

# 分布式视频信息存储网络的任务均衡分配<sup>\*</sup>

陈进才 张江陵 冯丹

(华中科技大学计算机学院 外存储系统国家专业实验室 武汉 430074)

**摘要** 考虑将广义超级市场模型(Generalized Supermarket Model, GSM)应用于分布式视频信息存储网络的任务均衡分配。设:①分布式视频信息存储网络由在地理上分布的存储服务结点、存储负载均衡调度结点和客户结点组成;②视频信息存储任务的到达是一泊松过程,任务粒度为视频流文件;③存储任务的执行时间服从指数分布;④系统在初始时刻的存储任务有限;⑤与存储任务的执行时间相比,存储任务的调度时间不计。在概率意义下,深入讨论了基于广义超级市场模型的负载均衡调度与随机服务选择两种情况下的平均存储任务消耗时间,分析结果表明:在大规模分布式视频信息存储网络中,采用广义超级市场模型的存储任务均衡调度可显著提高系统的平均存储任务消耗时间性能。

**关键词** 分布式系统,视频信息,存储网络,任务均衡分配

## Task Distribution Balancing for Distributed Storage Networks of Video Information

CHEN Jin-Cai ZHANG Jiang-Ling FENG Dan

(National Storage System Laboratory, College of Computer Science & Technology, Huazhong University of Science & Technology, Wuhan 430074)

**Abstract** The Generalized Supermarket Model (GSM) is used to the task distribution balancing for distributed storage networks of video information. Suppose that: ①a storage network of video information consists of certain distributed storage service nodes, a storage load balance-scheduling node, and client nodes; ②the arrivals of video information storage tasks can be regarded as a Poisson process, and the storage task granularity is of video stream file; ③the executing time for a storage task is exponentially distributed; ④at the initial time, the number of storage tasks is finite; ⑤compared with the executing time used for storage tasks, the scheduling time used for storage tasks is neglected. In the sense of probability, the mean time consumption of storage tasks in two cases, using Generalized Supermarket Model based scheduling strategy and random service node choice based scheduling strategy, is analyzed in detail. The result shows that: for large-scale distributed storage networks of video information, the performance of mean time consumption for storage tasks can be increased evidently by using of the GSM-based load balance scheduling strategy.

**Keywords** Distributed systems, Video information, Storage networks, Task distribution balancing

## 1 引言

在分布式视频信息存储网络中,视频信息存储任务的到达和存储任务的执行时间均具有随机性。直观而言,对于随机到达的存储任务,如果不考虑系统当时的负载状态,只是随机选择服务结点来进行存储,可能造成分布式视频信息存储网络负载的不平衡,致使系统的平均存储任务消耗时间增大。相比之下,如果按照分布式视频信息存储网络的动态负载状态,在一定的存储负载平衡意义下合理地选择服务结点进行存储,可有效降低系统的平均存储任务时间。

基于上述分析,本文考虑将广义超级市场模型(Generalized Supermarket Model, GSM)应用于分布式视频信息存储网络的动态任务分配。在概率意义下,分析和比较了在广义超级市场模型的负载均衡调度与随机服务选择两种情况下的平均存储任务消耗时间特性。结果表明:对于大规模分布式视频信息存储网络,采用广义超级市场模型的负载均衡调度与随机选择服务相比,可指数级地减小平均存储任务消耗时间。此亦说明,在大规模分布式视频信息存储网络中,对于随机到达的存储任务,按照随机选择服务结点的方式进行存储易造成

分布式视频信息存储网络负载的不平衡,从而导致系统的平均存储任务消耗时间增大。反之,如果按照分布式视频信息存储网络的动态负载状态,运用广义超级市场模型进行结点选择判决,可有效改善系统的负载均衡水平,从而降低了系统的平均存储任务消耗时间,进而验证了前述直觉的正确性。

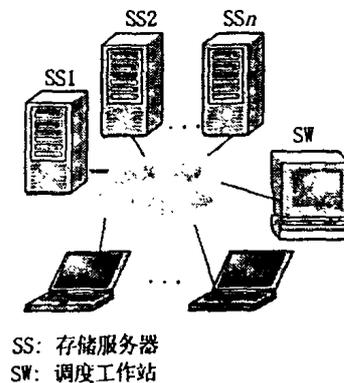


图1 视频信息存储网络

<sup>\*</sup> 本文得到国家自然科学基金(69973017)资助。陈进才 博士,副教授,主要研究方向:分布式系统、风格存储、智能代理、数字媒体技术;张江陵 教授,博士生导师,主要研究方向:计算机外存储系统、磁盘阵列、海量信息存储;冯丹 教授,博士生导师,主要研究方向:计算机外存储系统、磁盘阵列、海量信息存储。

## 2 系统的抽象服务模型

假设:①分布式视频信息存储网络由  $n$  个存储服务结点、1 个存储负载均衡调度结点及客户结点组成,如图 1 所示;②由客户提出视频信息存储请求,视频信息存储任务的到达是一泊松过程,速率为  $n\lambda$ ;③存储任务的执行时间服从参数为  $1/\mu$  的指数分布,  $\lambda < \mu$ , 任务粒度为视频流文件;④系统在初始时刻的存储任务有限;⑤与存储任务的执行时间相比,存储任务的调度时间不计(注意,我们所采用的任务粒度为视频流文件)。对于新到达的视频信息存储任务请求,经由存储负载均衡调度结点被重定向到合适的存储服务结点进行排队,等候任务的执行。显然,该系统等价于一个 M/M/n 服务模型<sup>[1]</sup>。

以下,详细分析在该 M/M/n 服务模型中引入基于 GSM 负载均衡调度后分布式视频信息存储网络的系统性能。

## 3 分布式视频信息存储网络的任务分配

### 3.1 预备知识

在文[2]中,Mitzenmacher M 博士提出了超级市场模型,用于描述超级市场的结算调度策略。该模型采用先选择部分服务器,再作进一步判优的 2 步调度机制。在此基础上,我们在文[3]中将该模型推广为:顾客按照速率为  $n\lambda$  的泊松分布到达收银区,收银区装有  $n$  台结算机,每台结算机旁配有相应的排队通道(通道容量无限);当顾客到达收银区时,首先观察  $n$  个队列的状态,从中选取最短者(观察、选择时间忽略不计),然后进入选定通道排队等候结算服务,结算机对于每个顾客的服务时间服从指数分布,均值为  $1/\mu$ ,其中  $\lambda < \mu$ 。称此模型为广义超级市场模型(Generalized Supermarket Model)。

设  $\chi_k(t)$  为系统在时刻  $t$  至少存在  $k$  个顾客的队列概率,则有

$$\begin{cases} \frac{1}{\mu} \frac{d\chi_k(t)}{dt} = \sigma(\chi_{k-1}(t) - \chi_k(t)) - (\chi_k(t) - \chi_{k+1}(t)), & k \geq 1, \\ \chi_0(t) \equiv 1, \\ \sum_{k=1}^{\infty} \chi_k(t) < \infty \end{cases} \quad (1)$$

这里,  $\sigma = \lambda/\mu < 1$ 。

以下,对于广义超级市场模型与式(1)不加区别。可以证明:

**命题 1(稳定性)** 广义超级市场模型是系统稳定的,即,若系统在初始时刻是顾客有限的,则在任意时刻的期望顾客数亦是有限的。

**命题 2(不动点的存在唯一性)** 广义超级市场模型存在唯一的不动点解。

$$\chi_k = \sigma^{k(k-1)/(n-1)}, k=0,1,2,\dots$$

**命题 3(解的收敛性)** 设广义超级市场模型在初始时刻队列为空,则其解收敛于系统的不动点。

关于广义超级市场模型,我们在文[3]中有详细论述,在此不予赘述。

### 3.2 动态存储负载均衡策略

根据第 2 节的假设,分布式视频信息存储网络符合广义超级市场模型的任务到达与服务特性,所以可将广义超级市场模型的调度策略应用于分布实现视频信息存储网络的存储负载均衡中,即,在分布式视频信息存储网络中,每个存储服务结点上设置一个等待该结点服务的存储任务队列;在存储负

载均衡调度结点中,维持与存储服务结点相对应的  $n$  维动态变化的存储任务队列长度信息向量;如果有一个新的视频信息存储任务到达,存储负载均衡调度结点首先对其进行队列分配,分配的原则是采用最短任务调度算法(Shortest Queue Routing, SQR)<sup>[4]</sup>,即选择具有最短长度的存储任务队列,任务在选中的存储队列中排队,等待对应存储服务结点的服务;当某存储服务结点的当前存储任务完成以后,从与其对应的存储任务队列中按照先进先出(FIFO)的原则得到最先进入的等待服务的存储任务,交由存储服务结点进行处理,该存储任务即从存储任务队列中移出,新的存储任务队列长度信息及时反映在存储负载均衡调度结点的  $n$  维存储任务队列长度向量中,如图 2 所示。

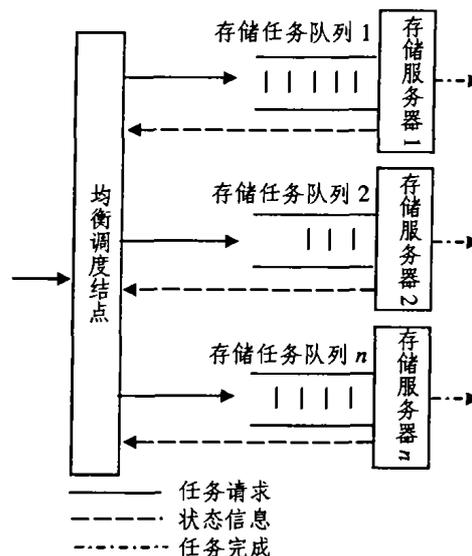


图 2 存储任务分配

显然,该存储负载均衡调度策略与广义超级市场模型是等价的,因此,式(1)和命题 1~命题 3 在分布式视频信息存储网络系统中依然成立。这里,  $\chi_k(t)$  对应于分布式视频信息存储网络在时刻  $t$  至少存在  $k$  个存储任务的队列的概率,广义超级市场模型中的顾客对应于分布式视频信息存储网络中的存储任务。

### 3.3 平均存储任务消耗时间

对于一个新到达的存储任务,它在系统中的消耗时间是随机变化的,其大小取决于系统在其进入时刻到服务完毕间的状态。然而,对于平均存储任务消耗时间,我们有:

**定理 1(平均存储任务消耗时间)** 在分布式视频信息存储网络中,假设:①系统采用广义超级市场模型的存储负载均衡调度策略;②系统在初始时刻所有存储服务结点的队列为空。则系统的平均存储任务消耗时间趋向于

$$T_s = \sum_{k=1}^{\infty} \sigma^{k(k-1)/(n-1)} / \mu$$

证明:很显然,存储任务在系统中的消耗时间包括:①在存储任务队列中的等待时间;②该任务的实际执行时间。存储任务在队列中等待时间均值取决于队列当时的长度。若队列的长度为  $k$ ,则该任务在相应队列中等待时间为  $k-1$  个存储任务的执行时间,加上本身所消耗的平均存储执行时间,即为存储任务在存储队列长度为  $k$  的情况下的总平均消耗时间:  $k/\mu$ 。故,存储任务的平均消耗时间为:

$$\begin{aligned} T(t) &= \sum_{k=1}^{\infty} k P(\text{队列为 } k) / \mu \\ &= \sum_{k=1}^{\infty} k (\chi_{k-1}^2(t) - \chi_k^2(t)) / \mu \end{aligned}$$

$$= \sum_{k=0}^{\infty} x_k^*(t) / \mu$$

由命题 3 知, 当  $t \rightarrow \infty$  时

$$\chi_k(t) \rightarrow \sigma^{(k-1)/(n-1)}, \text{ 所以, 当 } t \rightarrow \infty \text{ 时, } T(t) \rightarrow 1/\mu \sum_{k=0}^{\infty} \sigma^{k-1}$$

$$\sigma^{\frac{k-1}{n-1}} \cdot n = 1/\mu \sum_{k=1}^{\infty} \sigma^{\frac{k-n}{n-1}} \quad \square$$

显然, 在不考虑存储任务队列长度而按均匀分布随机选择存储任务队列的情况下, 系统整体所遵从的 M/M/n 服务模型便可分解为  $n$  个等价的 M/M/1 服务模型, 两者的单任务(这里为元任务, 元任务不可分解)执行时间指数分布参数一致, 皆为  $\mu$ 。所不同的是, 原 M/M/n 服务模型的任务到达速率为  $n\lambda$ , 而这里的 M/M/1 服务模型的任务到达速率为  $\lambda$ 。由 M/M/1 服务模型知, 任务的平均消耗时间为  $1/(\mu - \lambda)$ 。因此, 在不考虑存储任务队列长度而按均匀分布随机选择存储任务队列的情况下, 平均存储任务消耗时间为  $T_r = \frac{1}{\mu - \lambda}$ , 从而有:

**定理 2(均衡与随机调度间的关系)** 假设分布式视频信息存储网络满足定理 1 的条件, 则当  $\lambda \rightarrow \mu$  时

$$T_r = O(\log T_r)$$

证明:

当  $\lambda \rightarrow \mu$  时,  $\sigma = \lambda/\mu \rightarrow 1^-$ ,

$$\begin{aligned} \frac{T_r}{\log T_r} &= \frac{1/\mu \sum_{k=1}^{\infty} \sigma^{\frac{k-n}{n-1}}}{\log \frac{1}{\mu - \lambda}} = \frac{\sum_{k=1}^{\infty} \sigma^{\frac{k-n}{n-1}}}{\mu \log \frac{1}{\mu - \lambda}} \\ &= \frac{\sum_{k=1}^{\infty} \sigma^{\frac{k-n}{n-1}}}{\mu (\log \frac{1}{\mu} + \log \frac{1}{1-\sigma})} \sim \frac{\sum_{k=1}^{\infty} \sigma^{\frac{k-n}{n-1}}}{\mu \log \frac{1}{1-\sigma}} \\ &= \frac{\sum_{k=1}^{\infty} \sigma^{\frac{k-n}{n-1}}}{\mu \log \frac{1}{1-\sigma}} \sim \frac{\sum_{k=1}^{\infty} (\sigma^{\frac{1}{n-1}})^k \cdot \log(1 - \sigma^{\frac{1}{n-1}})}{-\mu \log(1 - \sigma^{\frac{1}{n-1}}) \cdot \log(1 - \sigma)} \end{aligned}$$

又

$$\begin{aligned} \frac{\log(1 - \sigma^{\frac{1}{n-1}})}{\log(1 - \sigma)} &\sim \frac{\frac{1}{n-1} \cdot \frac{1}{\sigma^{\frac{1}{n-1}-1}}}{\frac{1}{1-\sigma}} \\ &\sim \frac{1}{n-1} \cdot \frac{1-\sigma}{1 - \sigma^{\frac{1}{n-1}}} \sim \frac{1}{n-1} \cdot \frac{n-1}{\sigma^{\frac{1}{n-1}-1}} \rightarrow 1 \end{aligned}$$

令  $\sigma' = \sigma^{\frac{1}{n-1}}$ , 则

$$\frac{\sum_{k=1}^{\infty} (\sigma^{\frac{1}{n-1}})^k}{-\mu \log(1 - \sigma^{\frac{1}{n-1}})} = \frac{\sum_{k=1}^{\infty} (\sigma')^k}{-\mu \log(1 - \sigma')} \sim \frac{\sum_{k=1}^{\infty} \sigma^k}{-\mu \log(1 - \sigma)}$$

又  $\sigma^k$  非负且关于  $x$  单调递减, 故

$$\int_1^{N+1} \sigma^x dx \leq \sum_{k=1}^N \sigma^k \leq \int_0^N \sigma^x dx,$$

$$\int_1^{\infty} \sigma^x dx \leq \sum_{k=1}^{\infty} \sigma^k \leq \int_0^{\infty} \sigma^x dx$$

$$\leq 1 + \int_1^{\infty} \sigma^x dx,$$

$$\sum_{k=1}^{\infty} \sigma^k = \int_1^{\infty} \sigma^x dx + \alpha, \alpha \leq 1$$

令  $x' = n^x$ , 则

$$\int_1^{\infty} \sigma^x dx = \frac{1}{\log n} \int_1^{\infty} \frac{\sigma^{x'}}{x'} dx'$$

这里,  $\int_1^{\infty} \frac{\sigma^{x'}}{x'} dx'$  是第二类欧拉积分, 不能用初等函数表示, 而

$$\sum_{k=n+1}^{\infty} \frac{\sigma^k}{k} \leq \int_1^{\infty} \frac{\sigma^{x'}}{x'} dx' \leq \sum_{k=n}^{\infty} \frac{\sigma^k}{k}$$

$$\sum_{k=1}^{\infty} \frac{\sigma^k}{k} \leq \int_1^{\infty} \frac{\sigma^{x'}}{x'} dx' + \sum_{k=1}^n \frac{\sigma^k}{k} \leq \sum_{k=1}^{\infty} \frac{\sigma^k}{k} + \frac{\sigma^n}{n}$$

令  $f(\sigma) = \sum_{k=1}^{\infty} \frac{\sigma^k}{k}$ , 则  $f(\sigma)$  收敛且

$$f'(\sigma) = \sum_{k=0}^{\infty} \sigma^k = \frac{1}{1-\sigma},$$

$$f(\sigma) = -\log(1 - \sigma) + c$$

由  $f(0) = 0$  知,  $c = 0$

故

$$\int_1^{\infty} \frac{\sigma^{x'}}{x'} dx' + \sum_{k=1}^n \frac{\sigma^k}{k} = -\log(1 - \sigma) + \frac{\sigma^n}{n} \cdot \beta, \beta \leq 1,$$

$$\frac{\sum_{k=1}^{\infty} \sigma^k}{-\mu \log(1 - \sigma)}$$

$$\sim \frac{1}{\log n} (-\log(1 - \sigma) + \frac{\sigma^n}{n} \cdot \beta - \sum_{k=1}^n \frac{\sigma^k}{k}) + \alpha$$

$$\sim \frac{1}{\mu \log n}$$

所以,  $T_r = O(\log T_r)$  □

**结论** 本文详细讨论了将广义超级市场模型应用于分布式视频信息存储网络的动态任务均衡分配。假设分布式视频信息存储网络由  $n$  个存储服务结点、一个存储负载均衡调度结点和客户结点组成, 每个视频信息存储服务结点设置一个存储任务队列, 队列中的存储任务按照先进先出(FIFO)原则顺序执行, 任务完成后存储负载均衡调度结点及时收集新的任务队列长度信息并更新队列长度向量; 负载均衡调度结点对新到达的存储任务请求按照最短队列优先的方式重定向到存储服务结点并排队等待存储服务。结果表明: ①采用广义超级市场负载均衡调度策略的分布式视频信息存储网络是系统稳定的; ②与随机选择存储服务结点的方式相比, 在初始时刻所有存储任务队列为空的情况下, 按照广义超级市场模型进行动态负载均衡调度所得存储任务的平均消耗时间是指数下降的。

### 参 考 文 献

- 1 Robertazzi T G. Computer Networks and Systems: Queueing Theory and Performance Evaluation [M]. Berlin: Springer-Verlag, 1990
- 2 Mitzenmacher M. The power of two choices in randomized load balancing [J]. IEEE Trans. on Parallel and Distributed Systems, 2001, 12(10): 1094~1104
- 3 陈进才, 张江陵, 冯丹. GSM: 一种多服务器系统的动态负载均衡调度模型[J]. 小型微型计算机系统, 已录用, 待刊
- 4 林闯. 随机 Petri 网和系统性能评价[M]. 北京: 清华大学出版社, 2000