AMUN: A Practical Application Using the NILE Control System

R. Baker[1], *L. Zhou*[2], *J. Duboscq*[3]

[1] Cornell University, Ithaca, NY USA
[2] University of Florida, Gainesville, FL USA
[3] The Ohio State University, Columbus, OH USA

Abstract

AMUN (Advanced Monte Carlo Under NILE) is an application of the NILE Control System being used to generate Monte Carlo data for the CLEO experiment. The task of MC generation is divided into many small subtasks and distributed among a heterogeneous farm of processors. Individual processors can be added to or removed from the farm easily, allowing CLEO collaborators to loan their CPU power to the AMUN effort when the machines would otherwise be idle. Using the aggregated power of between 10 and 60 nodes, we have generated as many as 1 million CLEO II.V tau pair events per day.

Keywords: Data Handling, Data Management, Distributed Computing, NILE, CLEO, AMUN, Farm, Integration

1 Introduction

The NILE Project was begun in 1994 to develop modern distributed computing solutions for high energy physics experiments. In this paper, we report on a successful application of the NILE Control System for generation of CLEO Monte Carlo simulations. Event simulation is an ideal application for distributed computing because the task is easily divided into smaller, independent tasks. The processors used for this application are an ad-hoc collection of up to sixty Alpha processors, including some older, slow machines and newer, faster machines with processor speeds varying from 0.94 to 18.4 SpecInt95.

2 The NILE Control System

The NILE Control System schedules and manages the execution of a set of jobs over a variable, heterogeneous group of processors. The Control System is a controlling JAVA application which runs on one main processor and subordinate JAVA applications which run on every other processor. Inter-processor communication is handled via RMI. The control system periodically checks the status and availablility of the other processors.

The primary task of the control system is to manage the list of jobs to be run. Each job is divided into sub-jobs which are assigned to the available processors. In the case of Monte Carlo simulation, the division into sub-jobs is simply a matter of computing the number of events to be generated by each sub-job. The completion status of each sub-job is monitored and failed sub-jobs can be retried up to a user defined maximum number of retries. After the completion of all sub-jobs, a post processor is run to collect all of the generated events into a single output file.

3 Processor List

The collection of processors available for AMUN is constantly changing. The owner of a machine can make it available for AMUN on a set schedule so that the CPU can be used for Monte Carlo generation when it would otherwise be idle. In addition to predefined scheduling, machines can be added to or removed from the AMUN pocessor list interactively with simple command line entries. This flexible machine use scheduling minimizes the impact on users who voluntarily contribute their CPU time. This arrangement helps to optimize the utilization of the CPU resources at the laboratory. When a processor is removed from the processor list, the current sub-job is allowed a few minutes to complete to avoid causing an undue number of sub-job failures.

4 Experience and Performance

More than 25 million CLEO II.V Monte Carlo events have been generated by the AMUN system in approximately three months of operation, making a very significant contribution to the CLEO experiment. Depending on the number of machines available, total production has been as high as 1 million events per day. CPU owners have been very cooperative and have not experienced any significant inconvenience because of AMUN use of their machines. The NILE Control System has been very stable, running for weeks at a time without intervention.

As figure 1 shows, the performance of the AMUN system scales linearly with available CPU power over a wide range. The data for this figure were obtained while the AMUN farm was generating CLEO II.V τ lepton pair events. During the collection of this data, the number of processors in the farm varied between 9 and 24.

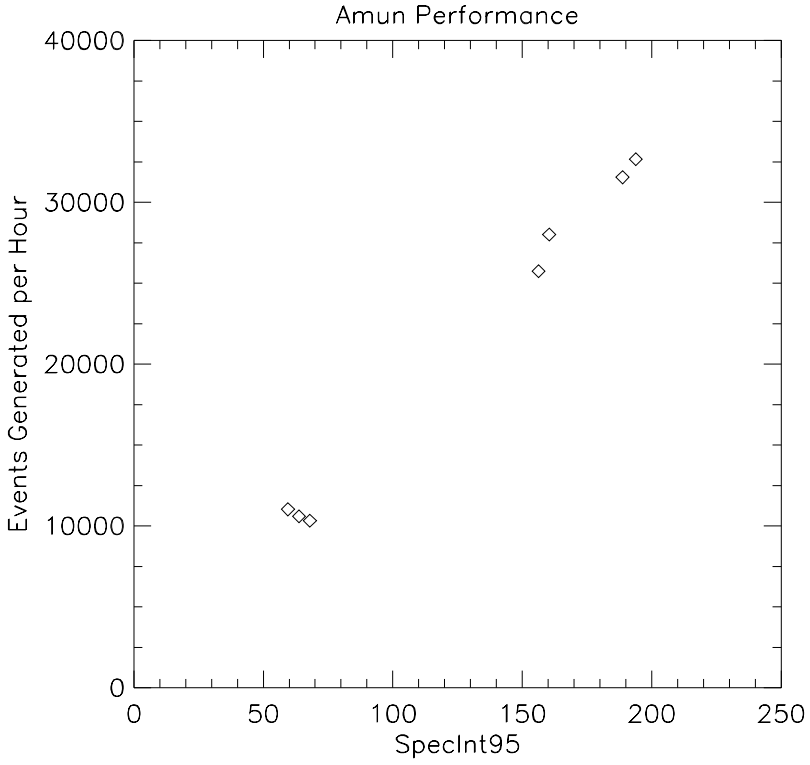

Figure 1: Event generation rate as a function of total farm CPU measured in SpecInt95 units. Events are CLEO II.V τ lepton pair decays.

5 Conclusions

AMUN (Advanced Monte Carlo Under Nile) is an application of the NILE Control System to the problem of CLEO Monte Carlo generation. The AMUN system has demonstrated that the NILE Control System is an efficient, scalable solution for distributed computing with a heterogeneous and variable set of processors. The Monte Carlo generated with AMUN has made a significant contribution to the CLEO experiment using CPU time that otherwise would have gone unused. For more information on the NILE Project, please see `http://www.nile.cornell.edu/`.