POLITECNICO DI MILANO

Dipartimento di Elettronica e Informazione

AN ENERGY-AWARE APPROACH FOR SERVICE PERFORMANCE EVALUATION

INTRODUCTION

The main idea presented here takes into account servers energy consumption to drive workload distribution, in which, services can be dynamically routed to different servers. For that we consider the intrinsic data centre servers heterogeneity, i.e. different computational power and energy consumption characteristics, classifying them from the slowest to the fastest at different energy consumption levels.

As each service has specific quality constraints to be satisfied, well described by Service Level Agreement (SLA), it is reasonable to think that some services do not need to be executed always into the fastest class of server. Services with non-critical performance constraints, such as response time, could be executed within slower servers to consume less power and to reduce their overall cost/price. Aiming to verify this dependency and to evaluate the impact of energy efficiency constraints in service centers, we present a system that is modeled as a closed queueing network with multiple-servers stations.

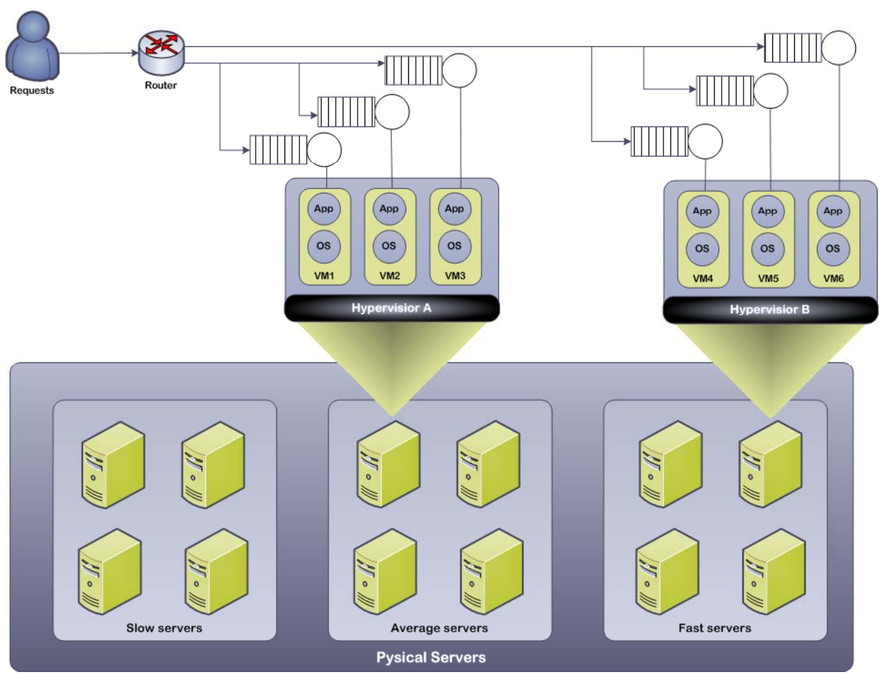


Fig.1 - Proposed model

MODEL

In order to perform analytic evaluations, we use queueing network modeling approach. System resources (such as hardware resources and their software queues) are represented as network of queues and different performance metrics is computed in simulated environments using Mean-Value Analysis (MVA) algorithm.

We start identifying the main components of the system and the interaction between hardware resources and software queues, which is done through Virtual Machines (VM). Each VM is represented as one server in our experiments and we assume the statement “One application - one server”. This means that one VM can only be assigned to one service at a given instance of time. We present the first set of simulated results using Java Modeling Tools (JMT), an open-source suite that supports common activities for performance evaluation.

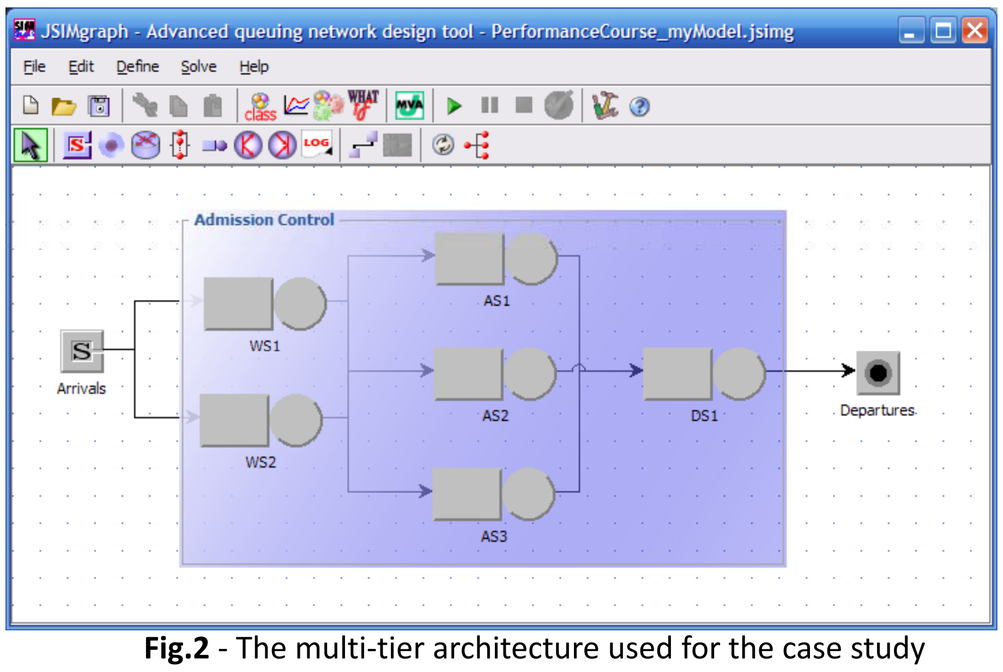


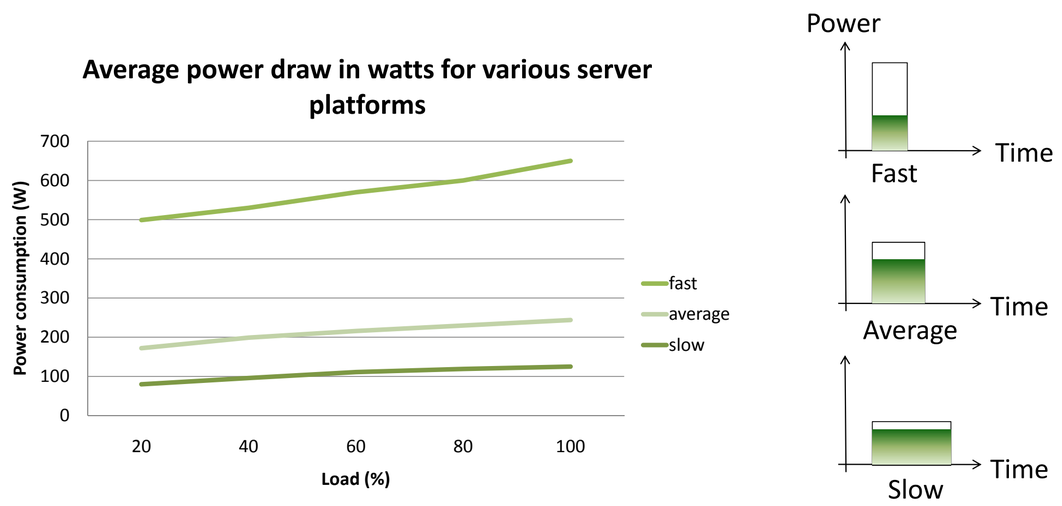
Fig.2 - The multi-tier architecture used for the case study

SIMULATION

The purpose of the simulation is to compare the differences between two different optimal network loads:

- *Experiment A*: considering only performance, which means, selecting always the server class with highest performance and highest energy consumption;
- *Experiment B*: taking into account of energy consumption of each class of servers, which means, selecting different classes through dynamic routing strategies. For both experiments we use an exponential distribution arrival rate λ_{ir} to model a Poisson process where the requests are independent from each other.

The total physical server energy consumption is split properly among its VMs based on components utilization rates. The evaluated scenario is a typical multi-tier web system model composed by a set of multi-class multi-server queues grouped as web servers (WS), application servers (AS), and database servers (DS) and divided into classes according to their characteristics.



ENERGY EVALUATION

In order to compare both experiments from energy point of view, we do need to put them in terms of energy consumption. For that, all servers utilization values are mapped into the server class and their utilization modes together with the number of servers in a specific class of server. we sum up the amount of energy consumed by each class of server according to their utilization mode and multiply it by the number of servers into that situation.

ALEXANDRE MELLO FERREIRA
PHD STUDENT
ferreira@elet.polimi.it

UNIVERSITÄT PASSAU
e 2010 Energy
www.e-energy-conf.org
1st Int'l Conf. on Energy-Efficient Computing and Networking
April 13-15, 2010
University of Passau, Germany

GAMES
Green Active Management of Energy in IT Service centres