



Performance Tests of CMSSW on the CernVM

Marko Petek (UERJ), Stephen Gowdy (CERN)

mpetek@cern.ch, gowdy@cern.ch

Introduction

The CERN Virtual Machine (CernVM) Software Appliance is a project developed in CERN with the goal of allowing the execution of the experiment's software on different operating systems in an easy way for the users.

To achieve this it makes use of Virtual Machine images consisting of a JEOS (Just Enough Operational System) Linux image, bundled with CVMFS, a distributed file system for software. This image can then be run with a proper virtualizer on most of the platforms available. It also aggressively caches data on the local user's machine so that it can operate disconnected from the network.

CMS wanted to compare the performance of the CMS Software running in the virtualized environment with the same software running on a native Linux box.

To answer this need a series of tests were made on a controlled environment during 2010-2011.

This work presents some results of those tests.

Tests Performed

In order to evaluate the overhead imposed by running CMSSW inside a Virtual Machine, it was decided to compare the running time of exactly the same script in a control computer both inside and outside the Virtual Machine.

To do this a dedicated control computer was configured. In order to avoid any kind of interference no other applications were running on this control computer, besides the ones needed to perform the tests.

The configuration of the environment was as follows:

- HP DC7900 with 4GBs of main memory and SLC5 as O.S.
- Oracle VirtualBox version 3.2.5
- CernVM version (JEOS) 2.1.0
- CMSSW 3_8_5

The maximum amount of memory allowed by VirtualBox was 2.5 GBs. In order to equate this with the LinuxBox environment, the size of memory used by SLC5 was limited to be also 2.5 GBs.

The scripts used to evaluate the performance of both environments were chosen from the list of scripts used to evaluate new releases of CMSSW (RelVal). Those scripts are well known and controlled.

Some of the scripts chosen had very fast times of execution, even when the number of events was increased. Other were very slow, demanding sometimes days of running in order to collect proper data.

After experimenting, some adequate scripts were selected to perform the tests.

Three runs were made for each script with 10, 30 and 50 events respectively.

One set of results was chosen to be presented in this work. It shows the same behavior observed throughout all tests performed.

New tests were conducted on the last months, now with scripts taking a very long time of executions scripts to stress the system as much as possible.

References

CernVM site: <http://cernvm.cern.ch/>.

Predrag Buncic, Jakob Blomer, Pere Mato, Carlos Aguado Sanchez, Leandro Franco, and Steen Klemer. "Cernvm - a virtual appliance for lhc applications". In Proceedings of the XII International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT08), 2008

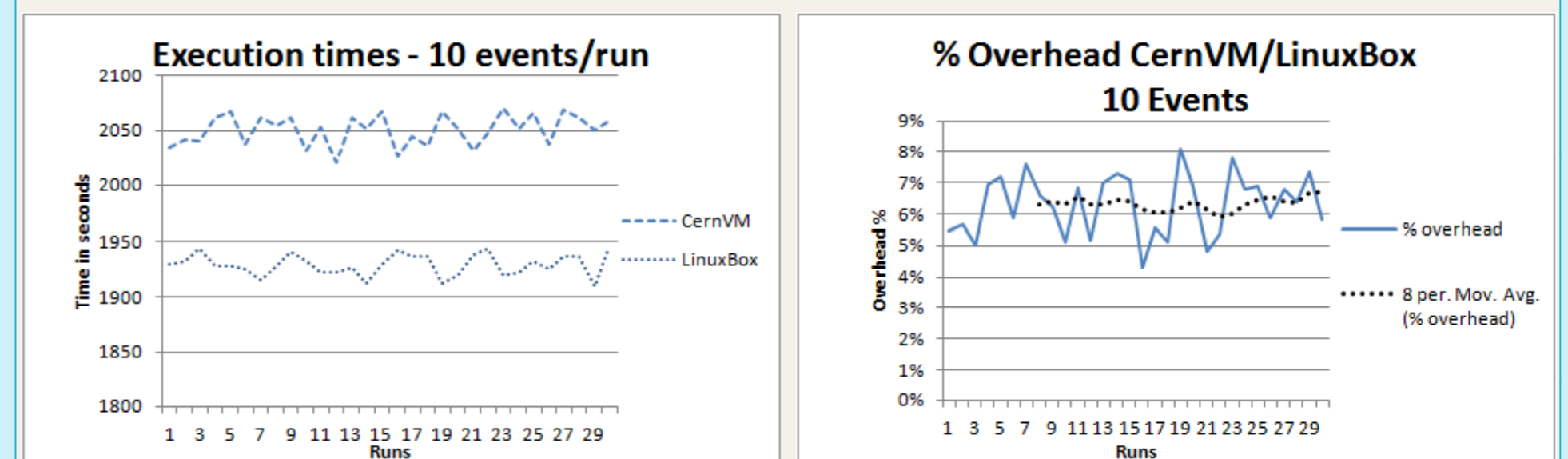
B Segal et al.; "LHC Cloud Computing with CernVM", Proceedings of the XIII. International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT10), Jaipur, 2010, PoS ACAT(2010)004

Performance Graphs

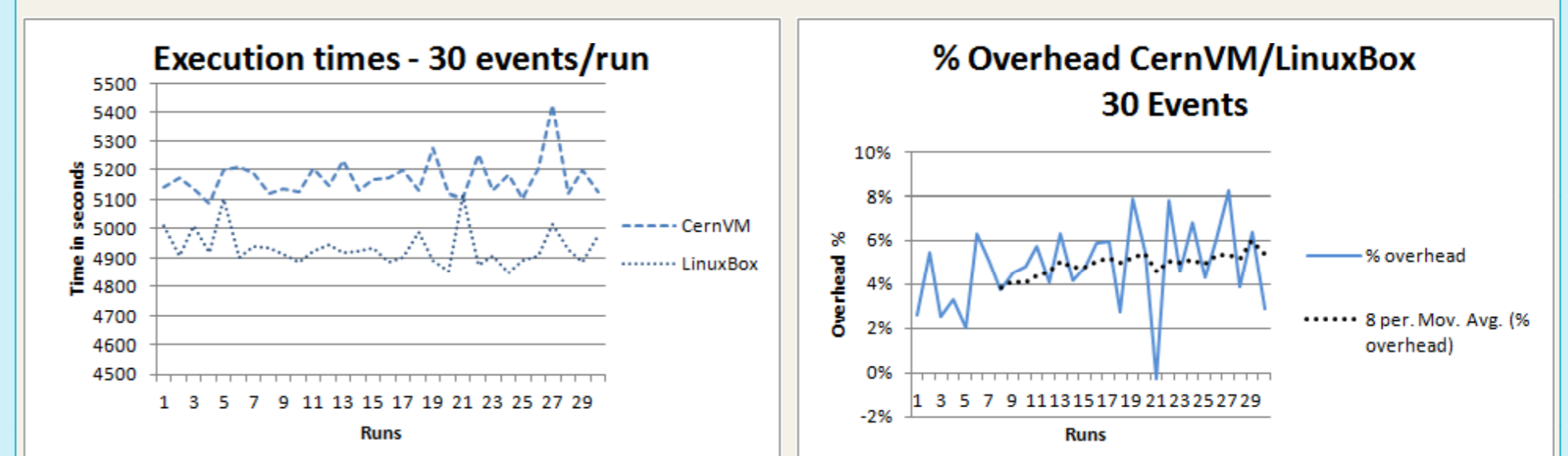
The graphs below show the performance of one test script submitted to runs of 10, 30 and 50 events.

The left graph has the Execution Time in Seconds, and the right graph plots the % of Overhead that CernVM had in relation to the plain LinuxBox.

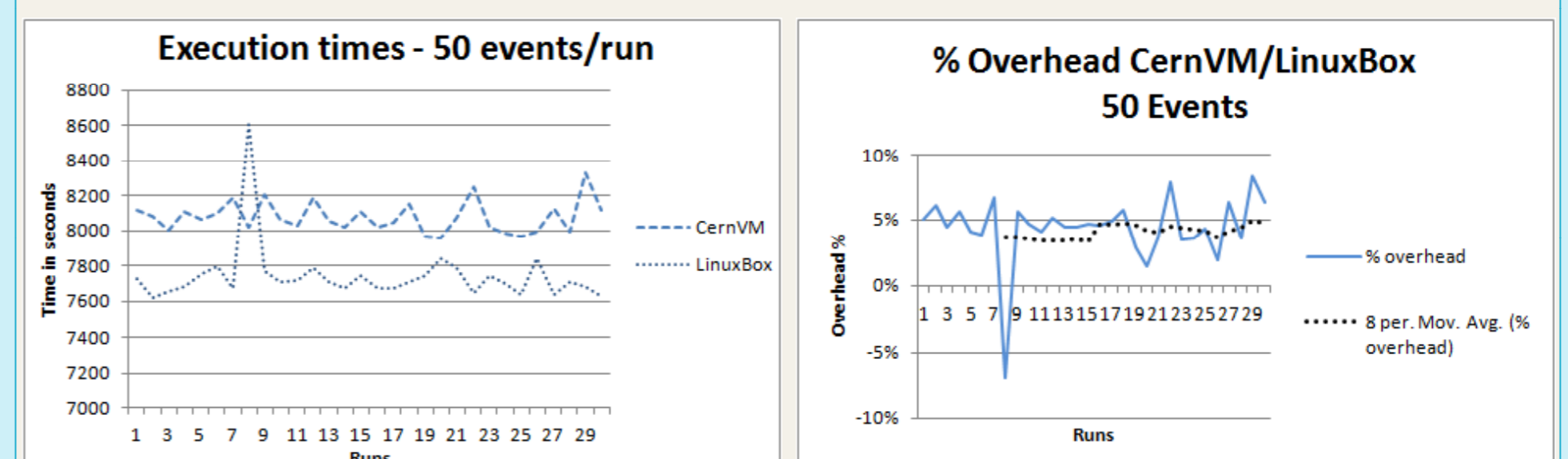
RUNS WITH 10 EVENTS



RUNS WITH 30 EVENTS



RUNS WITH 50 EVENTS



Based on the data represented by the above graphs, it is possible to see that the overhead due to the use of a Virtualizer is on the range of 4% to 6% when compared with a plain LinuxBox.

Tests performed in both Mac OS x and Microsoft Windows gave similar results.

The table below synthesizes the overhead according to the number of events analyzed.

Number of events	10	30	50
Overhead	6%	5%	4%

Average % of overhead among the three tests

Conclusion

Based on the performed tests and their results, it is fair to say that the average overhead presented by the usage of a Virtualizer is small enough to be surpassed by the advantages offered by the technology.

With a CernVM running in its desktop or laptop, a researcher can have a private testing environment accessible from any place, even without an internet connection.

The Virtual Machine also allows an easier configuration of the CMSSW environment that can be up and running in just a few minutes through a fast setup procedure. It also allows the submission of jobs for the Computing Grid and has a built-in feature (CVMFS) that keeps the new CMSSW releases always available and updated.

It is a cutting edge technology that can benefit all researchers of the CMS Collaboration.