2010 18th Annual IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems

Towards Collaborative Research and Education in Computer Architecture with the Archer System

Girish Venkatasubramanian, David Isaac Wolinsky and Renato J. Figueiredo
Advanced Computing and Information Systems Laboratory, University of Florida, Gainesville, Florida
girish@acis.ufl.edu, davidiw@acis.ufl.edu, renato@acis.ufl.edu

***Abstract*—Archer is a simulation environment and computing resource for research in the field of computer architecture. Archer facilitates the creation of an on-demand computing grid, the deployment of simulation tools on this grid and the batch-scheduling of large-scale simulation jobs on this grid. These features enable the use of Archer for simulation-based research as well as dissemination of tools and results among multiple research groups. This paper overviews and reports our experience in the use of Archer for collaborative research and for education.**

I. Introduction

Simulation is the primary research tool for evaluating new ideas [1] as well as the most-used teaching technique [2] in the field of computer architecture. However, with increasing complexity of computer systems, both human and IT resources required for conducting large-scale simulation studies, such as the time for developing simulation tools as well as the time and computing resources for running the simulations, increase significantly. Moreover, the nature of simulation-based research is becoming more distributed with different research teams at geographically diverse locations investigating issues in different computer sub-systems. Due to the inter-related nature of these issues being investigated, having access to the tools, models and simulation techniques developed by other teams will greatly benefit research groups. Sharing simulation tools and models enables re-creating the work of other teams and building upon it, leading to synergistic research [3]. However, such collaboration is hindered by the varied software stack requirements as well as the time involved in acquiring expertise in the tools of different research teams.

The Archer system can be used to meet these challenges. Archer [4] is a collaborative simulation environment and computing resource for simulation-based research and education in computer architecture. Archer, as shown in Figure 1, is a distributed system comprising of a shared resource pool of seed nodes (currently consisting of more than 450 nodes) connected using a self-configuring virtual network overlay, with each node running the Archer virtual machine (VM) image. Condor-based middleware [5] is deployed on this pool to support scheduling of simulation jobs. Using this architecture, Archer is able to provide a large number of cycles for simulations and facilitate the development, deployment and sharing of simulation tools and knowledge, thereby enabling collaborative research.

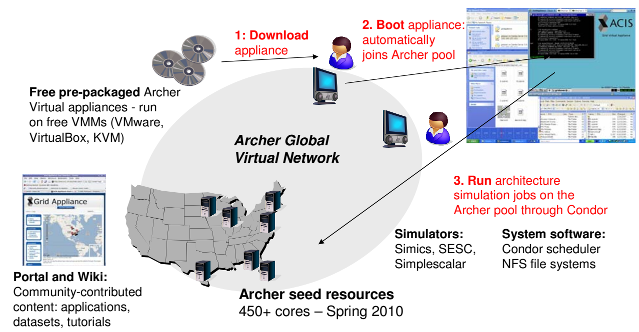

Figure 1. Using Archer for computer architecture simulations. Users can download the Archer VM and boot it using a virtualization software like VMware player or Virtual Box. Within minutes of booting up, the VM automatically joins the existing Archer pool. Users can then submit their jobs from their Archer VM to the nodes in the Archer pool.

II. Using Archer for Collaborative Research

Archer is designed to support three primary usability goals which make it a suitable platform for multiple research teams to collaborate and transfer tools and knowledge.

A. Providing a standard software environment

The Archer VM running on every node of the Archer pool, is built with a standard baseline Linux distribution, currently Ubuntu 9.10. Since the same VM image is used on every Archer node, tools and models developed on this VM is guaranteed software compatibility with every node in the pool. Using VMs also allows the deployment of pre-built simulators like Simics [6] for which the source code is not available.

One penalty associated with running simulations on a VM is a potential slowdown in the simulation speed. To evaluate this slowdown, Simics is used to create a single-core x86 machine running Xen 3.1.0, launch a transaction processing workload (TPCC) in it and checkpoint the machine. Starting from this checkpoint of the simulated machine, 1 billion instructions of TPCC workload are simulated on an Intel Xeon 2.0 GHz physical machine as well as on an Archer VM running on the physical machine. The experiment

1526-7539/10 $26.00 © 2010 IEEE
DOI 10.1109/MASCOTS.2010.66

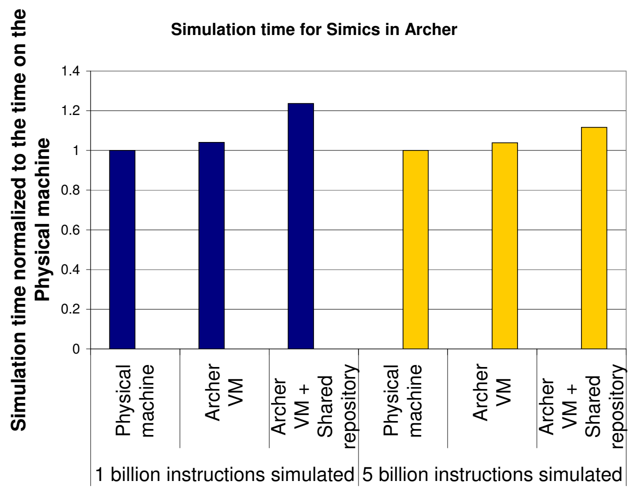

Figure 2. The simulation speed of Simics, with the checkpoint present on the same VM, reduces on on Archer VMs by less than 5%. Even when checkpoints from shared repositories are used over a 100Mbps link, the slowdown is about 11%. This slowdown is smaller for simulations which are less I/O-intensive.

is then repeated by increasing the number of simulated instructions to 5 billion and running the simulations on both the physical machine and the Archer VM. The comparison of the simulation times for both these cases, presented in Figure 2, shows a slowdown of less than 5% when using Archer. Even when the checkpoints are located on another Archer VM in the pool and accessed through a shared repository mechanism (described in Section II-C) over a 100 Mbps LAN, the slowdown for long-running simulations is about 11%. This slowdown will be smaller for CPU-bound simulations with smaller I/O component and with improvement in virtualization technology. Though the Archer software stack is packaged on VMs for convenient use in shared resources such as desktop grids, it can be installed on physical machines which are dedicated to an Archer pool. Such a set-up may be used for performance-critical simulation jobs or simulation tools requiring direct access to the physical hardware.

B. Creating on-demand resource pools

Simulation-based computer architecture research requires many CPU cycles. However, investing the money to create a dedicated computing pool for meeting this demand may not be feasible in cases of small-sized research groups or in classroom environments. Hence, Archer provides a community-driven pool of resources. A user can join Archer, simply by downloading and running the VM image from the Archer website [7]. On booting, the VM automatically makes the host machine on which it is running a part of the Archer pool. The user can then submit jobs to the pool using Condor, as shown in Figure 1. These jobs are scheduled on the available machines in the pool, with jobs belonging to the owner of a node getting scheduled on that node with higher priority. Archer VMs may also be used to create on-demand private clusters out of existing computing infrastructure, or even out of cloud-based computing resources.

C. Enabling collaboration using shared repositories

When an Archer VM boots up and joins the Archer pool, a predefined directory in that VM is network mounted via NFS and made visible to all the other nodes in the Archer pool. This directory can be used as a repository for sharing simulation models and workloads between different users in the pool. Archer also includes a web-based Wiki where users can post tutorials and guides, as well as a discussion forum. Both these help in transferring knowledge between users.

III. Case Study: Simics on Archer

Archer has been used by many research groups as a resource pool for simulation-based research and the list of resulting publications can be found at the Archer website [7]. Archer has also been used to create and store simulation configuration and execution templates for Simics [6], a full-system simulation framework, and to share these among many researchers at different institutions. For instance, tutorials for using Simics on Archer and for developing customized architectural models for Simics were created on the Archer Wiki. Script templates to set-up and run Simics simulation jobs were also created and placed in the shared repository of Archer nodes at University of Florida. Other research groups within the Archer collaboration used these to quickly start working with Simics on Archer. One group used GEMS as an add-on to Simics and shared their experience by writing a Wiki tutorial on using GEMS on Archer.

Archer has also been used for conducting undergraduate and graduate course projects at University of Florida. One example was a project aimed at characterizing the cache and Translation Lookaside Buffer behavior for virtualized workloads. For this project, a checkpoint of an x86 machine running Xen 3.0 with VMs running different benchmarks was created and, along with simulation scripts, stored in the shared repository of an Archer VM. Students joined Archer by running the Archer VM image on their personal computers, accessed the checkpoints from the repository and ran the simulations on the pool of Archer machines. By using Archer, two class sessions with 40 students who had little to no prior exposure to full-system simulation, batch schedulers or virtual machines were able to run more than 100 simulations per-person and successfully complete this project in about two months.

IV. Conclusion

This paper describes the use of Archer for simulation-based computer architecture research. Archer provides a readily-available pool for architectural simulation experiments and an easy infrastructure to share the tools, techniques and results from these simulation experiments. Real-life case studies showing the use of Archer amongst geographically distributed research groups as well as in a classroom environment are presented, demonstrating the viability of Archer as a system for collaborative research and education in computer architecture.

The deployed Archer infrastructure is open for use by computer architecture researchers and educators and can be accessed within minutes by following an introductory tutorial, posted on the Archer website [7]. The Archer software is open-source and is not limited to the computer architecture domain and the Archer VM can be used to create independent resource pools to deploy collaborative environments for other research communities.

Acknowledgment

This work is sponsored in part by the National Science Foundation under CRI collaborative awards 0751112, 0750847, 0750851, 0750852, 0750860, 0750868, 0750884, and 0751091. The authors also acknowledge the researchers at the seed Archer institutions including Florida State University, Northeastern University, University of Texas at Austin, Northwestern University, University of Minnesota, Cornell University and University of Wisconsin.

References

[1] E. Argollo *et al.*, "Cotson: infrastructure for full system simulation," *SIGOPS Operating Systems Review*, vol. 43, no. 1, pp. 52–61, 2009.

[2] G. S. Wolffe *et al.*, "Teaching computer organization/architecture with limited resources using simulators," *SIGCSE Bulletin*, vol. 34, no. 1, pp. 176–180, 2002.

[3] M. Oskin, "The revolution inside the box," *Communications of the ACM*, vol. 51, no. 7, pp. 70–78, 2008.

[4] R. J. O. Figueiredo *et al.*, "Archer: A community distributed computing infrastructure for computer architecture research and education," *Collaborative Computing: Networking, Applications and Worksharing*, vol. 10, pp. 181–192, 2009.

[5] M. Litzkow, M. Livny, and M. Mutka, "Condor- A Hunter of Idle Worksatations," in *ICDCS '88: Proceedings of the 8th IEEE International Conference on Distributed Computing Systems*, 1988.

[6] P. S. Magnusson *et al.*, "Simics: A full system simulation platform," *Computer*, vol. 35, no. 2, pp. 50–58, 2002.

[7] (2010, April) Archer. [Online]. Available: http://archer-project.org