**Title:**

**An Empirical Study on the Discrepancy between Performance Testing Results from Virtual and Physical Environments**

**Abstract:**Large software systems often undergo performance tests to ensure their capability to handle expected loads. These performance tests often consume large amounts of computing resources and time in order to exercise the system extensively and build confidence on results. Making it worse, the ever evolving field environments require frequent updates to the performance testing environment. In practice, virtual machines (VMs) are widely exploited to provide flexible and less costly environments for performance tests. However, the use of VMs may introduce confounding overhead (e.g., a higher than expected memory utilization with unstable I/O traffic) to the testing environment and lead to unrealistic performance testing results. Yet, little research has studied the impact on test results of using VMs in performance testing activities. In this thesis, we evaluate the discrepancy between the performance testing results from virtual and physical environments. We perform a case study on two open source systems -- namely Dell DVD Store (DS2) and CloudStore. We conduct the same performance tests in both virtual and physical environments and compare the performance testing results based on the three aspects that are typically examined for performance testing results: 1) single performance metric (e.g. CPU usage from virtual environment vs. CPU usage from physical environment), 2) the relationship between two performance metrics (e.g. correlation between CPU usage and I/O traffic) and 3) statistical performance models that are built to predict system performance. Our results show that 1) A single metric from virtual and physical environments do not follow the same distribution, hence practitioners cannot simply use a scaling factor to compare the performance between environments, 2) correlations among performance metrics in virtual environments are different from those in physical environments and 3) statistical models built based on the performance metrics from virtual environments are different from the models built from physical environments suggesting that practitioners cannot use the performance testing results across virtual and physical environments. In order to assist the practitioners leverage performance testing results in both environments, we investigate ways to reduce the discrepancy. We find that such discrepancy may be reduced by normalizing performance metrics based on deviance. Overall, we suggest that practitioners should not use the performance testing results from virtual environment with the simple assumption of straightforward performance overhead. Instead, practitioners and future research should investigate leveraging normalization techniques to reduce the discrepancy before examining performance testing results from virtual and physical environments.