

## Abstract

Virtualization technology provides ability for creating a virtual machine that can checkpoint and migrate its internal state transparently to guest operating systems and applications. However, it is nontrivial to provide checkpointing mechanisms with the same level of transparency for parallel applications running on top of a cluster of virtual machines. This paper proposes a layering Virtual Cluster Architecture (VCA) and the Virtual Cluster CheckPointing (VCCP) system, a thin-layer software system that performs checkpointing and recovery of a virtual cluster at the hypervisor-level without modifications to guest OS and applications. The VCCP system implements novel hypervisor-base Coordinated Checkpointing and Recovery protocols. Our correctness analysis shows that VCCP does not cause message lost or reordering. We also outline its limitations and show that VCCP should be applicable to applications that work with traditional process-level checkpointing. We have also implemented the VCCP prototype base on QEMU system and measured its performance on the NAS parallel benchmark. The experiments show the incorporation of the VCCP protocols to QEMU only generates low execution overheads. However, for some NAS kernel, we found that the execution times of some NAS kernels do not decrease as the number of VM increases due to the inefficiency of VDE switch network. We plan to implement a new network infrastructure in our future work. We also find that checkpointing performance depends on the time VM spend to save local VM state, flush time, and the time VM wait one another for local checkpointing to finish. Finally, we would like to highlight here that the emphasis of this work, at this initial stage, is to promote the transparency issue, while leaving performance to the future implementations of VCCP.