

# 无线虚拟网络中基于自回归滑动平均预测的在线自适应虚拟资源分配算法

唐伦 杨希希\* 施颖洁 陈前斌

(重庆邮电大学移动通信技术重点实验室 重庆 400065)

**摘要:** 该文针对无线虚拟化网络中业务的不确定和信息反馈的时延而引起虚拟资源分配不合理, 提出一种基于自回归滑动平均(ARMA)预测的在线自适应虚拟资源分配算法。首先, 该算法以保障虚拟网络队列上溢概率为目标对时频资源和缓存资源进行联合分配, 并建立虚拟网络总成本最小化的理论分析模型。其次, 考虑到虚拟网络对不同资源差异化的应用需求, 设计了一种多时间尺度的资源动态调度机制, 在长周期上基于ARMA模型的预测信息实现缓存资源的预留策略, 在短周期上基于利用大偏差原理推导的队列上溢概率对虚拟网络优先级排序, 并根据确定的优先级动态调度时频资源, 从而满足各虚拟网络的业务需求。仿真结果表明, 该算法可有效降低比特丢失率, 同时提升物理资源的利用率。

**关键词:** 无线虚拟化网络; 资源分配; 多时间尺度; 自回归滑动平均; 大偏差原理

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2019)01-0016-08

DOI: 10.11999/JEIT180048

## ARMA-prediction Based Online Adaptive Dynamic Resource Allocation in Wireless Virtualized Networks

TANG Lun YANG Xixi SHI Yingjie CHEN Qianbin

(Key Laboratory of Mobile Communication Technology, Chongqing University of Post and Telecommunications, Chongqing 400065, China)

**Abstract:** In order to solve the unreasonable virtual resource allocation caused by the uncertainty of service and delay of information feedback in wireless virtualized networks, an online adaptive virtual resource allocation algorithm proposed based on Auto Regressive Moving Average (ARMA) prediction. Firstly, a cost of virtual networks minimization is studied by jointly allocating the time-frequency resources and buffer space, while guaranteeing the overflow probability of each virtual network. Secondly, considering the different demand of virtual networks to different resources, a resource dynamic scheduling mechanism designed with multiple time scales, in which the reservation strategy of buffer space is realized based on the ARMA's prediction information in slow time scale and the virtual networks are sorted according to the overflow probability derived by the large deviation principle and dynamically schedules the time-frequency resources in fast time scale, so as to meet the service demand. Simulation results show that the algorithm can effectively reduce the bit loss rate and improve the utilization of physical resources.

**Key words:** Wireless virtualized networks; Resource allocation; Multiple time-scale; Auto Regressive Moving Average (ARMA); Large deviation principle

### 1 引言

随着智能终端的蓬勃发展, 多样化的应用场景对时延、可靠性、吞吐量的不同需求, 给现有网络带来了巨大的挑战<sup>[1]</sup>。无线网络虚拟化技术通过频

谱共享、基础设施虚拟化、空口虚拟化等技术实现对网络资源的统一管理, 在满足不同应用场景的业务需求的同时, 提高网络部署的灵活性和降低无线网络运行维护成本<sup>[2,3]</sup>。无线网络虚拟化的本质是将无线网络资源进行切割并重组为虚拟的网络资源, 然后根据各服务供应商的需求为其分配适量的虚拟网络资源, 以此形成多个共存于同一物理网络但逻辑上相互独立的虚拟网络, 从而实现物理资源的高效利用<sup>[4]</sup>。然而, 考虑到物理网络资源的有限性和虚拟网络业务需求的多样性, 需要设计一种有

收稿日期: 2018-01-15; 改回日期: 2018-09-26; 网络出版: 2018-10-19

\*通信作者: 杨希希 469519917@qq.com

基金项目: 国家自然科学基金(61571073)

Foundation Item: The National Natural Science Foundation of China (61571073)

效的虚拟资源分配方法来提升网络虚拟化性能<sup>[5]</sup>。

文献[6]分别为不同虚拟网络和服务用户设计了优先等级，并提出了一种启发式的控制机制来解决虚拟网络的接入控制和服务用户的资源动态调度问题。文献[7]为了实现网络资源的高效分配，将无线虚拟化网络中的缓存空间和无线带宽调度问题建立成一种博弈机制。文献[8]提出了在C-RAN场景下的虚拟化方案，并基于云的massive-MIMO(Multiple Input and Multiple Output)技术研究了频谱资源与天线资源之间的互补性和部分可替代性。文献[9]研究了在单蜂窝场景中的虚拟网络隔离方案，通过对各虚拟网络分配正交的频谱块来避免虚拟网络间的干扰，并对各频谱块上的发射功率进行了动态调整。文献[10]基于业务预测信息设计了一种可提高网络利用率同时满足不同网络切片的服务等级协议的虚拟资源分配算法。

尽管已有大量文献致力于无线虚拟化环境中的资源分配算法的研究，但多数工作都只停留在解决单一调度周期上的资源调度问题，而忽略了在时间域上网络的动态变化情况。考虑到网络状态的随机性和信息反馈的延迟引起不合理的虚拟网络资源分配问题，本文设计了一种基于自回归滑动平均(Auto Regressive Moving Average, ARMA)预测的虚拟资源动态调度算法。本文的贡献主要包括：(1)将时频资源和缓存资源作为虚拟化实体，建立基于虚拟网络总成本最小化的资源分配联合优化模型；(2)考虑到缓存资源的长期利用和瞬时可达的传输服务需求，设计了一种包含多种时间尺度的调度机制，分别在长周期和短周期上分配缓存资源和时频资源；(3)为了提高缓存资源的利用率以及降低租用成本，对各虚拟网络建立独立的ARMA预测模型并基于预测结果在长周期上执行缓存资源的预留策略；(4)利用大偏差原理推导出各虚拟网络的队列上溢概率，并根据溢出剩余时间或上溢概率对各虚拟网络进行动态排序，以保障队列上溢概率为前提在短周期上动态调度时频资源。

## 2 系统模型

### 2.1 系统架构

图1所示为基于虚拟化技术的资源动态配置的系统架构图。文中用 $\mathbf{K}$ 表示系统中的虚拟网络集合。

### 2.2 问题描述

对虚拟网络 $k, k \in \mathbf{K}$ ，令 $A_k(t) \in \mathbf{A} \triangleq \{0, 1, \dots, A_k\}$ 表示其服务用户在调度周期 $t$ 内到达的业务数量，其中 $A_k$ 为单个周期内可到达的最大业务量。 $A_k(t)$ 是一个随机独立同分布的过程。令 $D_k(t) \in \mathbf{D} \triangleq \{0,$

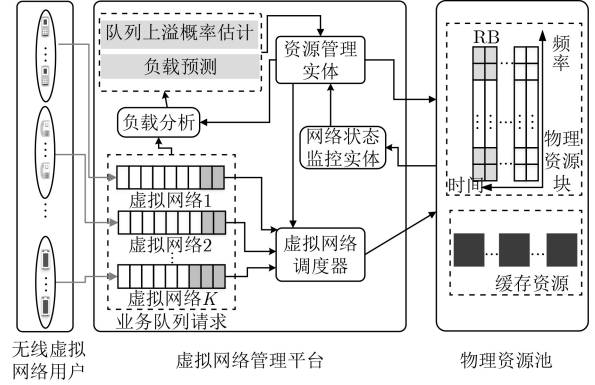


图1 系统架构

$1, \dots, D_k\}$ 表示虚拟网络 $k$ 在调度周期 $t$ 内离开的业务数量，其中 $D_k$ 为单个周期内可离开的最大业务量。另外，定义 $Q_k(t)$ 为虚拟网络 $k$ 在调度周期 $t$ 开始时的队列长度。因此，虚拟网络 $k$ 的队列更新模型为

$$Q_k(t+1) = \max\{Q_k(t) - D_k(t) + A_k(t), 0\} \quad (1)$$

由Little定理可得

$$\bar{Q}_k = \lambda_k \bar{L}_k \quad (2)$$

式中， $\bar{Q}_k$ 表示虚拟网络 $k$ 的平均队列长度， $\bar{L}_k$ 代表平均排队时延， $\lambda_k$ 为业务的到达速率。为了刻画业务请求与所分配的缓存资源的匹配程度，本文定义了各虚拟网络的队列上溢概率，具体可表示为

$$P_{\text{of}}^k = P(Q_k(t) > B_k), \forall k \quad (3)$$

式中， $B_k$ 代表虚拟网络 $k$ 在当前调度周期内租用的缓存资源大小。各虚拟网络的QoS需求可描述成式(4)中的队列上溢概率约束问题：

$$P(Q_k(t) > B_k) < \varepsilon_k, \forall k \quad (4)$$

为了最优地保证QoS需求、降低虚拟网络提供服务的总成本，本文的资源动态配置问题可建立成式(5)的数学模型：

$$\left. \begin{aligned} \min_{\mathbf{B}, \mathbf{x}} \sum_{k=1}^K \left[ \rho_k B_k + \alpha_k \sum_{n=1}^N x_{n,k} \right] \\ \text{s.t. C1: } \sum_{k=1}^K x_{n,k} = 1, \forall n \\ \text{C2: } P(Q_k(t) > B_k) < \varepsilon_k, \forall k \end{aligned} \right\} \quad (5)$$

式中，各虚拟网络的资源租用成本由两部分组成，第1部分代表租用缓存资源的费用，第2部分为使用时频资源的代价，令 $\rho_k, \alpha_k$ 分别代表虚拟网络 $k$ 租用缓存资源和时频资源的单价， $x_{n,k}$ 表示时频资源的分配指示， $x_{n,k} = 1$ 表示时频资源块 $n$ 配置给虚拟

网络 $k$ , 这里假设1个时频资源块只能配置给1个虚拟网络使用, 因此可得到约束C1。

### 3 多时间尺度资源调度策略

在实际应用中各虚拟网络的业务请求随机地到达, 而虚拟缓存资源的配置具有一定的延迟<sup>[11]</sup>, 在同一周期内确定缓存资源和时频资源的调度策略将是不现实、不准确的。因此, 本文设计了一种包含多种时间尺度的资源配置方式, 具体分配过程如图2所示。

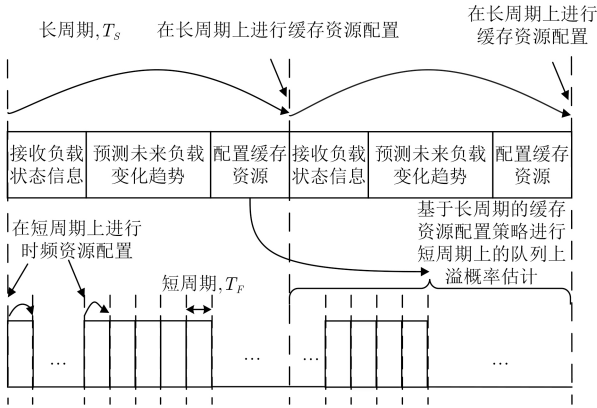


图2 多时间尺度的资源配置示意图

#### 3.1 基于ARMA预测的缓存资源预留策略

本文提出的多时间尺度资源配置框架, 需要设计有效的预测方法来实现各虚拟网络负载的准确预测, 辅助完成缓存资源的提前配置工作。常见的预测算法包括时间序列预测、神经网络、Markov模型、灰色预测模型等<sup>[12]</sup>。考虑到ARMA融合了回归分析和时间序列分析功能, 预测误差的方差更小<sup>[13]</sup>, 在此, 本文通过构建ARMA模型来预测各虚拟网络未来的负载状态。

将负载状态划分为 $L$ 个级别, 可以表示为从低到高的有限状态空间 $B_k(t) \in \{B^1, B^2, \dots, B^L\}$ 。基于ARMA模型, 各虚拟网络未来的平均负载状态为

$$y_k(t) = \varphi_1 y_k(t-1) + \dots + \varphi_p y_k(t-p) + \xi(t) - \theta_1 \xi(t-1) - \dots - \theta_q \xi(t-q) \quad (6)$$

式中,  $\{y_k(t-i) | i=1, 2, \dots, p\}$  代表虚拟网络 $k$ 在过去 $p$ 个长周期上的平均负载状态,  $\{\xi(t-i) | i=1, 2, \dots, q\}$  表示方差为 $\sigma^2$ 的白噪声,  $\{\varphi_i | i=1, 2, \dots, p\}$  和  $\{\theta_i | i=1, 2, \dots, q\}$  是待估计的模型参数。

考虑到各虚拟网络可能具有不同的负载特性, 将分别根据各虚拟网络的历史数据构造独立的预测模型, 具体过程如下:

第1步 估计参数 $\{\varphi_i | i=1, 2, \dots, p\}$  和  $\{\theta_i | i=1,$

$2, \dots, q\}$ 。利用对数处理后的观察序列 $y_k'$ 可以得到各虚拟网络的自协方差:

$$\gamma_k(i) = \frac{1}{I} \sum_{j=1}^{I-i} y_k'(i) y_k'(i+j) \quad (7)$$

式中,  $I$  表示可用观察序列规模。将式(6)两边同乘以 $y_k'(t-i)$ 再对其取均值可得到自协方差间的关系表达式:

$$\begin{aligned} \gamma_k(i) &= E\{y_k'(t) y_k'(t-i)\} \\ &= \varphi_1 E\{y_k'(t-1) y_k'(t-i)\} + \dots \\ &\quad + \varphi_p E\{y_k'(t-p) y_k'(t-i)\} \\ &\quad + E\{\xi(t) y_k'(t-i)\} \\ &\quad - \theta_1 E\{\xi(t-1) y_k'(t-i)\} - \dots \\ &\quad - \theta_q E\{\xi(t-q) y_k'(t-i)\} \end{aligned} \quad (8)$$

当 $i > q$ 时, 式(8)可改写成

$$\gamma_k(i) = \varphi_1 \gamma_k(i-1) + \varphi_2 \gamma_k(i-2) + \dots + \varphi_p \gamma_k(i-p) \quad (9)$$

基于式(9), 对于 $i = q+1, q+2, \dots, q+p$ 可得到方程组:

$$\left. \begin{aligned} \gamma_k(q+1) &= \varphi_1 \gamma_k(q) + \varphi_2 \gamma_k(q-1) + \dots \\ &\quad + \varphi_p \gamma_k(q+1-p) \\ \gamma_k(q+2) &= \varphi_1 \gamma_k(q+1) + \varphi_2 \gamma_k(q) + \dots \\ &\quad + \varphi_p \gamma_k(q+2-p) \\ &\vdots \\ \gamma_k(q+p) &= \varphi_1 \gamma_k(q+p-1) + \varphi_2 \gamma_k(q+p-2) + \dots \\ &\quad + \varphi_p \gamma_k(q) \end{aligned} \right\} \quad (10)$$

因此, 利用延伸的Yule-Walker方程可获得 $\{\varphi_i | i=1, 2, \dots, p\}$ 的估计值:

$$\begin{bmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \vdots \\ \gamma_{q+p} \end{bmatrix} = \begin{bmatrix} \gamma_k(q) & \gamma_k(q-1) & \dots & \gamma_k(q+1-p) \\ \gamma_k(q+1) & \gamma_k(q) & \dots & \gamma_k(q+2-p) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_k(q+p-1) & \gamma_k(q+p-2) & \dots & \gamma_k(q) \end{bmatrix} \cdot \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix} \quad (11)$$

接下来基于式(11)求解的 $\{\varphi_i | i=1, 2, \dots, p\}$ 估计 $\{\theta_i | i=1, 2, \dots, q\}$ 和噪声方差。由于 $y_k''(t) \triangleq y_k'(t) - \sum_{i=1}^p \varphi_i y_k'(t-i)$ 满足MA(Moving Average)模型, 因此可以近似地将 $y_k''(t-i)$ 看作MA( $q$ )的观察数据, 具体表示为

$$y_k''(t-i) = y_k'(t) - \varphi_1 y_k'(t-1) + \dots + \varphi_p y_k'(t-p), \quad t = p+1, p+2, \dots, I \quad (12)$$

同样地, 根据近似的观察数据可以得到相应的自协方差函数:

$$\begin{aligned}\gamma_k^*(i) &= E\{y_k^*(t)y_k^*(t-i)\} \\ &= E\left\{\left(-\sum_{j=0}^p \varphi_j y_k^*(t-j)\right)\left(-\sum_{l=0}^p \varphi_l y_k^*(t-i-l)\right)\right\} \\ &= \sum_{j,l=0}^p \varphi_j \varphi_l \gamma_k(i+l-j)\end{aligned}\quad (13)$$

利用式(13)的计算结果并结合MA模型的逆相关函数便可获得 $\{\theta_i|i=1,2,\dots,q\}$ 的估计值和白噪声 $\xi(t)$ 的方差 $\hat{\sigma}_\xi^2$ 。

第2步 确定各虚拟网络预测模型的阶数 $p, q$ 。本方案采用AIC(Akaike Information Criterion)准则来为各虚拟网络预测模型定阶。AIC函数可定义为

$$\text{AIC}(s) = \ln \hat{\sigma}_\xi^2 + (2s)/I \quad (14)$$

式中,  $s$ 为预测模型参数的总数, 即 $s = p + q + 1$ 。在一定 $p, q$ 变化范围内, 当AIC( $s$ )达到最小值时, 对应的值即为模型的阶数。

为了避免在新的长周期开始时发生大量数据丢失的情况, 本方案采用了静态与动态相结合的缓存资源预留方法。将ARMA模型预测的虚拟网络 $k$ 在下一长周期 $t$ 上的平均负载状态 $y_k(t)$ 与负载状态区间对比, 则可获得虚拟网络 $k$ 预留的静态部分 $B^{\text{st}}$ , 即 $B^{\text{st}} = B^l$ , 其中 $B^l$ 满足 $B^{l-1} < y_k(t) \leq B^l$ 。若在当前长周期结束时, 虚拟网络 $k$ 的瞬时队列长度 $Q_k(t)$ 远远大于 $B^l$ , 则需要为其预留动态部分 $B^{\text{dy}}$ , 令 $B^{\text{dy}} = B^{l'} - B^l$ , 其中 $B^{l'}$ 满足 $B^{l'-1} < Q_k(t) \leq B^{l'}$ , 因此, 虚拟网络 $k$ 预留的缓存资源为 $B_k(t) = B^{\text{st}} + B^{\text{dy}}$ 。当小周期上虚拟网络 $k$ 的瞬时队列长度 $Q_k(t')$ 满足 $B^{\text{st}} - Q_k(t') > A_k$ 时, 本方案考虑释放预留的动态部分, 具体流程如图3所示。

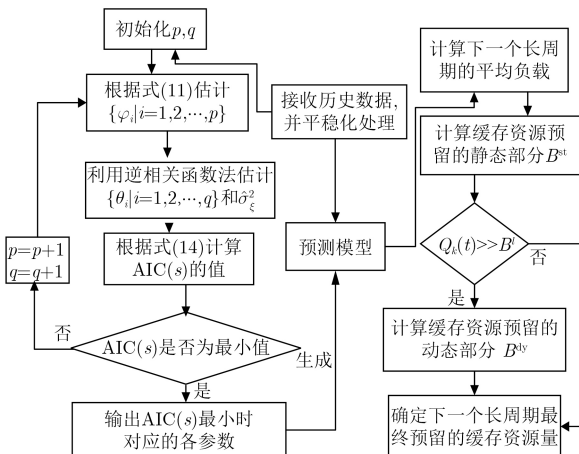


图3 长周期上基于ARMA预测的缓存资源预留策略流程图

### 3.2 具有QoS约束的时频资源调度策略

#### 3.2.1 队列上溢概率估计模型

定义 $\Delta_k(t) = A_k(t) - D_k(t)$ 为单周期上的队列增量, 其中 $\Delta_k(t)$ 的取值空间为 $\Delta_k(t) \in \{-D_k, \dots, 0, 1, \dots, A_k\}$ , 令 $\pi_k^d = P(\Delta_k(t) = d)$ 表示虚拟网络 $k$ 的队列变化的概率分布。

在第 $t$ 个短周期到第 $t+T$ 个短周期内, 其中 $T$ 代表预估周期, 虚拟网络 $k$ 的队列增量可表示为

$$\Delta_k(t+T) = \sum_{i=1}^T \Delta_k(t+i) \quad (15)$$

因此, 在第 $t+T$ 个短周期开始时, 虚拟网络 $k$ 的瞬时队列长度为

$$Q_k(t+T) = Q_k(t) + \Delta_k(t+T) \quad (16)$$

根据式(15), 式(16)可得虚拟网络 $k$ 在第 $t+T$ 个短周期上的队列上溢概率为

$$\begin{aligned}P_{\text{of}}^k(t+T) &= P(Q_k(t+T) > B_k) \\ &= P\left(Q_k(t) + \sum_{i=1}^T \Delta_k(t+i) > B_k\right) \\ &= P\left(\frac{\sum_{i=1}^T \Delta_k(t+i)}{T} > \frac{B_k - Q_k(t)}{T}\right) \\ &= P\left(\frac{\sum_{i=1}^T \Delta_k(t+i)}{T} > a_k\right)\end{aligned}\quad (17)$$

其中,  $a_k = (B_k - Q_k(t))/T$ 表示虚拟网络 $k$ 在未来 $T$ 个短周期内可接受的平均队列增长率, 令 $m_k = E\left[\sum_{i=1}^T \Delta_k(t+i)/T\right]$ 表示虚拟网络 $k$ 在未来 $T$ 个短周期内期望的平均队列增长率,  $E[\cdot]$ 是期望因子。式(17)的值越大则更容易发生队列溢出。

$\Delta_k(t)$ 满足一个有限瞬时生成的矩母函数 $G(\omega) = E[e^{\omega \Delta_k(t)}]$ 。若 $E[\Delta_k(t)] < a_k$ , 根据文献[14,15]可知序列 $\Delta_k(t)$ 满足大偏差原理, 因此, 当 $a_k > m_k$ , 利用Cramér's theorem可得<sup>[16]</sup>

$$\lim_{T \rightarrow \infty} \frac{1}{T} \lg P\left(\frac{\sum_{i=1}^T \Delta_k(t+i)}{T} > a_k\right) = -f(a_k) \quad (18)$$

式中,  $f(a_k)$ 为速率函数, 具体表示为

$$f(a_k) = \sup_{\omega > 0} \{a_k \omega - \lg G(\omega)\} \quad (19)$$

若已知 $\Delta_k$ 的概率分布 $\Delta_k \sim \left( \begin{matrix} -D_k, \dots, 0, 1, \dots, A_k \\ \pi_k^{-D_k}, \dots, \pi_k^0, \pi_k^1, \dots, \pi_k^{A_k} \end{matrix} \right)$ , 则 $\lg G(\omega)$ 可表示为

$$\lg G(\omega) = \lg \left\{ \sum_{d=-D_k}^{A_k} \pi_k^d e^{d\omega} \right\} \quad (20)$$

根据式(18), 对于足够大的 $T$ , 队列上溢概率可近似为

$$P_{\text{of}}^k(t+T) \approx e^{-Tf(a_k)} \quad (21)$$

尽管根据式(21)可得到队列上溢概率的估计值, 然而由于 $\Delta_k$ 的概率分布不可知, 将无法直接得到矩母函数 $G(\omega)$ 的值, 所以本文采用一种基于滑动窗口的方法来在线估计 $\pi_k^d$ .

假设滑动窗口的大小为 $T_w$ , 对于当前短周期 $t$ , 虚拟网络 $k$ 的观察向量可表示为 $\mathbf{W}_k(t) = [\Delta_k(t-1), \Delta_k(t-2), \dots, \Delta_k(t-T_w)]$ , 因此 $m_k$ 的估计值为

$$\hat{m}_k = \sum_{d=t-T_w}^{t-1} \Delta_k(d) / T_w \quad (22)$$

定义 $R_k^j$ 为在滑动窗口中 $\Delta_k(d) = j$ 发生的次数, 其发生概率可表示为 $\tilde{\pi}_k^d = R_k^j / T_w$ , 根据文献[17], 本文利用指数平滑法来柔和估计值, 具体表示为

$$\hat{\pi}_k^d(t) = \eta \hat{\pi}_k^d(t-1) + (1-\eta) \tilde{\pi}_k^d(t), \eta \in [0, 1] \quad (23)$$

### 3.2.2 时频资源动态调度策略

结合上述队列上溢概率的估计模型, 本文根据各虚拟网络的优先级依次为其配置适当的时频资源, 使其满足QoS需求。

对于任意的虚拟网络, 本文将分别计算处于 $\hat{m}_k \geq a_k$ 的溢出剩余时间 $T_k$ 和处于 $\hat{m}_k < a_k$ 的队列上溢概率 $P_{\text{of}}^k(t+T)$ , 其中虚拟网络 $k$ 的溢出剩余时间 $T_k$ 可近似为

$$T_k = \frac{B_k - Q_k(t)}{E[\Delta_k(t)]} \approx \frac{a_k}{\hat{m}_k} \quad (24)$$

式(24)的值越小, 则虚拟网络 $k$ 的优先级越高, 队列溢出上溢概率与阈值的差值 $P_{\text{of}}^k(t+T) - \varepsilon_k$ 越小, 则虚拟网络 $k$ 的优先级越低。

因此, 基于确定的虚拟网络优先级, 利用表1的算法1进行时频资源的调度。

## 4 性能仿真与结果分析

为了验证本文提出的基于负载预测的虚拟资源分配算法(virtual Resource Allocation based on

表1 算法1: 时频资源动态调度算法

- 
- (1) 在短周期 $t$ 上观察当前各虚拟网络队列状态 $Q_k(t)$ 、预留的缓存资源大小 $B_k$
  - (2) **for**  $k = 1; k < K; k++$  **do**
  - (3) 计算 $a_k$ , 根据式(22)估计 $\hat{m}_k$
  - (4) **if**  $\hat{m}_k \geq a_k$  **then**
  - (5) 加入虚拟网络集合 $\mathbf{K}_1$ , 根据式(24)估计溢出剩余时间 $T_k$
  - (6) **else**
  - (7) 加入虚拟网络集合 $\mathbf{K}_2$ , 执行黄金分割搜索算法估计 $P_{\text{of}}^k(t+T)$
  - (8) **end if**
  - (9) **end for**
  - (10) **while**  $\mathbf{K}_1 \neq \emptyset$  **do**
  - (11) 令 $m = 1$ , 选择虚拟网络 $k = \arg \min_{k \in \mathbf{K}_1} \{T_k\}$
  - (12) **while**  $A_k(t) > C_k(t)$  **do**
  - (13)  $m \leftarrow m + 1, C_k(t) \leftarrow mr,$   
 $N \leftarrow N - 1$
  - (14) **end while**
  - (15)  $\mathbf{K}_1 = \mathbf{K}_1 \setminus \{k\}$
  - (16) **end while**
  - (17) **while**  $\mathbf{K}_2 \neq \emptyset$  **do**
  - (18) 令 $m = 1$ , 选择虚拟网络 $k^* = \arg \max_{k^* \in \mathbf{K}_2} \{P_{\text{of}}^{k^*}(t+T) - \varepsilon_{k^*}\}$
  - (19) 重复步骤(12)—步骤(14)
  - (20)  $\mathbf{K}_2 = \mathbf{K}_2 \setminus \{k^*\}$
  - (21) **end while**
  - (22) **if**  $N \neq 0$  **then**
  - (23) **for**  $k = 1; k < K; k++$  **do**
  - (24) **if**  $C_k(t) < (Q_k(t) + A_k(t))$  **then**
  - (25) 加入虚拟网络集合 $\mathbf{K}_3$
  - (26) **end if**
  - (27) **end for**
  - (28) **while**  $\mathbf{K}_3 \neq \emptyset$  and  $N \neq 0$
  - (29) 令 $m = 1$ , 选择虚拟 $k'' = \arg \min_{k'' \in \mathbf{K}_3} \{\alpha_{k''}\}$
  - (30) **while**  $(Q_{k''}(t) + A_{k''}(t)) > (\bar{C}_{k''}(t) + C_{k''}(t))$  **do**
  - (31)  $m \leftarrow m + 1, \bar{C}_{k''}(t) \leftarrow mr, N \leftarrow N - 1$
  - (32) **end while**
  - (33)  $\mathbf{K}_3 = \mathbf{K}_3 \setminus \{k''\}$
  - (34) **end while**
  - (35) **end if**
- 

Load Forecasting, RALF), 本文将利用MATLAB工具进行仿真, 具体仿真参数设置如表2所示。

本节将分别与两种基本的缓存资源和时频资源调度策略进行对比。

(1)缓存资源对比方案: 采用两种静态的缓存资源预留方案SCBA (Static Conservative Buffer Allocation)和SABA (Static Abundant Buffer Allocation), 分别为各虚拟网络预留相对较少和较充

表 2 仿真参数设置

仿真参数	仿真值
虚拟网络数量	2,3,4,5,6
系统带宽	10 MHz (50 RBs)
短周期时长	1 ms
长周期时长	300 ms
负载到达过程	泊松分布
比特到达速率	$\lambda = 58.7$ kbit/ms
RB单价 $\alpha$	1.2, 2.0, 1.5 unit/RB
缓存资源单价 $\rho$	8, 6, 4 unit/kbit
队列上溢概率 $\epsilon$	0.13, 0.05, 0.12
滑动窗口大小 $T_w$	60 ms
平滑指数 $\eta$	0.7
仿真时间	6600 ms

裕的缓存资源。

(2)时频资源对比方案：与HS (Hard Slice)和SFP (Slice with Fixed Prioritization)两种时频资源调度方案进行对比，其中HS方案为各虚拟网络配置固定数量的时频资源，SFP方案根据固定的优先顺序为各虚拟网络动态配置时频资源数量。

接下来对以上算法从资源租用总成本、资源利用率、比特丢失率3方面进行评估。从图4可以看出，本文所提的RALF算法的资源租用总成本在数量较多时才出现明显的增长，而其他4种方案始终保持相近的速率递增。这是因为本文所提的RALF算法采用了基于负载预测的缓存资源预留机制，提前进行了针对性的缓存资源配置。

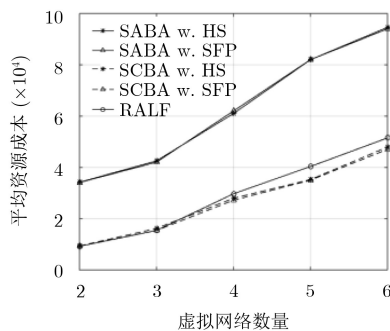


图 4 不同方案平均资源成本

图5, 图6显示了当虚拟网络数量较小时，各方案的比特丢失率都处于较为理想的水平。当数量增加，各虚拟网络开始竞争有限的时频资源，资源利用率不断增高，然而由于资源的不足将出现更加严重的比特丢失。尽管SABA方案的比特丢失率较低，但从图5可看出其在资源利用率上性能最差，而SCBA的资源利用率处于较高的水平，但由于其缓存资源配置的不充足，在比特丢失率上的性能十

分不理想。HS的资源利用率明显低于SFP，这是因为HS为每个虚拟网络配置固定数量的时频资源而忽略了各虚拟网络的真实需求，造成了部分物理资源处于空闲的状态。而本文所提的RALF算法在上述两方面均具有较为理想的效果，简单地说，RALF算法的优势体现在它的自适应调节能力，针对未来的负载变化提前进行缓存资源的预留，同时根据队列上溢概率动态调整各虚拟网络的优先级，并优先满足溢出概率高的虚拟网络的传输服务，并作出动态的调整。

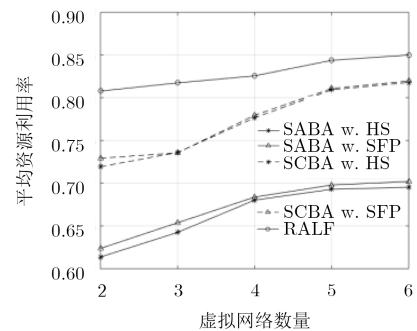


图 5 不同方案平均资源利用率

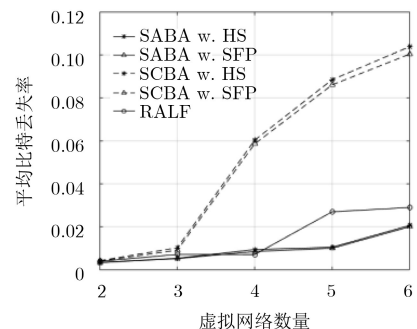
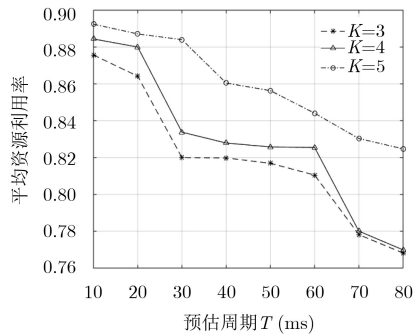
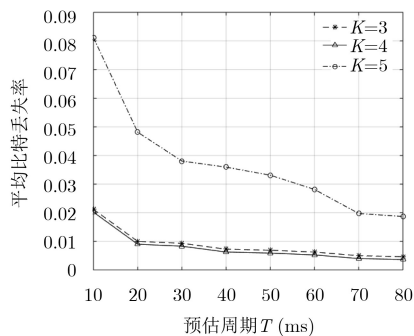


图 6 不同方案平均比特丢失率

进一步，本节设计了仿真实验研究短周期上队列上溢概率的预估周期 $T$ 对系统性能的影响。从图7, 图8可以观察到，无论虚拟网络数量为何值，随着预估周期 $T$ 的增加，系统的缓存资源与时频资源的加权平均利用率和平均比特丢失率在某种程度上都有所下降。且当虚拟网络数量 $K$ 为3, 4时，两种情况下的资源利用率和比特丢失率随预估周期 $T$ 的变化趋势很接近。而当虚拟网络数量 $K$ 增至5时，资源利用率显著上升，但是平均比特丢失率也随之增大。也由式(21)可知，预估周期越大，队列上溢概率的估计值将会越精确，所以为了最优地保障各虚拟网络的队列上溢概率约束，理论上应当选择较大的 $T$ 对其进行估计。然而从图7看出预估周期 $T$ 从10增至80时，系统的平均资源利用率大约会降低0.35%。因此，为了保证系统各方面的性能都

图7 不同 $T$ 对应的平均资源利用率图8 不同 $T$ 对应的平均比特丢失率

处于良好的状态, 需要合理地选取预估周期 $T$ 的值。

最后, 图9(a), 9(b), 9(c)分别刻画了3个虚拟网络在连续22个长周期上平均负载值与预测值的拟合效果。可以观察到, 每个虚拟网络的实际平均负载状态与预测的平均负载波动情况基本保持一致。

## 5 结束语

本文针对无线虚拟化网络中业务的不确定性和信息反馈的时延而引起虚拟资源分配不合理, 提出

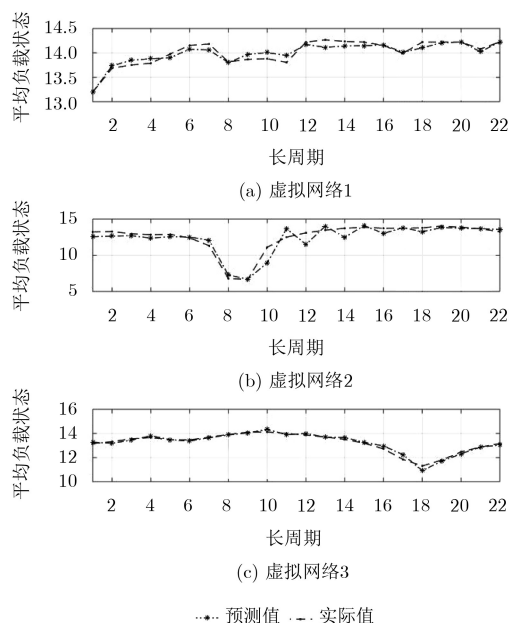


图9 各虚拟网络的平均负载实际值与预测值比较

了一种基于负载预测的动态虚拟资源分配算法。该算法将缓存资源和时频资源作为虚拟化的载体, 并考虑到不同资源的差异化特征, 设计了一种多时间尺度的混合调度机制, 以最小化资源租用成本为目标, 分别在长周期和短周期上对缓存资源和时频资源进行动态的调度。仿真结果表明, 本文算法可有效降低资源租用成本, 同时提升资源利用率。但是, 本文设计的缓存资源预留策略相对较为简单, 为了能更好地适应不同虚拟化应用场景的缓存需求, 后续工作将会对缓存资源调度进行更加精确的建模。

## 参考文献

- [1] AGIWAL M, ROY A, and SAXENA N. Next generation 5G wireless networks: A comprehensive survey[J]. *IEEE Communications Surveys & Tutorials*, 2017, 18(3): 1617–1655. doi: 10.1109/COMST.2016.2532458.
- [2] KALIL M, AL-DWEIK A, SHARKH M A, *et al.* A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks[J]. *IEEE Access*, 2017, 5: 20814–20827. doi: 10.1109/ACCESS.2017.2746666.
- [3] ZHANG Haijun, LIU Na, CHU Xiaoli, *et al.* Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges[J]. *IEEE Communications Magazine*, 2017, 55(8): 138–145. doi: 10.1109/MCOM.2017.1600940.
- [4] RAHMAN M M, DESPINS C, and AFFERS S. Design optimization of wireless access virtualization based on cost & QoS trade-Off utility maximization[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(9): 6146–6162. doi: 10.1109/TWC.2016.2580505.
- [5] SALLENTO O, PEREZ-ROMERO J, FERRUS R, *et al.* On radio access network slicing from a radio resource management perspective[J]. *IEEE Wireless Communications*, 2017, 24(5): 166–174. doi: 10.1109/MWC.2017.1600220WC.
- [6] JIANG Menglan, CONDOLUCI M, and MAHMOODI T. Network slicing management & prioritization in 5G mobile systems[C]. The 22th European Wireless Conference, Oulu, Finland, 2016: 1–6.
- [7] ZHU Qixuan and ZHANG Xi. Game-theory based buffer-space and transmission-rate allocations for optimal energy-efficiency over wireless virtual networks[C]. 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, USA, 2015: 1–6. doi: 10.1109/GLOCOM.2015.7417845.
- [8] AHMADI H, MACALUSO I, GOMEZ I, *et al.*

- Substitutability of spectrum and cloud-based antennas in virtualized wireless networks[J]. *IEEE Wireless Communications*, 2017, 24(2): 114–120. doi: [10.1109/MWC.2016.1500303WC](https://doi.org/10.1109/MWC.2016.1500303WC).
- [9] LEANH T, TRAN N, NGO D T, *et al.* Resource allocation for virtualized wireless networks with backhaul constraints[J]. *IEEE Communications Letters*, 2017, 21(1): 148–151. doi: [10.1109/LCOMM.2016.2617307](https://doi.org/10.1109/LCOMM.2016.2617307).
- [10] SCIANCALEPORE V, SAMDANIS K, COSTA-PEREZ X, *et al.* Mobile traffic forecasting for maximizing 5G network slicing resource utilization[C]. IEEE INFOCOM 2017-IEEE Conference on Computer Communications, Atlanta, USA, 2017: 1–9. doi: [10.1109/INFOCOM.2017.8057230](https://doi.org/10.1109/INFOCOM.2017.8057230).
- [11] CHU Yenming, HUANG Nenfang, and LIN Shengsiung. Quality of service provision in cloud-based storage system for multimedia delivery[J]. *IEEE Systems Journal*, 2014, 8(1): 292–303. doi: [10.1109/JSYST.2013.2257338](https://doi.org/10.1109/JSYST.2013.2257338).
- [12] AMIRI M and MOHAMMAD-KHANLI L. Survey on prediction models of applications for resources provisioning in cloud[J]. *Journal of Network & Computer Applications*, 2017, 82: 93–113.
- [13] 李捷, 刘先省, 韩志杰. 基于ARMA的无线传感器网络流量预测模型的研究[J]. 电子与信息学报, 2007, 29(5): 1224–1227. LI Jie, LIU Xianxing, and HAN Zhijie. Research on the ARMA based traffic prediction algorithm for wireless sensor network[J]. *Journal of Electronics & Information Technology*, 2007, 29(5): 1224–1227.
- [14] MANDJES M. Large Deviations for Gaussian Queues: Modelling Communication Networks[M]. Chichester: Wiley, 2007: 55–60.
- [15] DEMBO A and ZEITOUNI O. Large Deviations Techniques and Applications[M]. Berlin: Springer, 2010: 303–304.
- [16] YANG Jian, RAN Yongyi, CHEN Shuangwu, *et al.* Online source rate control for adaptive video streaming over HSPA and LTE-Style variable bit rate downlink channels[J]. *IEEE Transactions on Vehicular Technology*, 2016, 65(2): 643–657. doi: [10.1109/TVT.2015.2398515](https://doi.org/10.1109/TVT.2015.2398515).
- [17] GARDNER E Jr. Exponential smoothing: The state of the art—Part II[J]. *International Journal of Forecasting*, 2006, 22(4): 637–666. doi: [10.1016/j.ijforecast.2006.03.005](https://doi.org/10.1016/j.ijforecast.2006.03.005).
- 唐 伦: 男, 1973年生, 教授, 博士生导师, 主要研究方向为新一代无线通信网络、异构蜂窝网络、软件定义无线网络等。  
杨希希: 女, 1992年生, 硕士生, 研究方向为网络虚拟化。  
施颖洁: 女, 1993年生, 硕士生, 研究方向为网路切片。  
陈前斌: 男, 1967年生, 教授, 博士生导师, 主要研究方向为个人通信、多媒体信息处理与传输、下一代移动通信网络、异构蜂窝网络。