CERN Summer Student Report

BOINC-VM: LHC Volunteer Cloud Project

Rohit Yadav (Supervisor: Ben Segal)

May 19 – July 16, 2010

Table of Contents

1. Introduction

The "LHC Volunteer Cloud Project" has been in development at CERN, over the last 12 months. This project proposes to use volunteer computing as a cheap resource for the CERN LHC experiments, capable of supporting event simulation and reconstruction jobs, conveniently accessed by the experiments' existing job production facilities. It is an integrated cross-platform solution supporting Windows, Mac and Linux and is based on BOINC [1] distributed computing system and uses the CernVM [2] virtual machine.

Previously, work was done in running applications in virtual machines [3], PyBoinc [4], VMController [5], VMWrapper [6]. The project's progress has recently been summarised in a conference paper presented in Jaipur, India at ACAT 2010. [7]

2. BOINC-VM

BOINC-VM [8] [9] is an open-source general-purpose integrated cloud computing solution that is aimed to provide cross-platform distributed computing platform using virtualization as the key technology and can be used for harnessing volunteer computing resources like laptops, desktops, server farms etc. for CPU intensive scientific applications.

BOINC-VM exploits existing open source softwares: BOINC and VirtualBox; PyBoinc, VMWrapper, and VMController which requires chirp and libraries like twisted, stomper, netifaces, coilmq, and simplejson.

2.1 BOINC

BOINC [1] is the most popular open source middleware supporting volunteer computing

projects. BOINC was developed by the famous SETI@home project team at UC Berkeley and is the basis for CERN's existing LHC@home project, currently used for LHC beam tuning studies.

BOINC consists of a server that hosts the project and a client which is installed by a volunteer on her machine. When volunteers attach their BOINC clients to such projects, the BOINC server sends them jobs (science application and data files) which are executed on the volunteer machine and the results of which are sent back to the server; based on the amount of CPU-time volunteered, a volunteer is awarded credit in form of a virtual currency called 'cobblestone'. The volunteers can form teams and compete with other teams, for this virtual money.

During the development of this project, a BOINC server (using the standard server VM image) was deployed at *cvmappi09.cern.ch* and test projects like http://cvmappi09.cern.ch/test/, http://cvmappi09.cern.ch/pytest/ with sample applications were created to test automatic configuration of newly attached clients, trickle messaging for partial creditting and PyBoinc based test application.

2.2 Hypervisor: VirtualBox

VirtualBox is a free and open source hypervisor software that is supported on a wide variety of operating system including Windows, Mac and Linux. [11]

2.3 Virtual Machine appliance: CernVM

CernVM [2] is a CERN-developed system for managing and supporting LHC physics computing code running in virtual machines. Virtualisation allows the exact computing environment of each LHC experiment to be reproduced on all the diverse volunteer platforms found in a BOINC community. CernVM also offers a Cloud Interface "Co-Pilot" to connect groups of virtual machines to each LHC experiment's standard job production system. Thus by combining BOINC, CernVM and Co-Pilot, a "Volunteer Cloud" of BOINC machines can be made available to the experiments without changing their software or their production procedures.

2.4 VMController

Originally written by David Garcia Quintas in 2009, VMController [5] is an open-source general-purpose host-guest VM communication system for VirtualBox. During the project it was tested on different operating systems: Windows, Mac and Linux, and many new interfaces were proposed, some interfaces were fixed and bugs were reported. The packaging issue was resolved by making a python script that created python bundles, also called 'eggs' which can be found on project's download page [10].

2.5 PyBoinc and VMWrapper

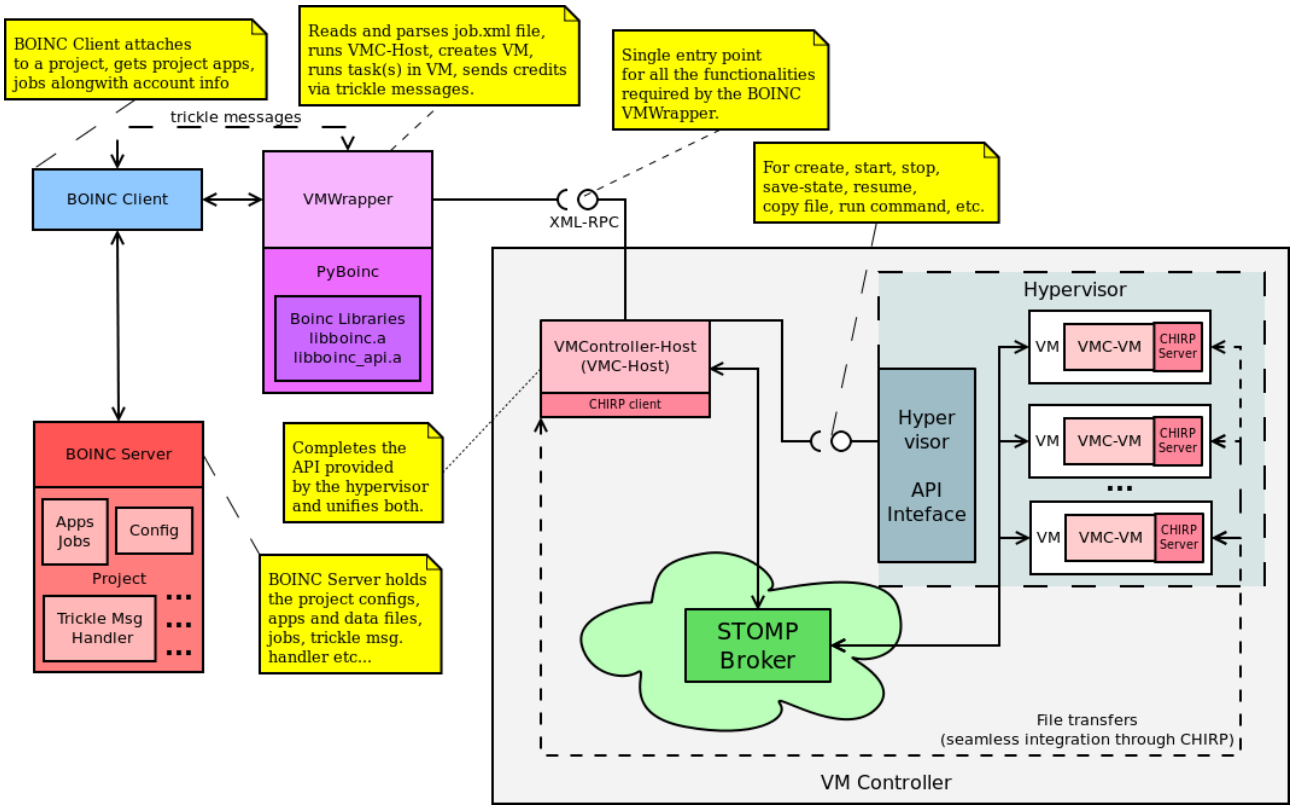
Originally written by David J. Weir in 2009, PyBoinc [4] provides python APIs for BOINC which are used in VMWrapper [6]. During the project, PyBoinc bindings were built and tested for all the three operating systems, which can be found on project's download page [10]. As of current state, x86 (32-bit Intel) builds are available for Windows, Mac and Linux and x86_64 (64-bit Intel) builds are available only for Linux (SLC5 and Ubuntu).

VMWrapper [6] is based on BOINC Wrapper which allows applications to run in guest VM as well as in the host. It was originally written in ppyython by Jarno Rantala in 2009. During the project, the code for computing credit for the resources used was implemented, using Trickle Messaging API, which involved writing a Trickle Message handler at the server and using the *boinc_send_trickle_up* API in the wrapper. And, the code was simplified and split into multiple

files.

2.6 Design

The following diagram is an approximation that captures the whole system:



3. Conclusions and Further Work

A proof of concept system ran successfully on both 32 and 64 bit Ubuntu Linux in standalone mode. It involved in creating a virtual machine and booting it. The rest was handled by the Co-Pilot agent. This concludes that this system can work and has potential.

The major challenges ahead involves development of VMController to include some new sets of interfaces, support for VMWare, and ruggedisation and testing of the system on all major operating systems and on 32 and 64 bit computers. Followed by packaging and deployment of this system on Windows, Mac and Linux.

4. Acknowledgements

Many thanks to my supervisor Ben Segal for all the advice and support, I have had during the summer. It has been an honour to work with him. Secondly, I want to thank all the people who have helped me with my project: David Garcia Quintas, Artem Harutyunyan, Carlos Aguado-Sanchez, Predrag Buncic, David Weir. Also, I want to thank everyone that were in my office 32-rb-15, at the time. Lastly, thanks to everyone at CERN who made this experience possible. I'll always remember the work culture, the brainstorming and discussions, and the friends I made there.

5. References

[1] David Anderson et al, “BOINC: Berkeley Open Interface for Network Computing”: http://boinc.berkeley.edu/

[2] Predrag Buncic et al, CernVM: http://cern.ch/cernvm

[3] Ben Segal, David Weir, Kevin Reed et al, “Running apps in virtual machines”, 2008: http://boinc.berkeley.edu/trac/wiki/VmApps

[4] David J. Weir, “PyBoinc: A python API for BOINC”, 2009: http://plato.tp.ph.ic.ac.uk/~djw03/boinc-python/documentation-0.3.1/

[5] David Garcia Quintas, “VMController: A host<->guest VM communication system for Virtual Box”, 2009: http://boinc.berkeley.edu/trac/wiki/VirtualBox, http://code.google.com/p/vmcontroller/

[6] Jarno Ranatal, “VMWrapper”, 2009: http://boinc.berkeley.edu/trac/wiki/VmApps

[7] Ben Segal, Predrag Buncic et al, “LHC Cloud Computing with CernVM ”, *Proceedings of Science*, 2010 : http://www.cern.ch/ben/ACAT_paper.pdf

[8] BOINC-VM project: http://code.google.com/p/boincvm/

[9] BOINC-VM wiki: http://code.google.com/p/boincvm/w/list

[10] BOINC-VM downloads: http://code.google.com/p/boincvm/downloads/list

[11] VirtualBox, open-source hypervisor: http://virtualbox.org