# Accepted Manuscript

Efficient image restoration of virtual machines with reference count based rewriting and caching

Jiwei Xu, Tao Wang, Xiaozhao Xing, Wenbo Zhang, Hua Zhong

PII: S0167-739X(17)30121-8
DOI: http://dx.doi.org/10.1016/j.future.2017.06.026
Reference: FUTURE 3524

To appear in: *Future Generation Computer Systems*

Received date : 20 January 2017
Revised date : 15 May 2017
Accepted date : 25 June 2017

Please cite this article as: J. Xu, T. Wang, X. Xing, W. Zhang, H. Zhong, Efficient image restoration of virtual machines with reference count based rewriting and caching, *Future Generation Computer Systems* (2017), http://dx.doi.org/10.1016/j.future.2017.06.026

# Efficient image restoration of virtual machines with reference count based rewriting and caching

Jiwei Xu[a], Tao Wang[a,*], Xiaozhao Xing[b], Wenbo Zhang[a], Hua Zhong[a]

[a]*Institute of Software, Chinese Academy of Sciences, Beijing, China, 100190*
[b]*Institute of Scientific and Technical Information of China, Beijing, China, 100038*

**Abstract**

Virtual machine (VM) image backups have duplicate data blocks distributed in different physical addresses, which occupy a large amount of storage space in a cloud computing platform [2, 11]. Deduplication is a widely used technology to reduce the redundant data in a VM backup process. However, deduplication always causes the fragmentation of data blocks, which seriously affects the VM restoration performance. Current approaches often rewrite data blocks to accelerate image restoration, but rewriting could cause significant performance overhead because of frequent I/O operations. To address this issue, we have found that the reference count is a key to the fragmentation degree from a series of experiments. Thus, we propose a reference count based rewriting method to defragment VM image backups, and a caching method based on the distribution of rewritten data blocks to restore VM images. Compared with existing studies, our approach has no interfere to the deduplication process, needs no extra storage, and efficiently improves the performance of VM image restoration. We have implemented a prototype to evaluate our approach in our real cloud computing platform OnceCloud. Experimental results show that our approach can reduce about 57% of the dispersion degree of data blocks, and accelerate about 51% of the image restoration of virtual machines.

*Keywords:* virtual machine image, image restoration, reference count, data deduplication, data defragment

*Corresponding author: Tao Wang
*Email address:* wangtao@iscas.ac.cn (Tao Wang)