

# 异构环境下非均质性最小冗余分布式云存储系统

胡力文<sup>1</sup>, 李启明<sup>2</sup>

(<sup>1</sup> 江苏财会职业学院 信息系, 江苏 连云港 222061; <sup>2</sup> 上海海事大学 信息工程学院, 上海 201306)

**摘要:** 为提高分布式云存储系统构建的合理性, 提出一种异构环境下考虑非均质性的最小冗余分布式云存储系统构建方法. 首先, 开发了异构环境中, 进行数据可用性计算的分析框架. 由于在存储节点增长时数据可用性计算非常复杂, 提出了一种以较少计算代价的方法来估计真实值的蒙特卡洛方法. 其次, 由于决定在每个主机上分配的最佳冗余量是很难计算的, 提出了一种基于粒子群算法 (PSO) 的分配方法. 最后, 提供了一个简单的迭代算法来确定保证不同存储应用程序的数据可用性所需的最小冗余. 实验结果显示, 所提方法可实现数据冗余和存储成本的降低.

**关键词:** 异构环境; 非均质性; 最小冗余; 分布式; 云存储

## Minimum Redundancy Distributed Cloud Storage System Considering Heterogeneity in Heterogeneous Environment

HU Li-wen<sup>1</sup>, LI Qi-ming<sup>2</sup>

(<sup>1</sup> Information Department, Jiangsu College of Finance & Accounting, Lianyungang 222061, China; <sup>2</sup> Information Engineering College, Shanghai Maritime University, Shanghai 201306, China)

**Abstract:** In order to improve the rationality of distributed cloud storage system construction, a minimum redundant distributed cloud storage system construction method under heterogeneous environment is proposed. Firstly, an analytical framework for computing data availability in heterogeneous environments is developed. Because the computation of data availability is very complicated when the storage nodes are growing, a Monte Carlo method is proposed to estimate the real values with less computational cost. Secondly, since it is difficult to compute the optimal redundancy allocation for each host, a method based on particle swarm optimization (PSO) is proposed. Finally, a simple iterative algorithm is provided to determine the minimum redundancy required to ensure data availability for different storage applications. Experimental results show that the proposed method can reduce data redundancy and storage cost.

**Key words:** heterogeneous environments; heterogeneity; minimum redundancy; distributed; cloud storage

**作者简介:**

胡力文女, (1983-), 硕士研究生, 讲师. 研究方向为计算机软件技术理论. E-mail: 61661605@qq.com.

李启明男, (1982-), 博士, 副教授. 研究方向为大数据分析、高性能计算.