

Phase Analysis and Prediction for Dynamic Resource Provisioning

Jian Zhang and Renato Figueiredo
ACIS Lab, University of Florida

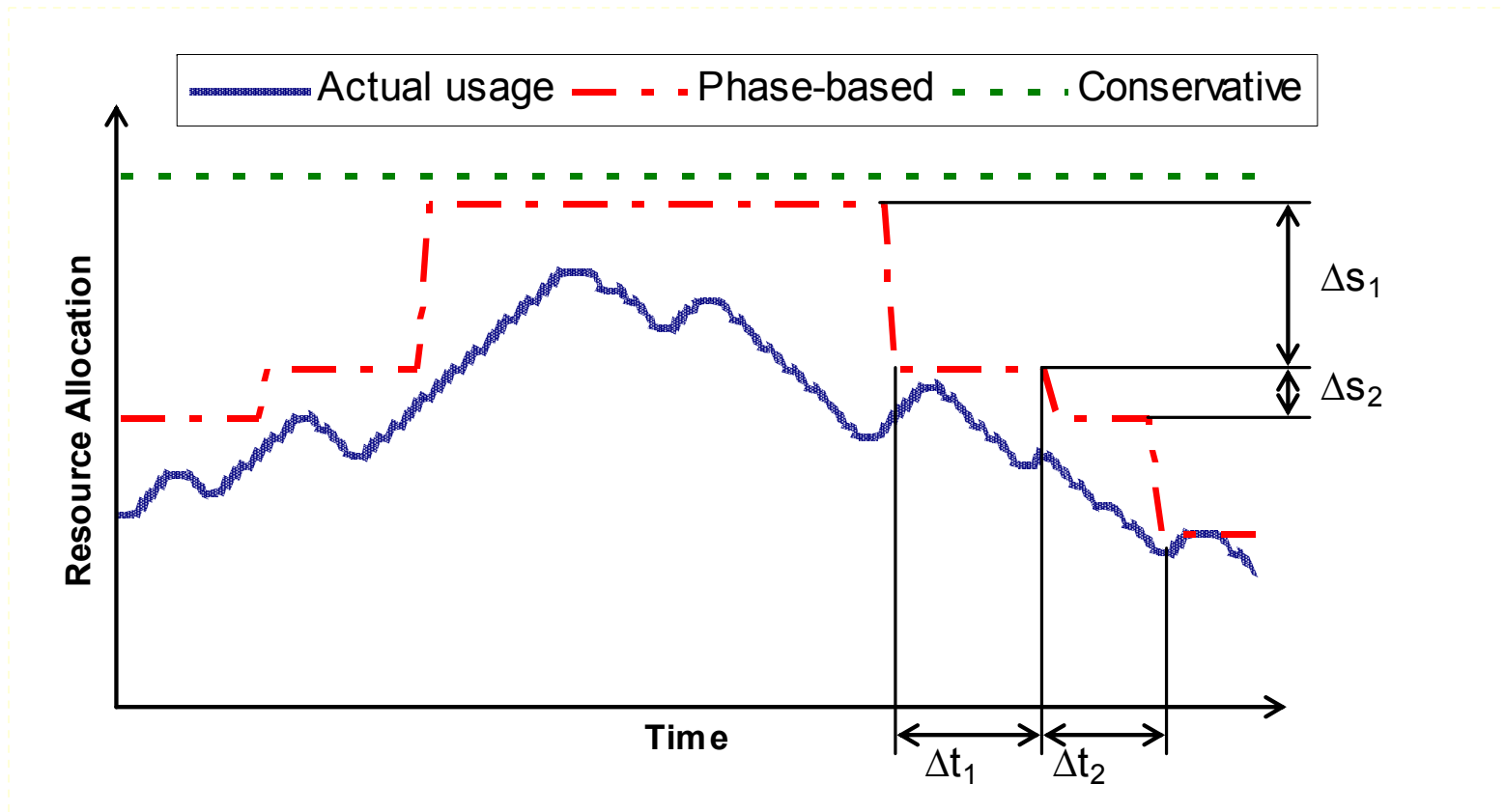
Mazin Yousif and Robert Carpenter
Intel Corporation

June 12, 2007

Motivation

- Provide computing resources as a utility and charge the users for a specific usage
 - Exp. Amazon's Elastic Compute Cloud (EC2)
 - User incentive: Request no more than the amount of resources that an application needs
- Present a need to adapt the resource provisioning to the changing workload
→ minimize the costs to the users

Problem Statement

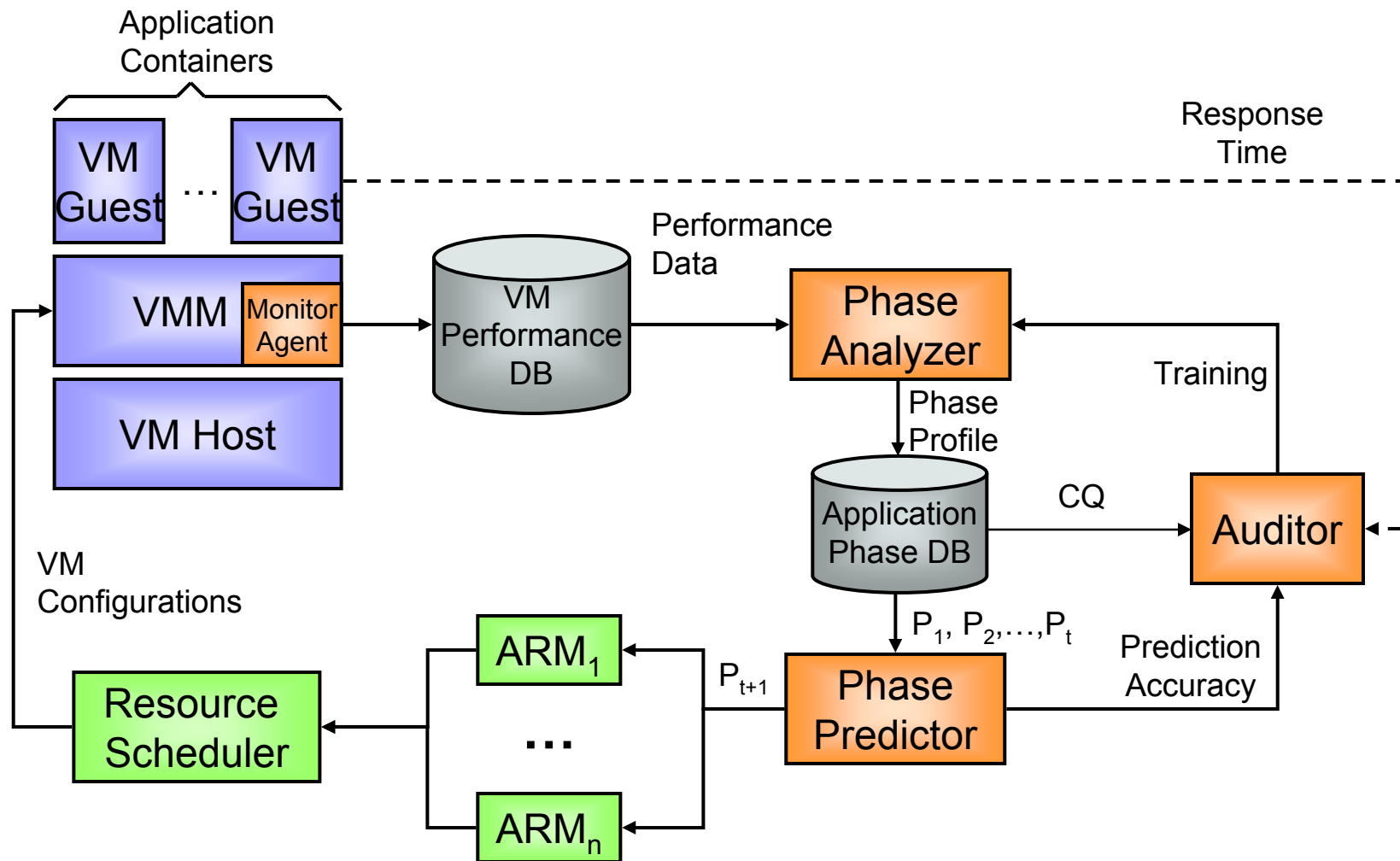


How to determine the *time* and *space* granularities of the resource allocation?

Approach

- Use *virtual machines (VMs)* as *resource container* to multiplex data center resources
- Develop analytical tools to automatically discover the *similarities* and *changes* in the application's resource consumption
 - **Phase:** A set of intervals within an application's execution that have similar system-level resource consumption behavior
 - **Phase analysis:** Use *k-mean clustering* based algorithms to partition a set of intervals into phases and determine the best number of phases
 - **Phase prediction:** Use time series prediction and *classification* algorithms to predict future execution phases

Learning Aided Application Phase Characterization Prototype

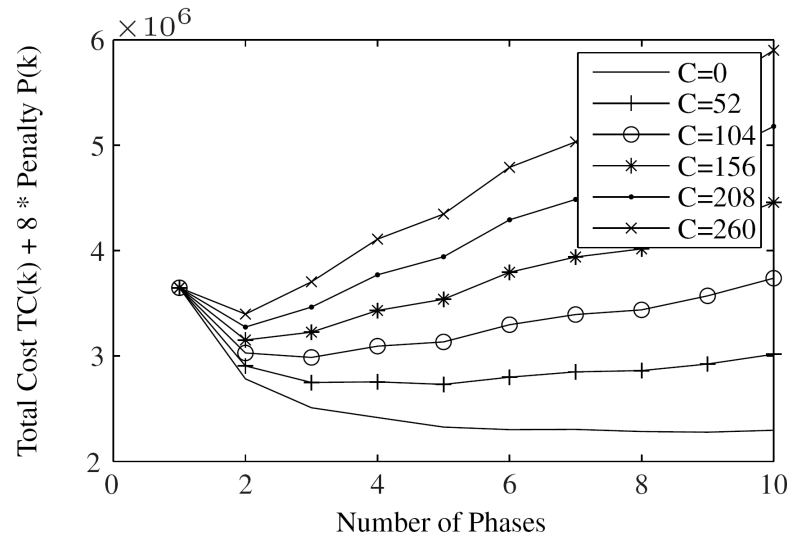


VMM: Virtual Machine Monitor
ARM: Application Resource Manager

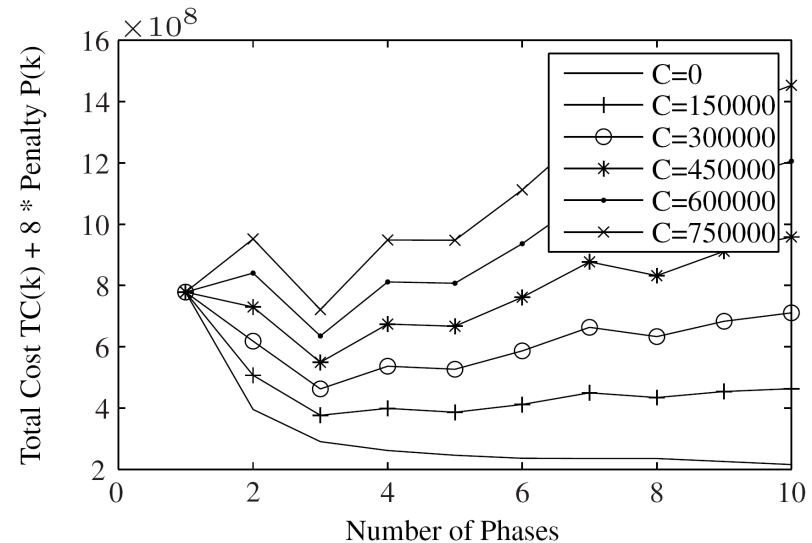
VM: Virtual Machine
CQ: Clustering Quality

DB: Database
 P_t : Phase ID at time t

Experimental Results



(a) CPU_user of SPECseis96



(b) Bytes_out of WC'98 log replay

Figure Phase Profile ($C_p=8$)

- Cost Model:

$$TC(k) = R(k) + C * TR(k) + C_p * P(k)$$

$R(k)$ -- Resource cost

$TR(k)$ -- Re-provisioning cost

$P(k)$ -- Mis-prediction penalty $P(k)$

C and C_p -- The relative unit cost of phase transition, and mis-prediction penalty with regard to the unit cost of resource reservation

Thank you