doi:10.11959/j.issn.1000-436x.2015077

Cloud-P2P 云存储结构的模型建立与性能分析

金顺福[1,2]，王晨飞[1,2]，陈玲玲[3]，霍占强[4]

(1. 燕山大学 信息科学与工程学院，河北 秦皇岛 066004；2. 河北省计算机虚拟技术与系统集成重点实验室，河北 秦皇岛 066004；
3. 中国联合网络通信有限公司秦皇岛市分公司，河北 秦皇岛 066000；4. 河南理工大学 计算机科学与技术学院，河南 焦作 454000)

摘 要：基于 Cloud-P2P 云存储结构，针对云中心和 P2P 节点存储层数据副本的访问机制，考虑节点存储层数据副本的修复过程，建立一个三维连续时间 Markov 链模型。使用矩阵几何解方法导出该模型的稳态解，并给出节点存储层传输率，数据访问延迟和副本修复率等系统性能指标的表达式。通过数值实验和系统仿真定量刻画数据副本数等系统参数对 Cloud-P2P 云存储结构性能的影响。构造利润函数，进行用户存储层副本数的优化设置。

关键词：云存储；P2P；三维 Markov 链；矩阵几何解；性能指标

中图分类号：TP393 **文献标识码**：A

Modeling and analysis of Cloud-P2P storage architecture

JIN Shun-fu[1,2], WANG Chen-fei[1,2], CHEN Ling-ling[3], HUO Zhan-qiang[4]

(1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China;
2. The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004, China;
3. The China United Network Communications Co., Ltd., Qinhuangdao Branch, Qinhuangdao 066000, China;
4. College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: Based on the Cloud-P2P storage architecture, taking into account the data access mechanism with both the data center and the peer-based layer, considering the repair process of the replica, a three-dimensional Markov chain model was constructed. The steady-state analysis of the model is obtained by using a matrix-geometric method. Then, the performance measures in terms of peer-based traffic load, the average access time, and the repair rate are given. Moreover, numerical results with analysis and simulation are provided to demonstrate the influence of the system parameters on the system performance. By constructing a benefit function, the number of replicas in peer-based layer is optimized.

Key words: cloud storage; P2P; three-dimensional Markov chain; matrix-geometric; performance measures

1 引言

云存储是在云计算的概念上延伸出来的概念。云存储系统通过集群应用、网格技术或分布式文件系统等功能，将网络中大量不同类型的存储设备通过应用软件集合起来协同工作，对外提供数据存储和业务访问功能[1]。在面对如今的海量数据应用结合高速互联网服务，云存储所表现出的巨大优势使其被工业界和学术界认为是未来最富有前景的应用之一。

在云存储环境中，具有相似或相同喜好的用户在云端存储的数据具有很高的相似度，甚至存储相同的数据。随着使用云存储服务用户数量的不断增加，大量相同的数据将会存储在云端服务器上，这将对云端存储资源造成不必要的浪费，因此需要对云端数据进行合理的管理[2]。

另一方面，如果有大量用户访问云存储端的一个数据节点，而其他数据节点处于闲置状态，将会造成通信和存储资源的极大浪费，同时，用户下载数据所花费的时间也会延长。因此云节点间的负载

收稿日期：2013-12-13；**修回日期**：2014-01-19
基金项目：国家自然科学基金资助项目（11201408, 61472342）；河北省自然科学基金资助项目（2012203093）
Foundation Items: The National Natural Science Foundation of China (11201408, 61472342); The Natural Science Foundation of Hebei Province (2012203093)

均衡就变得十分必要。针对云存储系统中数据重复问题，Google 公司提出了 MapReduce 模型[3]。由于 MapReduce 模型对重复数据的检测基于数据集和中间结果缓存的大小，文献[4]提出了一种带有修剪策略的签名策略，缩减了数据集和中间结果缓存的大小，提高了数据检测的精确度，减少了有效数据的丢失率。文献[5]提出了一种基于时间序列的预测算法，云存储系统使用该算法对云节点的负载进行预测，并将负载压力分配给其他云节点，有效地实现了云节点之间的负载均衡。在文献[6]提出了一个新的数据管理模型，将数据分成若干个子块，并将每个子块的名称列表存储在索引名称服务器 INS(index name server)上，当用户进行数据上传时，首先与 INS 上的索引数据进行比对，再确定是否上传数据，有效地减少了云端的数据冗余。当用户进行数据访问时，INS 根据用户的数据传输速率，为用户合理地分配云端节点，保证了节点间的负载均衡，提高了带宽和存储资源的利用率。

众所周知，数据可靠性是评价一个存储系统性能优劣的重要指标之一，如何保证云存储系统的可靠性也成为国内外学者研究的重要课题。文献[7]提出了一种副本主动检查(proactive replica checking)模型。模型主要由 2 部分构成，一部分为 PRCR 节点，主要用于副本的管理；另一部分为用户接口，主要为请求数据的用户分配 PRCR 节点，通过动态地检查数据的可用性，以保证云存储系统的高可靠性。文献[8]从资源分配角度，考虑云存储环境中的资源优先级、QoS 及存储空间等因素对系统可靠性的影响，建立了一个云存储系统可靠性的分析模型，将分析模型运用到现有的云存储系统中，对模型的有效性进行了验证。文献[9]根据不同地点用户访问同一云节点数据频率的高低，提出一种副本备份算法，提高了数据的可靠性，减少了用户数据的访问时间。

由于云存储服务一般由服务器集群组成，在今后的发展中将面临着 2 方面的挑战[10]，一方面使系统搭建费用和管理成本都集中在服务商，另一方面由于断电，自然灾害及人为攻击等因素增加了数据中心的风险性。这种云存储服务在遇到灾难性事件后很可能造成数据丢失，进而导致严重的后果。为此，一些研究者开始使用 P2P 技术作为一种补充加入到云存储结构中[11~13]。P2P 网络强调去中心化，弱化了服务器的概念。这类网络成本低，容灾性好，可扩展性强，且具有较强的负载均衡能力。因此将闲散的用户存储空间以 P2P 的形式组织起来，构成节点存储层，并加入到由服务商提供的云存储服务器集群中，形成一个 Cloud-P2P 云存储结构[14]，是一种较为成功的实践。

文献[15]提出了一种采用 P2P 技术构建的云存储服务系统 AmazingStore。利用网络中用户节点闲置的带宽和存储资源，辅助云中心节点工作，降低了云中心节点的压力，描述了 AmazingStore 系统中数据的存储形式，给出了维护可用副本数的计算公式，并验证了 AmazingStore 的高可靠性和高下载速度。文献[16]针对 Cloud-P2P 的内容分发网络中 Peer 节点和云端节点的带宽最优分配问题，建立了一个细粒度性能模型(OBPA)，使用快速收敛迭代算法对模型进行了解析，并对模型的有效性进行了分析。文献[17]建立了一个 P2CP(peer to cloud and peer)模型，并引入了分等级的安全机制以保证数据的安全性，不但从理论上证明了在传输速度上 P2CP 优于传统 P2P，并且从实验角度验证了 P2CP 的数据可用性和数据安全性。文献[18]针对如何充分利用 Cloud-P2P 中移动节点的带宽资源问题，提出了一个叫做 NCTLD 机制。NCTLD 机制包括 2 个算法，即 NCS 算法和 DLT 算法。NCS 算法帮助移动节点找到可以提供服务的对等节点，DLT 算法保证了移动节点拥有足够的下载速度。NCTLD 机制改善了移动节点的下载速度，缓解了云中心的带宽压力。

在已有的研究成果中，针对 Cloud-P2P 云存储系统的研究主要集中在资源分配算法、数据的可靠性和安全性等方面，并没有考虑到 Cloud-P2P 网络中副本数量对云存储系统性能的影响。在 Cloud-P2P 云存储系统中，节点存储层的数据副本数是十分重要的系统参数[19]。节点存储层副本节点的数量越多，响应用户数据请求的节点就越多，云中心服务器接收到的用户数据请求就越少，数据的传输速度就越快。另一方面，节点存储层的副本节点越多，云中心所要监测的副本节点也就越多，跟踪服务器消耗的网络带宽和存储资源消就越大。

为此，本文考虑 Cloud-P2P 云存储网络中副本数量对云存储系统性能的影响，结合 Cloud-P2P 云存储系统中用户访问数据的机制及节点存储层副本节点的修复过程，建立了一个三维连续时间的 Markov 链，使用矩阵几何解方法给出系统性能指标的表达式。构建利润函数，通过数值实验对单个数

据副本数进行优化设置。

2　Cloud-P2P 云存储结构及系统模型

2.1　Cloud-P2P 云存储结构分析

Cloud-P2P 云存储结构包括节点存储层和云中心 2 部分。节点存储层主要由运行在用户节点上的客户端程序和用户的共享存储空间组成，当云中心备份服务器发生故障时，节点存储层所提供的副本备份仍可以满足用户的使用。

云中心由服务提供商的服务器集群组成，主要有以下 2 个功能，一是对用户上传的数据进行备份并提供下载服务，一是对节点存储层上的副本进行组织和管理。云服务器采用定时器机制对节点存储层副本节点的在线情况进行周期性地监控，当节点存储层中一个数据实例的所有副本节点都不在线时，云中心响应用户数据请求并提供服务。

在 Cloud-P2P 云存储系统中，将云存储技术和 P2P 技术结合起来，使云中心和节点存储层的功能相互补充，提高了用户数据的可靠性。另一方面，部分数据访问任务由节点存储层提供，使云中心服务器的压力得到了有效的缓解，Cloud-P2P 云存储结构如图 1 所示。

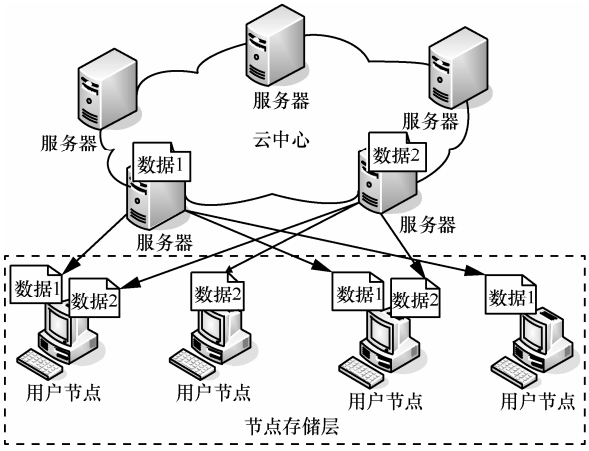

图 1　Cloud-P2P 云存储结构

构架 Cloud-P2P 云存储系统需要考虑以下 2 个方面的内容，一是如何对节点存储层中的数据副本进行有效地组织和管理，二是如何把用户的数据请求合理地分配给云中心和节点存储层的数据节点。

针对以上问题，在 Cloud-P2P 云存储系统中，云中心服务器和节点存储层的所有的副本节点按数据内容划分成不同的数据实例，每个数据实例具有唯一的实例 ID。处于云端的跟踪服务器存有每个数据实例对应的用户存储层副本节点的地址信息，并使用定时器机制对副本节点的在线情况进行周期性的监测。当用户请求某一数据实例时，处于云中心的跟踪服务器首先在节点存储层查找是否存在在线的副本节点。如果有副本节点在线，则将该用户请求交给副本节点处理，否则交给云中心进行处理。这种调度方式提高了节点存储层副本节点的资源利用率，并降低了云中心的带宽消耗，有效地减轻了云端服务器的数据传输压力。

2.2　系统模型建立

在 Cloud-P2P 云存储结构中，假设不同数据实例中的数据相互独立，节点存储层针对一个数据实例维护副本的数量为 H 。设定副本的修复阈值为 m ，当系统中在线副本数小于或等于 $H-m$ 时，系统开始对副本进行修复。

假设在 t 时刻，节点存储层的副本可能处于正在修复、等待修复和可以使用 3 个状态之一。令 $n_c(t)=i$ ， $i\in\{0,1,2,\cdots\}$ 表示 t 时刻用户向 Cloud-P2P 云存储系统发送数据请求的个数； $n_s(t)=j$ ， $j\in\{0,1,2,\cdots,H\}$ 表示 t 时刻节点存储层可以使用的副本个数； $n_r(t)=k$ ， $k\in\{0,1,2,\cdots\}$ 表示 t 时刻正在修复的副本个数。那么 $\{n_c(t),n_s(t),n_r(t)\}$ 构成了一个三维连续时间随机过程，其状态空间为 $\Omega=\{(i,j,k)\mid i\in\{0,1,2,\cdots\},(j,k)\in\{(0,H),(1,H-1),\cdots,(H-m,m),(H-m+1,0),(H-m+1,1),\cdots,(H-m+1,m-1),\cdots,(H-1,0),(H-1,1),(H,0)\}\}$ 。

假设节点存储层中所有副本的在线时间相互独立并都服从指数分布，在线时间参数用 $\mu_s(\mu_s>0)$ 表示。节点存储层上所有副本节点的修复时间相互独立并服从指数分布，副本节点修复参数表示为 $\lambda_s(\lambda_s>0)$ 。数据请求的到达率服从 $\lambda_c(\lambda_c>0)$ 的指数分布。云中心服务器对处理用户数据请求的服务率服从指数分布，服务率表示为 $\mu_{c1}(\mu_{c1}>0)$ 。节点存储层单个副本节点处理用户数据请求的服务率服从指数分布，服务率用 $\mu_{c2}(\mu_{c2}>0)$ 表示。显然， $\{n_c(t),n_s(t),n_r(t)\}$ 是一个三维连续时间 Markov 链。

3　模型解析与性能指标

3.1　转移率矩阵

如果当前时刻有 i 个用户向 Cloud-P2P 系统发送数据请求，则称当前时刻系统处于 i 水平。考虑在水平 i 下，系统节点存储层可用副本数 j 的变化过程如下。

1) 设当前时刻系统中可用副本的个数为 $j(0 \leqslant j \leqslant H-m)$，下一时刻系统中可用副本的个数可能由于副本节点离线而减少一个，转移率为 $j\mu_s$，也可能由于系统完成一个副本的修复使可用副本数增加一个，转移率为 $(H-j)\lambda_s$，还可能由于没有节点离线并且修复未完成而保持不变，转移率为 $-j\mu_s-(H-j)\lambda_s$。用 $\boldsymbol{T}_0$ 表示下一时刻可用副本数 $j(0 \leqslant j \leqslant H-m)$ 的状态转移率矩阵，则 $\boldsymbol{T}_0$ 可以写成一个三对角的 $(H-m+1)$ 阶方阵如式(1)所示。

$$\boldsymbol{T}_0=\begin{pmatrix} -H\lambda_s & H\lambda_s & & & \\ \mu_s & -\mu_s-(H-1)\lambda_s & (H-1)\lambda_s & & \\ & \ddots & \ddots & \ddots & \\ & & (H-m-1)\mu_s & -(H-m-1)\mu_s-(m+1)\lambda_s & (m+1)\lambda_s \\ & & & (H-m)\mu_s & -(H-m)\mu_s-m\mu_s \end{pmatrix} \tag{1}$$

设当前时刻系统中可用副本的个数为 $j=H-m$，此时有 m 个副本正处于正在修复状态。如果下一时刻系统中可用的副本数增加一个，说明有一个副本修复完成。令 $\boldsymbol{T}_1$ 表示系统中可用副本的个数由 $j=H-m$ 转移至 $j=H-m+1$ 的转移率矩阵，则 $\boldsymbol{T}_1$ 为一个 $(H-m+1)\times m$ 矩阵，如式(2)所示。

$$\boldsymbol{T}_1=\begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & m\lambda_s \end{pmatrix} \tag{2}$$

2) 设当前时刻系统中可用副本的个数为 $j=H-m+1$，如果下一时刻系统中可用副本的个数由于副本节点发生离线而减少一个。令 $\boldsymbol{T}_2$ 表示可用副本的个数由 $j=H-m+1$ 转移至 $j=H-m$ 的转移率矩阵，则 $\boldsymbol{T}_2$ 可以表示成 $m\times(H-m+1)$ 矩阵，如式(3)所示。

$$\boldsymbol{T}_2=\begin{pmatrix} 0 & \cdots & 0 & \cdots & (H-m+1)\mu_s \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & (H-m+1)\mu_s \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & (H-m+1)\mu_s \end{pmatrix} \tag{3}$$

3) 设当前时刻系统中可用副本的个数为 $j(H-m+1 \leqslant j \leqslant H)$，如果下一时刻有一个副本节点修复完成使系统中可用副本的个数增加一个。令 $\boldsymbol{P}_{j,j+1}$ 表示系统中可用副本的个数由 j 转移至(j+1)的转移率矩阵，则 $\boldsymbol{P}_{j,j+1}$ 可以表示成一个 $(H-j+1)\times(H-j)$ 矩阵，如式(4)所示。

$$\boldsymbol{P}_{j,j+1}=\begin{pmatrix} 0 & \cdots & 0 \\ \lambda_s & & \\ & \ddots & \\ & & (H-j)\lambda_s \end{pmatrix} \tag{4}$$

如果下一时刻系统中可用副本的个数保持不变，说明既没有副本节点发生离线，也没有副本修复完成。令 $\boldsymbol{P}_{j,j}$ 表示系统中可用副本的个数保持不变的转移率矩阵，则 $\boldsymbol{P}_{j,j}$ 可以表示成 $(H-j+1)$ 阶方阵，如式(5)所示。

$$\boldsymbol{P}_{j,j}=-diag(j\mu_s)-diag(0,\lambda_s,2\lambda_s,\cdots,(H-j)\lambda_s) \tag{5}$$

其中，$diag(x)$ 表示对角线元素为 x 的对角阵。

4) 设当前时刻系统中可用副本的个数为 $j(H-m+2 \leqslant j \leqslant H)$，如果下一时刻系统中可用副本的个数由于副本节点发生离线而减少一个，令 $\boldsymbol{P}_{j,j\text{-}1}$ 表示系统中可用的副本数由 j 转移到(j-1)的转移率矩阵，则 $\boldsymbol{P}_{j,j\text{-}1}$ 可以表示成 $(H-j)\times(H-j+1)$ 的矩阵，如式(6)所示。

$$\boldsymbol{P}_{j,j-1}=\begin{pmatrix} j\mu_s & & 0 \\ & \ddots & \vdots \\ & & j\mu_s \quad 0 \end{pmatrix} \tag{6}$$

令 $\boldsymbol{T}$ 表示同一水平下系统中可用副本数的转移率矩阵。综合式(1)~式(6)，可以把矩阵 $\boldsymbol{T}$ 写成一个分块三对角形式的 $((m+1)m/2+H-m+1)$ 阶方阵，如式(7)所示。

$$\boldsymbol{T}=\begin{pmatrix} \boldsymbol{T}_0 & \boldsymbol{T}_1 & & & \\ \boldsymbol{T}_2 & \boldsymbol{P}_{(H-m+1),(H-m+1)} & \boldsymbol{P}_{(H-m+1),(H-m+2)} & & \\ & \boldsymbol{P}_{(H-m+2),(H-m+1)} & \boldsymbol{P}_{(H-m+2),(H-m+2)} & \boldsymbol{P}_{(H-m+2),(H-m+3)} & \\ & & \ddots & \ddots & \ddots \\ & & & \boldsymbol{P}_{(H-1),(H-2)} & \boldsymbol{P}_{(H-1),(H-1)} \quad \boldsymbol{P}_{(H-1),H} \\ & & & & \boldsymbol{P}_{H,(H-1)} \quad \boldsymbol{P}_{H,H} \end{pmatrix} \tag{7}$$

令 $\boldsymbol{B}_i$ 表示系统由 i 水平到 $i-1$ 水平的转移率矩阵，$\boldsymbol{B}_i$ 可表示成式(8)所示的形式。

$$\boldsymbol{B}_i=\begin{cases}diag(\mu_{c2},\min(i,j)\mu_{c1},\cdots,\min(i,j)\mu_{c1}), & 1\leqslant i<H\\ diag(\mu_{c2},\mu_{c1},2\mu_{c1},\cdots,H\mu_{c1}), & i\geqslant H\end{cases} \tag{8}$$

由式(8)可知，当 $i\geqslant H$ 时，$\boldsymbol{B}_i$ 中的元素值与 i 的取值无关。当 $i\geqslant H$ 时，可将 $\boldsymbol{B}_i$ 用 $\boldsymbol{B}$ 表示。

令 $\boldsymbol{C}$ 表示系统从 i 水平转移到 $i+1$ 水平的转移率矩阵，$\boldsymbol{C}$ 可以表示成式(9)所示的形式。

$$\boldsymbol{C}=diag(\lambda_c,\cdots,\lambda_c) \tag{9}$$

令 $\boldsymbol{A}_i$ 表示系统由 i 水平到 i 水平的转移率矩阵。对应于不同的水平 i 和可用副本的个数 j，矩阵 $\boldsymbol{A}_i$ 如表 1 所示。

表 1　　矩阵 $\boldsymbol{A}_i$

(i,j)	$\boldsymbol{A}_i$
$i=0$	$diag(-\lambda_c)+\boldsymbol{T}$
$1\leqslant i\leqslant H-1, j=0$	$diag(-\lambda_c-\mu_{c2})+\boldsymbol{T}$
$1\leqslant i\leqslant H-1, 1\leqslant j\leqslant H$	$diag(-\lambda_c-\min(i,j)\mu_{c1})+\boldsymbol{T}$
$i\geqslant H, j=0$	$diag(-\lambda_c-\mu_{c2})+\boldsymbol{T}$
$i\geqslant H, 1\leqslant j\leqslant H$	$diag(-\lambda_c-j\mu_{c1})+\boldsymbol{T}$

由表 1 可知，当 $i\geqslant H$ 时，$\boldsymbol{A}_i$ 中的元素值与 i 的取值无关。当 $i\geqslant H$ 时，将 $\boldsymbol{A}_i$ 表示为 $\boldsymbol{A}$。

令 $\boldsymbol{Q}$ 表示三维连续时间 Markov 链 $\{n_c(t),n_s(t),n_r(t)\}$ 的转移率矩阵，综合式(8)和式(9)及表 1，$\boldsymbol{Q}$ 可以表示为如下分块三对角形式

$$\boldsymbol{Q}=\begin{pmatrix}\boldsymbol{A}_0 & \boldsymbol{C} & & & & & \\ \boldsymbol{B}_1 & \boldsymbol{A}_1 & \boldsymbol{C} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \boldsymbol{B}_{H-1} & \boldsymbol{A}_{H-1} & \boldsymbol{C} & & \\ & & & \boldsymbol{B} & \boldsymbol{A} & \boldsymbol{C} & \\ & & & & \ddots & \ddots & \ddots\end{pmatrix} \tag{10}$$

3.2　稳态分布

定义三维连续时间 Markov 链 $\{n_c(t),n_s(t),n_r(t)\}$ 的稳态分布为 $\pi_{i,j,k}$，则有

$$\pi_{i,j,k}=\lim_{t\to\infty}P\{n_c(t)=i,n_s(t)=j,n_r(t)=k\} \tag{11}$$

设稳态下系统处于 i 水平的概率向量为 $\boldsymbol{\pi}_i$，如式(12)所示。

$$\boldsymbol{\pi}_i=(\pi_{i,0,H},\pi_{i,1,(H-1)},\ldots,\pi_{i,(H-m),m},\pi_{i,(H-m+1),0},\pi_{i,(H-m+1),1},\ldots,\pi_{i,(H-m+1),(m-1)},\ldots,\pi_{i,(H-1),0},\pi_{i,(H-1),1},\pi_{i,H,0}) \tag{12}$$

系统稳态向量 $\boldsymbol{\pi}$ 可由 $\boldsymbol{\pi}_i$ 构成，表示如下

$$\boldsymbol{\pi}=(\boldsymbol{\pi}_0,\boldsymbol{\pi}_1,\boldsymbol{\pi}_2,\ldots) \tag{13}$$

特别地，当系统中可用副本的个数为 $H=0$ 时，$\boldsymbol{Q}$ 可以表示为式(14)的形式。

$$\boldsymbol{Q}=\begin{pmatrix}-\lambda_c & \lambda_c & & \\ \mu_{c2} & -\lambda_c-\mu_{c2} & \lambda_c & \\ & \ddots & \ddots & \ddots\end{pmatrix} \tag{14}$$

此时，Cloud-P2P 云存储结构退化为只有云中心服务器为用户提供服务，三维连续时间 Markov 链 $\{n_c(t),n_s(t),n_r(t)\}$ 退化为一个生灭过程。由文献[20]可知，系统的稳态分布如式(15)所示。

$$\boldsymbol{\pi}_i=\pi_{i,0,0}=(1-\lambda_c/\mu_{c2})(\lambda_c/\mu_{c2})^i \tag{15}$$

当系统中可用副本的个数 $H\geqslant 1$ 时，云中心和用户存储层同时为用户提供服务。由矩阵几何解法[21]，系统的稳态分布可以有方程组(16)给出。

$$\begin{cases}(\boldsymbol{\pi}_0,\boldsymbol{\pi}_1,\boldsymbol{\pi}_2,\ldots,\boldsymbol{\pi}_H)B[\boldsymbol{R}]=0\\ \sum_{i=0}^{H-1}\boldsymbol{\pi}_i\boldsymbol{e}+\boldsymbol{\pi}_H(\boldsymbol{I}-\boldsymbol{R})^{-1}\boldsymbol{e}=1\\ \boldsymbol{\pi}_i=\boldsymbol{\pi}_H\boldsymbol{R}^{i-H},\ i\geqslant H\end{cases} \tag{16}$$

其中，$B[\boldsymbol{R}]$ 可表示为式(17)所示的形式。

$$B[\boldsymbol{R}]=\begin{pmatrix}\boldsymbol{A}_0 & \boldsymbol{C} & & & \\ \boldsymbol{B}_1 & \boldsymbol{A}_1 & \boldsymbol{C} & & \\ & \ddots & \ddots & \ddots & \\ & & \boldsymbol{B}_{H-1} & \boldsymbol{A}_{H-1} & \boldsymbol{C}\\ & & & \boldsymbol{B} & \boldsymbol{RB}+\boldsymbol{A}\end{pmatrix} \tag{17}$$

$\boldsymbol{R}$ 是二次方程 $\boldsymbol{R}^2\boldsymbol{B}+\boldsymbol{R}\boldsymbol{A}+\boldsymbol{C}=0$ 的谱半径小于 1 的最小非负解。设 $\boldsymbol{R}$ 的初值为 0，将 $\boldsymbol{R}^2\boldsymbol{B}+\boldsymbol{R}\boldsymbol{A}+\boldsymbol{C}=0$ 变形为 $\boldsymbol{R}=-(\boldsymbol{C}+\boldsymbol{R}^2\boldsymbol{B})\boldsymbol{A}^{-1}$。采用迭代方法求得 $\boldsymbol{R}$，代入式(16)，即可求出系统的稳态分布。

3.3　性能指标

节点存储层传输率 β 定义为节点存储层响应用户数据请求的概率，即节点存储层在线副本节点个数不为 0 的概率。β 可表示成式(18)表示的形式。

$$\beta=\begin{cases}0, & H=0\\ \left(1-\sum_{i=0}^{\infty}\pi_{i,0,H}\right), & H\geqslant 1\end{cases} \tag{18}$$

数据访问延迟定义为从用户发出数据请求开始到用户接收到数据所需要的时间。系统的平均访问延迟 E 可以表示成式(19)所示的形式。

$$E=\begin{cases}\dfrac{1}{\mu_{c2}-\lambda_c}, & H=0\\ \dfrac{1}{\lambda_c}\sum_{i=0}^{\infty} i\left(\sum_{j=1}^{H-m}\pi_{i,j,(H-j)}+\sum_{j=H-m+1}^{H}\sum_{k=0}^{H-j}\pi_{i,j,k}\right), & H\geqslant 1\end{cases} \tag{19}$$

副本修复率 Λ_s 定义为单位时间内被系统成功修复的副本个数。系统正在修复 k 个副本的概率为 $\sum_{i=0}^{\infty}\sum_{j=0}^{H}\pi_{i,j,k}$。当 $0\leqslant j\leqslant H-m$ 时，$k=H-j$，当 $H-m+1\leqslant j\leqslant H$ 时，$k=0,1,2,\cdots,H-j$。Λ_s 的表达式如式(20)所示。

$$\Lambda_s=\begin{cases}0, & H=0\\ \lambda_s\left(\sum_{i=0}^{\infty}\sum_{j-1}^{H-m}(H-j)\pi_{i,j,(H-j)}+\sum_{i=0}^{\infty}\sum_{j=H-m+1}^{H}\sum_{k=0}^{H-j}k\pi_{i,j,k}\right), & H\geqslant 1\end{cases} \tag{20}$$

4 数值实验与系统优化

4.1 数值实验

在数值实验中，令副本的在线时间服从参数为 μ_s 的指数分布，$\mu_s=0.001$。节点存储层副本节点对用户数据请求服从服务率为 μ_{c1} 的指数分布，$\mu_{c1}=0.4$。云中心对用户数据请求的服务率 μ_{c2} 服从指数分布，$\mu_{c2}=0.8$。副本的修复阈值 $m=1$。

设用户数据请求的到达率为 $\lambda_c=0.6$，将副本节点修复参数 λ_s 分别设置为 0.006，0.007 和 0.008，则节点存储层传输率 β 随副本数量 H 的变化趋势如图 2 所示。

由图 2 可知，当副本个数 H 较少，如 $H=1$ 时，随着副本个数 H 的增加，用户数据请求由节点存储层副本节点服务到的概率增大，节点存储层传输率 β 呈上升趋势。当副本数量增加到一定值，如 $H=3$ 时，随着副本数量 H 的继续增加，节点存储层传输率 β 增大的趋势减缓，并逐渐趋近 100%。这表明，当系统中副本个数 H 较少时，增加节点存储层副本数量可以有效减小云中心服务器的数据传输压力，降低云端服务器的资源消耗。但副本数量增加到一定值之后，节点存储层传输率接近 100%。此时，副本数的继续增加对云端服务器的影响不大。

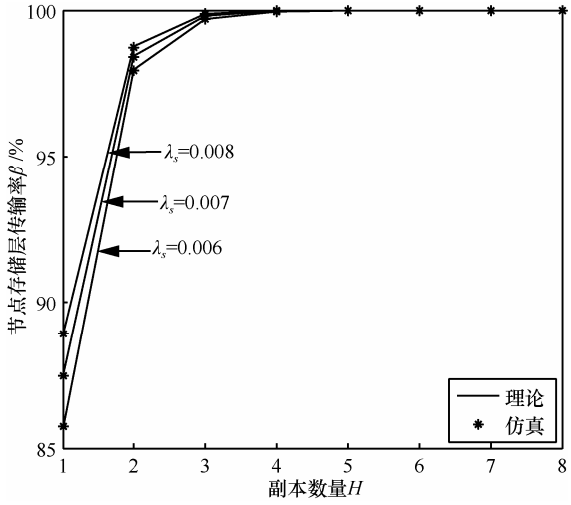

图 2 节点存储层传输率 β 随副本数量 H 的变化趋势

设副本节点修复参数为 $\lambda_s=0.006$，用户数据请求的到达率 λ_s 分别为 0.4，0.5 和 0.6，副本数量 H 影响平均访问延迟 E 变化的趋势图如图 3 所示。

由图 3 可知，当副本数量为 0 时，用户数据请求全部由云中心处理，由于云中心带宽良好，数据传输能力强，所以平均访问延迟 E 很低。随着副本数量 H 的逐渐增加，用户的数据请求被节点存储层副本节点服务到的概率增大，由于用户存储层副本结点的传输较差，因此平均访问延迟 E 呈上升趋势。随着副本数量 H 的继续增大，节点存储层的传输能力逐渐增强，平均访问延迟 E 呈下降趋势，并逐渐趋于稳定。

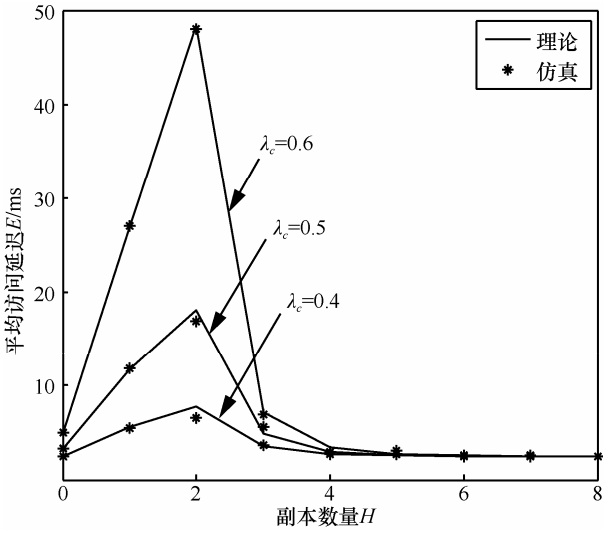

图 3 平均访问延迟 E 随副本数量 H 的变化趋势

设用户数据请求的到达率为 $\lambda_c = 0.6$，将副本节点修复参数 λ_s 分别设置为 0.006、0.007 和 0.008，副本修复率 Λ_s 随副本数量 H 的变化趋势如图 4 所示。

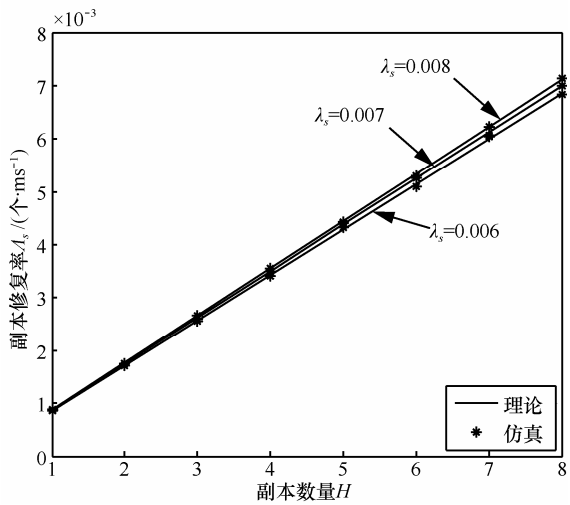

图 4　副本修复率 Λ_s 随副本数量 H 的变化趋势

由图 4 可知，随着副本数 H 的增加，系统修复副本的开销逐渐增大，副本的修复率略呈增大趋势。对于相同的副本数 H，随着副本修复速度的增加，系统的修复率呈增大趋势，且随着副本数的增加，修复率增加的趋势愈明显。

4.2　系统优化

结合图 2~图 4 中的数值实验结果可以得知，节点存储层副本节点数量越多，节点存储层的数据传输能力越强，云中心承担的数据传输压力越小。另一方面，副本节点数量越多，管理副本花费的开销也越大，因此，在 Cloud-P2P 云存储结构中，需要综合考虑系统的多项性能指标对副本数进行优化配置。

设当系统中副本数量为 H 时，将节点存储层传输率，平均访问延迟和副本修复率分别设置为 $\beta(H)$，$E(H)$ 和 $\Lambda_s(H)$。节点存储层传输率每提升 1%，由于减少云中心带宽开销而带来的利润定义为 f_1。每降低一个单位时间(ms)的数据访问延迟而获得的利润定义为 f_2。单位时间内每完成一个副本节点的修复所消耗的带宽开销定义为 f_3；系统中每增加一个节点存储层副本节点引起的存储开销定义为 f_4。云中心跟踪服务器监测节点存储层副本节点的开销定义为 f_5。由以上条件，构造如下的利润函数 $F(H)$。

$$F(H) = f_1\left(\beta(H) - \beta(0)\right) + f_2\left(E(0) - E(H)\right) - f_3\left(\Lambda_s(H) - \Lambda_s(0)\right) - (f_4 + f_5)H$$

设定副本修复时间的参数 $\lambda_s = 0.006$，$f_1 = 4$，$f_2 = 3$，$f_3 = 2$，$f_4 = 1$，$f_5 = 5$。针对数据请求的到达率 λ_c 分别为 0.4，0.5 以及 0.6 的情况，系统的利润函数 $F(H)$ 随副本数量 H 的变化趋势如图 5 所示。

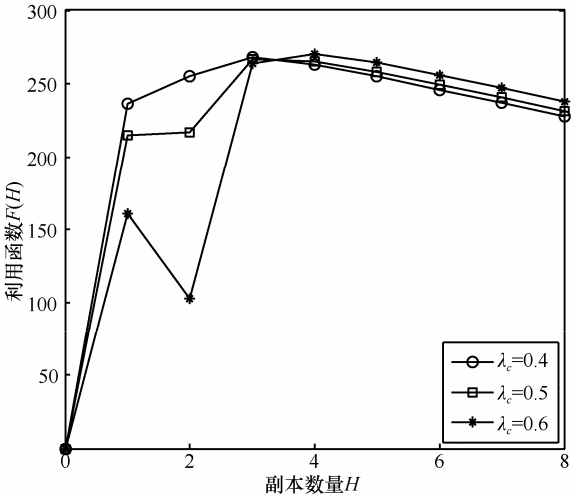

图 5　利润函数 $F(H)$随副本数量 H 的变化趋势

从图 5 中可以看出，对于较小的数据请求到达率，如 $\lambda_c = 0.4$ 和 $\lambda_c = 0.5$，当副本数量较少时，随着副本数量的增多，节点存储层为用户提供的带宽逐渐增大，云端的带宽压力减小，利润函数 $F(H)$ 呈增大趋势；随着节点存储层副本数量的不断增多，维护副本的开销逐渐增大，利润函数 $F(H)$ 呈减小趋势。而对于较大的数据请求的到达率，如 $\lambda_c = 0.6$，当副本数量较少时，随着副本数量的增多，如副本数量由 1 增加到 2，维护副本所需要的成本也不断增加，但减少云端带宽开销获得的利润小于维护副本节点所需要付出的成本，因此利润函数 $F(H)$ 呈减小趋势；随着副本节点的增多，如副本数量由 2 增加到 3，节点存储层为用户提供的带宽逐渐增大，云端的带宽压力减小，减少云端开销获得的利润大于维护副本节点所需要付出的成本，利润函数 $F(H)$ 呈增大趋势；副本节点数继续增多，如 $H > 3$，维护副本的开销逐渐增大，利润函数 $F(H)$ 呈减小趋势。

对于所有的数据请求到达率 λ_c，均存在最优的副本数量 H^*，使系统利润取得最大值。当 λ_c 分别为 0.4 和 0.5 时，最优副本数量为 $H^* = 3$，此时系统的最大利润分别为 267.925 8 和 266.232；当 $\lambda_c = 0.6$

时，最优副本数量为 $H^*=4$，系统的最大利润为 270.408 8。

5 结束语

在 Cloud-P2P 云存储结构中，将用户的闲散存储空间以 P2P 的形式组织起来，作为云中心服务器的辅助层，增加了数据的安全性并降低了系统的构建费用。基于 Cloud-P2P 云存储结构中用户访问数据机制及节点存储层副本节点的修复过程，考虑数据请求数，节点存储层可用副本数以及正在修复的副本数，构建了一个三维连续时间 Markov 链模型。使用矩阵几何解方法，给出了各性能指标的表达式。结合数值实验和系统仿真，定量分析了节点存储层副本数对系统性能指标的影响。综合多方面的性能要求，建立了一个利润函数，并给出了数据副本数的优化设置方案。

随着 4G 移动网络的不断普及，移动终端设备的性能不断增强，云存储技术和 P2P 技术向移动互联网扩展成为今后发展的趋势。同时，随着固定移动网络融合(FMC)技术的提出，基于下一代网络的 Cloud-P2P 技术成为本课题继续研究的重点。

参考文献：

[1] 张龙立. 云存储技术探讨[J]. 电信科学, 2010, 8(A): 71-74.
ZHANG L L. Cloud storage technology[J]. Telecommunications Science, 2010, 8(A):71-74.

[2] WU Y T, LEE W T, LIN Y S, *et al*. Dynamic load balancing mechanism based on cloud storage[A]. Computing, Communications and Applications Conference[C]. Hong Kong, China, 2012. 102-106.

[3] DEAN J, GHEMAWAT S. MapReduce: Simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1):107-113.

[4] RONG C, LU W, DU X, *et al*. Efficient and exact duplicate detection on cloud[J]. Concurrency Computation Practice and Experience, 2013, 25(15):2178-2206.

[5] LI S P, WONG M H. Data allocation in scalable distributed database systems based on time series forecasting[A]. IEEE International Congress on Big Data[C]. Santa Clara, USA, 2013. 17-24.

[6] WU Y T, LEE W T, LIN C F, *et al*. Cloud storage performance enhancement by realtime feedback control and de-duplication[A]. Wireless Telecommunications Symposium[C]. London, United Kindom, 2012. 1-5.

[7] LI W, YANG Y, CHEN J, *et al*. A cost effective mechanism for cloud data reliability management based on proactive replica checking[A]. IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing[C]. Ottawa, Canada, 2012. 564-571.

[8] FARAGARDI H R, SHOJAEE R, TABANI H, *et al*. An analytical model to evaluate reliability of cloud computing systems in the presence of QoS requirements[A]. International Conference on Computer and Information Science[C]. Niigata, Japan, 2013. 315-321.

[9] JOOLAHLUK J, WEN Y F. Reliable and available data replication planning for cloud storage[A]. IEEE International Conference on Advanced Information Networking and Applications[C]. Barcelona, Spain, 2013. 772-779.

[10] SAVU L. Cloud computing: Deployment models, delivery models, risks and research challenges[A]. International Conference on Computer and Management[C]. Wuhan, China, 2011. 1-4.

[11] SONG J, DENG H J. NOVA: A P2P-Cloud VoD system for iptv with collaborative pre-deployment module based on recommendation scheme[A]. International Conference on Materials Science and Information Technology[C]. Nanjing, China, 2013. 1566-1570.

[12] JACOB C. Cost and profit driven Cloud-P2P interaction[J]. Peer-to-Peer Networking and Applications, 2015, 8(2): 244-259.

[13] BABAOLU O, MARZOLLA M, TARMBURINI M. Design and implementation of a P2P cloud system[A]. Proceedings of the ACM Symposium on Applied Computing[C]. Trento, Italy, 2012. 412-417.

[14] HANNA K, ALBERTO M. P2P and Cloud: A marriage of convenience for replica management[A]. International Workshop on Self-Organizing System[C]. Delft, Netherlands, 2012. 60-71.

[15] YANG Z, ZHAO B Y, XING Y, *et al*. AmazingStore: Available, low-cost online storage service using cloudlets[A]. International Workshop on Peer-to-Peer Systems[C]. San Jose, USA,2010. 1-5.

[16] LI Z, ZHANG T, HUANG Y, *et al*. Maximizing the bandwidth multiplier effect for hybrid Cloud-P2P content distribution[A]. IEEE International Workshop on Quality of Service[C]. Coimbra, Portugal, 2012. 1-9.

[17] ZHE S, SHEN J. A high performance peer to cloud and peer model augmented with hierarchical secure communications[J]. Journal of Systems and Software, 2013, 86(7): 1790-1796.

[18] CONG X, SU S, SHUANG K, *et al*. An efficient server bandwidth costs decreased mechanism towards mobile devices in cloud-assisted P2P-VoD System[J]. Peer-to-Peer Networking and Applications, 2014, 7(2):175-187.

[19] 张启飞, 张尉东, 李文娟等. 基于对等网络的面向小文件的云存储系统[J]. 浙江大学学报(工学版), 2013, 47(1): 8-14.
ZHANG Q F, ZHANG W D, LI W J, *et al*. Cloud storage system for small file based on P2P[J]. Journal of Zhejiang University (Engineering Science), 2013, 47(1): 8-14.

[20] KHAZAEI H, MISIC J, MISIC V. Performance analysis of cloud computing centers using M/G/m/m+r queuing systems[J]. IEEE Transactions on Parallel and Distributed Systems, 2013, 23(5): 936-943.

[21] NIGEL G, MALGORZATA M, REILLY O, *et al*. Second-order Markov reward models driven by QBD processes[J]. Performance Evaluation, 2012, 69(9): 440-455.

作者简介：

金顺福（1966-），女，朝鲜族，内蒙古满洲里人，燕山大学教授、博士生导师，主要研究方向为网络资源分配与优化、排队论研究与应用等。

王晨飞（1988-），男，河北肃宁人，燕山大学硕士生，主要研究方向为网络资源分配与优化。

陈玲玲（1984-），女，河北保定人，中国联合网络通信有限公司秦皇岛市分公司工程师，主要研究方向为网络资源分配与优化。

霍占强（1979-），男，河北邯郸人，河南理工大学副教授、硕士生导师，主要研究方向为计算机系统及网络的性能分析、离散时间排队理论、计算机网络协议的分析。