**TwinHype: A Novel Approach to Reduce Cloud Downtime**

**Lei Lu (VMWare Inc.), Xing Gao (College of William and Mary), Jidong Xiao (Boise State University)**

**INTRODUCTION**
Apply security patches to hypervisors requires VMM module reloading which incurs service downtime and non-trivial network work traffic due to VM live migration to a different host.

**Traditional approach:**
- Source VM Running
- Migrating stage
- Target VM Runs
- Patching source VMM
- Migrating stage
- PAUSED
- Source VM Runs

**Proposed approach:**
- Source VM Running
- Migrating stage
- Target VM Runs
- Patching source VMM
- Migrating stage
- PAUSED
- Source VM Runs

**GOAL**
- Reducing cloud service downtime due to hypervisor level system maintenance e.g. software upgrading, patching et c.
- Avoiding non-trivial network traffic incurred by the existing virtual machine migration solutions

**APPROACH**
- Runs two hypervisors on the same physical machine for software upgrading.
- When upgrading, the virtual machines on one hypervisor can be migrated onto the second one on the same host to reduce service downtime and network traffic.
- After the upgrading, the virtual machines can then be migrated back to the first hypervisor.

**IMPLEMENTATION**
- Normally, KVM exposes its virtualization capabilities via a character device /dev/kvm and tools like QEMU can interact with it via its APIs.
- We modify current KVM module code so that it can run with a second KVM module /dev/kvm2.
  - Supporting VMM coexistence
  - Enable/disable CR4.VMXE bit in a cooperative way
  - Add a per-cpu variable to avoid making processors to enter VMX mode multiple times.
- We add an option to QEMU so that it can create VMs using /dev/kvm2
- We ensure that no code dependency exists between the two VMM modules, so that we can load/unload one KVM module without affecting a second one.

**NEXT STEPS**
- Performance overhead analysis for the two migrations and its improvement
- Performance overhead study for running two KVMs at the same time

**CONCLUSION**
We present TwinHype, a new solution for cloud providers, not only enabling cloud providers to upgrade hypervisors without incurring significant downtime to their customers, but also helping them avoid the non-trivial network traffic caused by traditional migration schemes. We believe that this new model will benefit cloud customers as well as cloud providers.