存脏页面数。

对虚拟机内存脏页面预测变量选取，本文考虑以下能反映虚拟机应用特征环境的因素。

- $X_1$ : 迁移不同迭代周期产生的内存脏页面数量。
- $X_2$ : 迁移不同迭代周期虚拟机内存使用率。
- $X_3$ : 迁移不同迭代周期虚拟机 CPU 使用率。
- $X_4$ : 迁移不同迭代周期虚拟机网络吞吐率。
- $X_5$ : 上一次迭代的时间。
- $X_6$ : 迁移不同迭代周期宿主机的网络使用情况。

设迁移过程中总共进行  $S$  次迭代，每次内存脏页面产生数量都与上个迭代周期的应用特征和上一次迭代时间有关。基于  $G(1,N)$ 的内存脏页面预测模型建立过程如下。

**Step1** 给出灰色预测序列  $X^{(0)}$ ，令  $X^{(1)}$ 为  $X^{(0)}$ 的 AGO 序列。

$$X^{(0)} = \begin{bmatrix} x_{1(1)}^0 & x_{1(2)}^0 & \cdots & x_{1(S)}^0 \\ x_{2(1)}^0 & x_{2(2)}^0 & \cdots & x_{2(S)}^0 \\ \cdots & \cdots & \cdots & \cdots \\ x_{6(1)}^0 & x_{6(2)}^0 & \cdots & x_{6(S)}^0 \end{bmatrix},$$

$$X^{(1)} = \begin{bmatrix} \sum_{j=1}^1 x_{1(j)}^0 & \sum_{j=1}^2 x_{1(j)}^0 & \cdots & \sum_{j=1}^S x_{1(j)}^0 \\ \sum_{j=1}^1 x_{2(j)}^0 & \sum_{j=1}^2 x_{2(j)}^0 & \cdots & \sum_{j=1}^S x_{2(j)}^0 \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{j=1}^1 x_{6(j)}^0 & \sum_{j=1}^2 x_{6(j)}^0 & \cdots & \sum_{j=1}^S x_{6(j)}^0 \end{bmatrix} \quad (1)$$

**Step2** 求解  $z_1^{(1)}$ 。

$z_1^{(1)}$ 为  $x_1^{(1)}$ 的近邻值生成的序列，令  $z_1^{(1)}$ 为  $x_1^{(1)}$ 的均值序列，所以  $z_1^{(1)}(k) = \frac{1}{2}(x_1^{(1)}(k) + x_1^{(1)}(k-1))$ ，则

$$z_1^{(1)} = (z_1^{(1)}(2), z_1^{(1)}(3), \dots, z_1^{(1)}(S)) \quad (2)$$

**Step3** 建立灰微分方程。

$$x_1^{(0)}(k) + az_1^{(1)}(k) = b_2x_2^{(1)}(k) + b_3x_3^{(1)}(k) + \cdots + b_6x_6^{(1)}(k) \quad (3)$$

其中， $k = 1, 2, \dots, S$ ， $a$ 为发展系数， $b_i (i = 1, 2, 3, \dots, 6)$ 为驱动系数。

**Step4** 建立白化微分方程。

$$\frac{dx_1^{(1)}(k)}{dt} + a_1^{(1)}(k) = b_2x_2^{(1)}(k) + b_3x_3^{(1)}(k) + \cdots + b_6^{(1)}(k) \quad (4)$$

**Step5** 建立基于  $GM(1,N)$ 方程组。令参数列为

$$\hat{a} = [-a, b_2, b_3, \dots, b_6]^T, \quad Y = X_1^{(0)} = [x_{1(1)}^0 \ x_{1(2)}^0 \ \cdots \ x_{1(S)}^0]$$

$$B = \begin{bmatrix} z_1^{(1)}(1) & x_2^{(1)}(1) & \cdots & x_6^{(1)}(1) \\ z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_6^{(1)}(2) \\ \cdots & \cdots & \cdots & \cdots \\ z_1^{(1)}(S) & x_2^{(1)}(S) & \cdots & x_6^{(1)}(S) \end{bmatrix} \quad (5)$$

则  $GM(1,N)$ 灰微分方程为  $Y = B\hat{a}$ ，利用最小二乘法求解  $\hat{a}$

$$\hat{a} = (B^T B)^{-1} B^T Y \quad (6)$$

**Step6** 求解出  $GM(1,N)$ 的离散响应函数。将解得参数代入微分方程，则  $GM(1,N)$ 离散响应函数为

$$\hat{x}_1^{(1)}(k) = \left[ x_1^{(0)}(1) - \frac{b_2}{a}x_2^{(1)}(k) - \frac{b_3}{a}(k) - \cdots - \frac{b_6}{a}x_6^{(1)}(k) \right]$$

$$e^{-a(k-1)} + \frac{b_2}{a} x_2^{(1)}(k) + \frac{b_3}{a} x_3^{(1)}(k) + \dots + \frac{b_6}{a} x_6^{(1)}(k) \quad (7)$$

其中,  $k=1,2,3,\dots,S$ 。又因为  $\hat{x}_1^{(0)}(1) = x_1^{(0)}(1)$ , 累减还原计算式为

$$\hat{x}_1^{(0)}(k) = \hat{x}_1^{(1)}(k) - \hat{x}_1^{(1)}(k-1) \quad (8)$$

**Step7 残差修正**

灰色预测模型有时候预测出波动较大的数据会影响数据列发展规律, 因此需要对  $GM(1,N)$  进行模型修正。令  $\mathbf{f}^{(0)} = (\xi_1^{(0)}(1), \xi_1^{(0)}(2), \dots, \xi_1^{(0)}(S))$  为原始残差序列,  $\xi_1^{(0)}(k) = |x_1^{(0)}(k) - \hat{x}_1^{(0)}(k)|$ 。若  $k_0$  满足以下条件

- 1)  $\forall k \geq k_0, \xi_1^{(0)}(k)$  的符号相同;
- 2) 当  $S - k_0 \geq 4$  时, 可称  $(|\xi_1^{(0)}(k_0)|, |\xi_1^{(0)}(k_0+1)|, |\xi_1^{(0)}(k_0+2)|, \dots, |\xi_1^{(0)}(k_0+S)|)$  为可建模残差尾段, 记为

$$\mathbf{f}^{(0)} = (|\xi_1^{(0)}(k_0)|, |\xi_1^{(0)}(k_0+1)|, |\xi_1^{(0)}(k_0+2)|, \dots, |\xi_1^{(0)}(k_0+S)|) \quad (9)$$

对  $\mathbf{f}^{(0)}$  建立  $GM(1,1)$  模型, 求出其参数列  $\hat{\mathbf{a}} = [a_\xi, b_\xi]^T$ , 计算出  $\xi_{k_0+1}^{(0)}$  的模拟值

$$\xi_{(k+1)}^{(0)} = (-a)(\xi_{k_0}^{(0)} - \frac{b_\xi}{a_\xi})e^{-a_\xi(k-k_0)} \quad (10)$$

则  $GM(1,N)$  离散响应函数的修正式为

$$\hat{x}_1^{(1)}(k) = \begin{cases} \left[ x_1^{(0)}(1) - \frac{1}{a} \sum_{i=2}^6 b_i x_i^{(1)}(k) \right] e^{-a(k-1)} + \frac{1}{a} \sum_{i=2}^6 b_i x_i^{(1)}(k), & k < k_0 \\ \left[ x_1^{(0)}(1) - \frac{1}{a} \sum_{i=2}^6 b_i x_i^{(1)}(k) \right] e^{-a(k-1)} + \frac{1}{a} \sum_{i=2}^6 b_i x_i^{(1)}(k) \pm \xi_{(k)}^{(0)}, & k \geq k_0 \end{cases} \quad (11)$$

本节在感知虚拟机应用特征环境下, 建立预测内存脏页面产生数量的模型, 能够在采集少量数据样本的情况下, 得到较好的预测模型, 减少了系统的额外开销。同时, 通过残差修正对误差波动较大的结果进行动态调整, 有效保证了模型的准确性。本节通过在感知应用特征环境下的内存脏页面预测模型进行内存脏页面数预测, 预测出的内存脏页面数量将会作为下一节感知应用特征和网络环境的网络预留调整算法的一部分输入数据。

**5 应用特征感知的网络带宽预留调整算法**

在第 4 节中, 在感知应用特征环境下, 预测

虚拟机迁移过程中产生的内存脏页面数量, 得到每次迭代需传输的数据量。内存脏页面传输效率不仅与传输的数据量大小有关, 还与网络带宽有关。因此在迁移过程中, 根据不同内存脏页面数量, 进行有效的网络带宽管理对提高迁移性能有重要意义。

由于网络密集型应用会占用较多的网络带宽, 所以, 迁移过程中迁移虚拟机与虚拟机的应用会竞争网络带宽资源, 严重影响迁移数据的传输和虚拟机应用的性能。因此, 针对不同应用特征, 感知网络环境: 得到虚拟机应用所需网络带宽和系统网络带宽使用情况, 结合网络带宽环境进行网络调整, 提出了一个应用特征感知的网络带宽预留调整算法 (NRAC, network bandwidth reserve adjustment algorithm for different application characteristic)。通过收集虚拟机中应用所需网络带宽信息, 判断虚拟机是否属于网络密集型应用虚拟机。如果迁移虚拟机不是网络密集型应用虚拟机, 根据预测的内存脏页面数量, 判断目前空闲带宽是否满足内存脏页面数量的传输带宽 (即网络带宽是否大于内存脏页面率)。如果满足就不进行调整; 否则, 对网络带宽进行调整: 优先保证非网络密集型应用迁移, 延迟同一时间网络密集型虚拟机迁移。如果迁移的是网络密集型应用虚拟机, 根据历史网络传输流量数据进行应用带宽预留。将现在空闲网络带宽减去应用预留带宽就为现在虚拟机迁移所分配的带宽。NRAC 算法符号解释如表 2 所示。

**表 2 NRAC 算法符号解释**

符号	解释
$De(n, i)$	虚拟机 $n$ 第 $i$ 次迭代中产生的内存脏页面率
$Nu(i)$	第 $i$ 次迭代中网络带宽使用量
$Napp(n, k+1)$	使用基于时间序列的预测 $k+1$ 时刻的虚拟机 $n$ 的应用所需网络带宽
$N_{vm}(n, k)$	虚拟机的 $k$ 个网络流量历史记录
$Net_{vm}(n, j)$	虚拟机 $n$ 在 $j$ 时刻采集的网络流量值
$Nf(i)$	第 $i$ 次迭代的空闲网络带宽
$Th_r$	网络密集型应用的判断阈值, 网络流量值不小于 $Th_r$ 时, 为网络密集型应用虚拟机
$N_t$	网络带宽总量

NRAC 算法的伪代码如下:

**输入** 通过第 4 节预测模型预测虚拟机  $n$  在第  $i$  次迭代过程中的内存脏页面数  $D_p(n, i)$ ; 虚拟机  $n$ ;

第  $i$  次迭代时间  $T(i)$ ;

输出 在第  $i$  次迭代过程中分配的网络带宽为

$N_C(i)$ ;

begin

1) 获取  $N_u(i)$  和  $N_i$ ;

2) 得到  $N_f(i) \leftarrow N_i - N_u(i)$ ;

3) 获取虚拟机  $n$  的  $k$  个网络流量记录  $N_{vm}(n, k)$ ;

4) 计算虚拟机  $n$  第  $i$  次迭代的内存脏页面率:

$$D_e(n, i) \leftarrow \frac{D_p(n, i)}{T(i)};$$

5)  $N_C(i) \leftarrow N_f(i)$ ;

//判断虚拟机  $n$  是否为非网络密集型应用虚拟机

6) if (  $\sum_{x=1}^K Net_{vm}(n, x) / k \leq Th_r$  ) then

7) if ( $N_f(i) < D_e(n, i)$ ) then

8) 进行网络调整: 延迟将要迁移的网络密集型虚拟机进行迁移, 优先迁移非网络密集型应用虚拟机;

9) else then

10)  $N_C(i) \leftarrow D_e(n, i)$ ;

11) end if

12) else then

// 预留应用带宽

13) 获取  $N_{app}(n, k+1)$ ;

14) if ( $N_f(i) > (N_{app}(n, k+1) + D_e(n, i))$ ) then

15)  $N_C(i) \leftarrow N_{app}(n, k+1) + D_e(n, i)$ ;

16) end if

17) end if

18) return  $N_C(i)$ ;

19) end

本算法根据不同应用特征产生的内存脏页面率, 结合网络环境进行网络带宽调整。网络预留调整原理如图 3 所示。

从图 3 中可以看出, 该算法以虚拟机应用所需的网络带宽情况判断虚拟机是否为网络密集型应用。如果不是网络密集型虚拟机, 根据灰色预测模型预测该虚拟机产生的脏页面数, 计算脏页面产生速率, 以此判断是否有足够带宽进行该非网络密集型虚拟机迁移。如果不够, 则优先迁移该非网络密

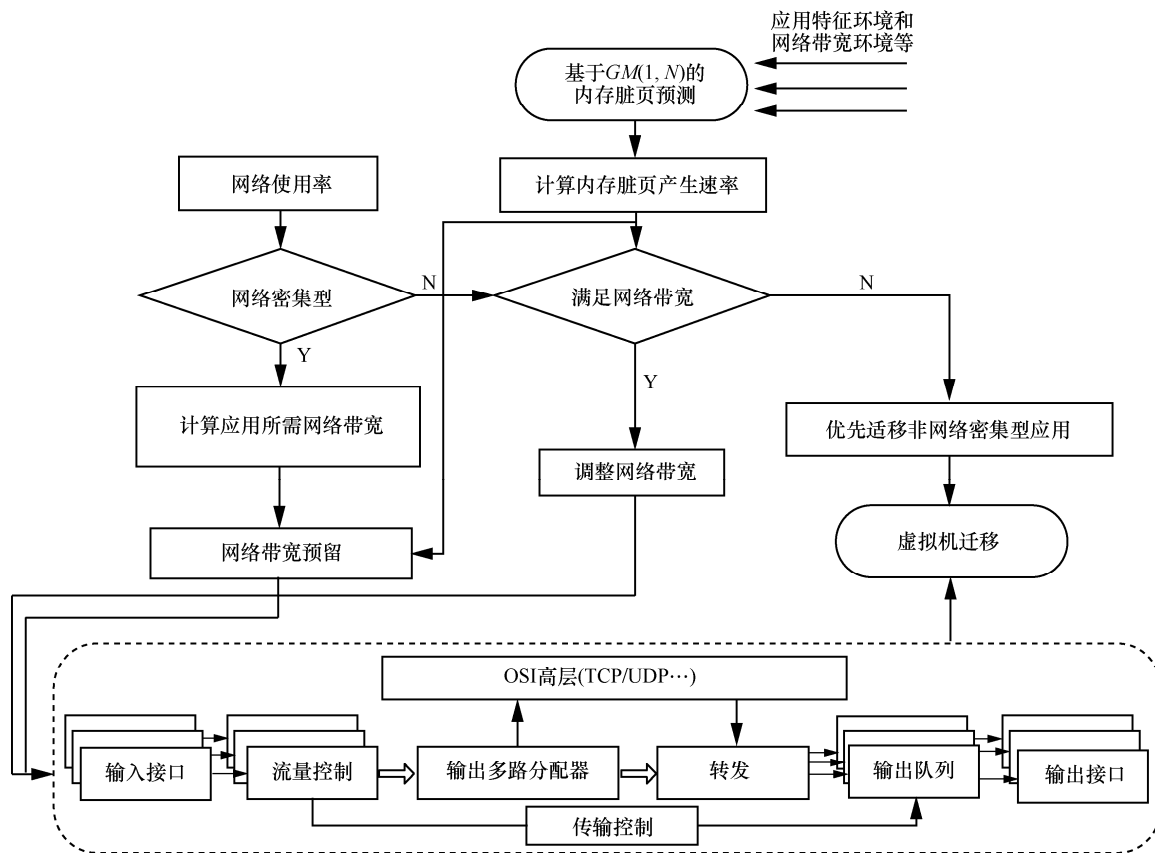


图 3 网络预留调整原理

密集型应用，推迟其他网络密集型虚拟机的迁移。反正有足够的网络带宽就分配给该虚拟机。在网络密集型虚拟机的情况下，如果剩余网络带宽能够满足虚拟机迁移和网络应用所需的带宽的话，则对网络带宽进行预留。预留后剩下的网络带宽为分配给迁移所需的网络带宽。

流量调整和控制主要由流量控制模块实现，即 Traffic control 模块<sup>[15]</sup>，主要分为输入接口、流量控制、输入多路分配器、转发、输出队列和输出接口 6 个模块。输入接口负责接收数据分组，并把数据分组传送到流量控制接口模块。流量控制接口负责筛选数据分组，将符合规定的数据分组传输到输入多路分配器模块中并把不符合规定的数据分组丢弃。输入多路分配器对输入的数据分组进行判断分析，决定数据分组的流向。如果数据分组为本主机的则将数据分组交给高层；否则，将数据分组传输给转发模块进行转发处理；通过查看路由表流，决定数据分组的下一跳。接着，数据分组会把传输到输出队列模块中，该模块会对数据分组进行排队，按照队列顺序将数据分组传输到输出接口。输出接口按照将数据分组传输给下一跳。流量控制主要在输出队列排列时处理和实现。

为了减少网络密集型应用虚拟机和应用之间对网络资源的竞争，本文为网络密集型应用虚拟机迁移预留其应用所需要的网络带宽资源，合理调整网络带宽资源，减少了网络拥塞发生的可能性。在含网络密集型应用迁移时，不仅迁移本身需要占用网络带宽，而且虚拟机的应用也需要占用带宽。在网络带宽分配不合理的情况下，容易加剧应用和虚拟机迁移对网络资源的竞争，这会导致迁移过程中数据传输率下降，从而导致迭代时间变长，而迭代时间变长又会导致产生的内存脏页面数量变多，这样很难进入迁移的 stop-and-copy 阶段，严重的甚至会导致迁移失败。本算法对含网络密集型应用，结合网络环境进行网络带宽预留，能够减少迁移过程中网络带宽的竞争，提高迁移效率，减少迭代时间，降低迁移时间，从而达到了提高迁移性能的目的。同时，优先保证非网络密集型虚拟机迁移，合理调配了网络资源、提高了网络的利用率。

## 6 实验

本文在基于 KVM 虚拟化技术的广西大学云教

育资源管理平台中对 NRAC 方法有效性进行验证。在实验中，对规格为 4 GB 内存、4 核 CPU、20 GB 磁盘的虚拟机进行迁移，搭建共享存储池。以不带任何负载的虚拟机迁移为基准，然后分别在虚拟机中加载内存密集型应用负载、CPU 密集型应用负载和网络密集型应用负载。

以不带任何负载的虚拟机迁移为基准，然后分别在虚拟机中加载内存密集型应用负载、CPU 密集型应用负载和网络密集型应用负载，通过实验分析含不同应用特征的虚拟机和网络带宽对产生内存脏页面和宕机时间的影响。实验结果如图 4 和图 5 所示。

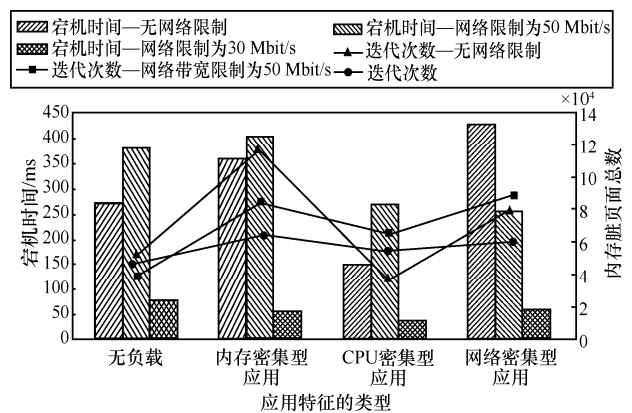


图 4 不同应用特征和网络带宽限制下的宕机时间和迭代次数

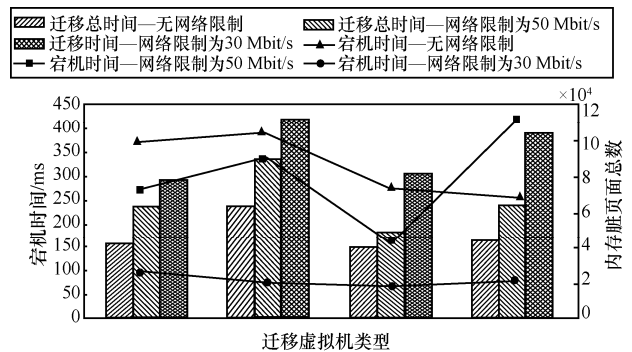


图 5 不同应用特征和网络带宽限制下的迁移时间和宕机时间

本文在虚拟化服务器 1 和虚拟化服务器 2 中部署 NRAC 方法。按照第 2 节的应用场景，依次对 VM1、VM2、VM3 和 VM4 使用 NRAC 方法进行在线迁移，收集虚拟机应用特征和迁移性能信息以及宿主机的资源使用情况，每种场景执行 15 次。本节实验与 KVM 系统默认的 pre-copy 在线迁移方法进行对比。图 6~图 8 为 NRAC 方法与 pre-copy 在线迁移方法在不同网络带宽限制条件下迁移时间的对比实验结果。

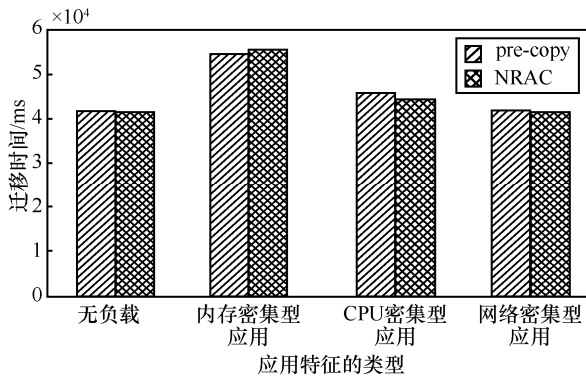


图 6 无网络带宽限制情况下的迁移时间

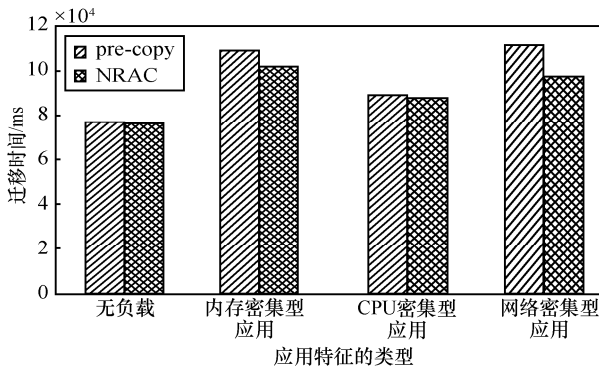


图 7 网络带宽限制在 50 Mbit/s 情况下的迁移时间

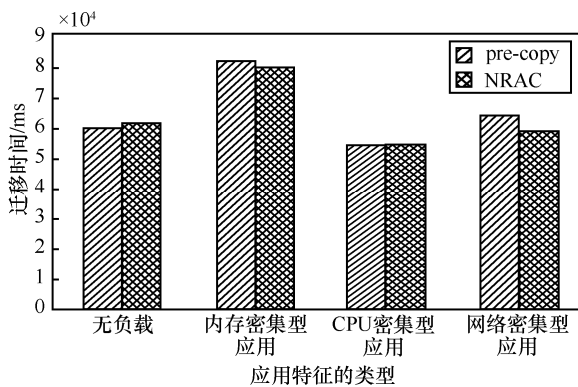


图 8 网络带宽限制在 30 Mbit/s 情况下的迁移时间

1) 在无网络带宽限制情况下,不同应用特征环境下的迁移时间。

图 6 为无网络带宽限制情况下的迁移时间的实验结果。在各应用特征类型下,同一应用特征的虚拟机使用 pre-copy 在线迁移方法和 NRAC 方法进行在线迁移,其迁移时间基本相近。这是因为在无网络带宽限制情况下,有充分的网络带宽资源,虚拟机迁移不受网络带宽限制。同一应用特征类型下,资源使用率相近,每次迭代产生的内存脏页面数量基本相近,所以迁移总时间也是相近的。

2) 网络带宽限制在 50 Mbit/s 以内的情况下,不同应用特征环境下的迁移时间。

在 50 Mbit/s 网络带宽限制下,如图 7 所示,在无负载和 CPU 网络密集应用特征情况下,同一应用特征下的 NRAC 方法和 pre-copy 在线迁移方法的迁移时间是相近的。而同内存密集应用特征或者同网络密集型应用特征下, NRAC 方法比 pre-copy 在线迁移方法所用的迁移时间更少。NRAC 方法根据网络带宽使用情况能够有效调整网络带宽,提高网络传输率,减少每次迭代的时间,从而内存脏页面产生量也会相对减少。因此,在 50 Mbit/s 网络带宽限制下, NRAC 方法能够降低含网络密集型应用和含内存密集型应用的虚拟机迁移总时间。

3) 网络带宽限制在 30 Mbit/s 以内的情况下,不同应用特征环境下的迁移时间。

图 8 为在网络带宽限制在 30 Mbit/s 的情况下各应用特征下的迁移时间。网络带宽限制在 30 Mbit/s 的情况下,网络带宽资源存在竞争,在相同的应用特征下, NRAC 方法的迁移时间比 pre-copy 在线迁移方法的少,尤其是虚拟机含内存密集型应用和含网络密集型应用的情况下。这是因为 NRAC 方法能够有效预测每次迭代产生的内存脏页面的数量,计算内存脏页面率,根据内存脏页面率和网络带宽使用情况进行网络带宽的预留调整。同一时间内,保证含非网络密集型应用的虚拟机优先迁移同时为含网络密集型应用的虚拟机预留应用所需的网络带宽,减少了网络带宽的竞争,有效提高每次迭代的网络传输率,这样减少了每次迭代的时间。因此,在网络带宽竞争较严重的情况下, NRAC 方法与 pre-copy 在线迁移方法相比,能够有效降低迁移时间。

综上所述,在无网络带宽限制或网络带宽竞争较小的情况下, pre-copy 在线迁移方法和 NRAC 方法迁移时间是相近的。在网络带宽资源存在较大竞争情况下, NRAC 方法能提高网络带宽资源利用率,减少网络带宽资源的竞争,提高网络传输效率,减少迭代的时间,从而减少该次内存脏页面产生数量和迁移时间。

### 7 结束语

传输的数据量和网络性能是影响在线迁移性能的重要因素。而目前预测虚拟机内存脏页面方法中大部分都没有考虑到虚拟机应用特征环境对虚

拟机内存脏页面的影响。在网络性能优化中, 大部分方法都是通过减少传输数据量来提高网络性能, 常采用重复内存页面数据删除和内存页面压缩、抑制相似内存页面产生<sup>[12]</sup>等方法。与本文方法相比, 这些方法会带来较大的额外资源开销如 CPU 资源开销, 同时这些方法的效率还依赖于删除算法、压缩和解压算法的效率。不仅如此, 还未考虑网络带宽环境对传输内存脏页面的影响。

针对以上在线迁移过程中存在的问题, 本文首先通过实验验证虚拟机迁移与虚拟机的应用特征和网络带宽有关。与网络带宽调整方法相比, 本文能够感知不同的应用特征环境进行网络带宽调整。在感知不同虚拟机的应用特征环境下, 使用  $GM(1,N)$  模糊预测方法进行内存脏页面数量的预测, 并通过残差修正保证预测方法的准确度。该方法能在较少的数据样本的情况下, 获得较准确的内存脏页面数量预测值。在得到内存脏页面数量预测模型之后, 根据内存脏页面数量对网络带宽环境进行预留调整。网络带宽预留调整不仅考虑了虚拟机迁移过程中的网络流量还针对虚拟机中含网络密集型的应用所需的网络带宽, 为应用预留所需的网络带宽。通过对比实验, 相比传统的 pre-copy 在线迁移方法, 本文方法在含网络密集型应用或内存密集应用的虚拟机迁移过程中, 能有效减少网络资源的竞争, 提高网络传输效率, 减少迁移时间。

### 参考文献:

- [1] XU F, LIU F, LIU L. iAware: making live migration of virtual machines interference-aware in the cloud[J]. IEEE Transactions on Computers, 2014, 63(12): 3012-3025.
- [2] ALDHALAAN A, MENASCÉ D A. Analytic performance modeling and optimization of live VM migration[C]//European Workshop on Performance Engineering(EPEW 2013). 2013: 28-42.
- [3] CUCKLE H, WALD N. The impact of screening for open neural tube defects in England and wales[J]. Cluster Computing, 2011, 16(2): 171-182.
- [4] ZHENG J, NG T S E, SRIPANIDKULCHAI K. Pacer: a progress management system for live virtual machine migration in cloud computing[J]. IEEE Transactions on Network & Service Management, 2013, 10(4): 369-382.
- [5] DESHPANDE U, KULKARNI U, GOPALAN K. Inter-rack live migration of multiple virtual machines[C]//The 6th international workshop on Virtualization Technologies in Distributed Computing. 2012: 19-26.
- [6] DESHPANDE U, WANG X, GOPALAN K. Live gang migration of virtual machines[C]//The 20th international symposium on High performance distributed computing. 2011:135-146.
- [7] SVÅRD P, HUDZIA B, TORDSSON J. Evaluation of delta compression techniques for efficient live migration of large virtual machines[J]. ACM Sigplan Notices, 2011, 46(7): 111-120.
- [8] WOOD T, RAMAKRISHNAN K, SHENOY P. CloudNet: dynamic pooling of cloud resources by live WAN migration of virtual machine[J]. IEEE/ACM Transactions on Networking, 2011, 46(7): 121-132.
- [9] HU L, ZHAO J, XU G. HMDC: live virtual machine migration based on hybrid memory copy and delta compression[J]. Applied Mathematics, 2013,7(2L): 639-646.
- [10] FORSMAN M, GLAD A, LUNDBERG L. Algorithms for automated live migration of virtual machines[J]. Journal of Systems and Software, 2015, 101(3): 110-126.
- [11] KOTO A, YAMADA H, OHMURA K. Towards unobtrusive VM live migration for cloud computing platforms[C]//The Asia-Pacific Workshop on Systems. 2012: 7.
- [12] 陈阳, 怀进鹏, 胡春明. 基于内存混合复制方式的虚拟机在线迁移机制[J]. 计算机学报, 2011, 34(12): 2278-2291.  
CHEN Y, HUAI J P, HU C M. Live migration of virtual machines based on hybrid memory copy approach[J]. Chinese Journal of Computers, 2011, 34(2): 2278-2291.
- [13] DESHPANDE U, KEAHEY K. Traffic-sensitive live migration of virtual machines[C]//15th IEEE/ACM International Symposium on Cluster/Cloud and Grid Computing (CCGrid). 2015: 51-60.
- [14] 刘诗海, 孙宇清, 刘古月. 面向业务特征的自适应虚拟机迁移带宽分配算法[J]. 计算机学报, 2013, 36(9): 1816-1825.  
LIU S H, SUN Y Q, LIU G Y. an adaptive bandwidth allocation algorithm for virtual machine migration based on service features[J]. Chinese Journal of Computers, 2013, 36(9): 1816-1825.

### 作者简介:



李湘 (1991-), 女, 广西贺州人, 广西大学硕士生, 主要研究方向为软件工程、网络分布式计算。



陈宁江 (1975-), 男, 广西南宁人, 博士, 广西大学教授, 主要研究方向为软件工程、网络分布式计算、云计算、大数据等。

杨尚林 (1992-), 男, 山东滕州人, 广西大学硕士生, 主要研究方向为软件工程、网络分布式计算。

李华 (1994-), 男, 广西北海人, 广西大学硕士生, 主要研究方向为网络软件工程。