

DCWoRMS - a tool for simulation of energy efficiency in Grids and Clouds

Krzysztof Kurowski, Ariel Oleksiak, Wojciech Piatek, Tomasz Piontek

Abstract

Keywords:

1. Introduction

...

The remaining part of this paper is organized as follows. In Section 2 we give a brief overview of the current state of the art concerning modeling and simulation of distributed systems, like Grids and Clouds, in terms of energy efficiency. Section 3 introduces the main features of DCWoRMS. In particular, it introduces our approach to workload and resource management, presents the concept of energy efficiency modeling and explains how to incorporate a specific application performance model into simulations. Section 4 discusses energy models adopted within the DCWoRMS. In Section 5 we present some experiments that were performed on our testbed and then repeated using DCWoRMS to evaluate the correctness of the simulation environment. Section 6 focuses on the role of DCWoRMS within the CoolE-mAll project. Final conclusions and directions for future work are given in Section 7.

2. Related Work

The growing importance of energy efficiency in information technologies led to significant interest in energy saving methods for computing systems. Therefore, intelligent resource management policies are gaining popularity when considering the energy efficiency of IT infrastructures. Nevertheless, studies of impact of scheduling strategies on energy consumption require a large effort and are difficult to perform in real distributed environments. To

overcome these issues extensive research has been conducted in the area of modeling and simulation tools. As a result, a wide variety of simulation tools emerged. The following section contains a short summary of existing simulators that address the green computing issues in distributed infrastructures.

2.1. GreenCloud

GreenCloud [5] is a C++ based simulation environment for energy-aware data cloud computing data centers. It was developed as an extension of the NS2 network simulator. GreenCloud allows researchers to observe and evaluate data centers performance and study their energy-efficiency, focusing mainly on the communications within a data center. Along with the workload distribution, it offers users a detailed, fine-grained modeling of the energy consumed by the elements of the data center.

To deliver information about the energy usage, GreenCloud distinguishes three energy consumption components: computing energy, communicational energy, and the energy component related to the physical infrastructure of a data center. This approach enables modeling energy usage associated with computations, network operations and cooling systems. In GreenCloud, the energy models are implemented for every simulated data center entity (computing servers, core and rack switches). Moreover, due to the advantage in the simulation resolution, energy models can operate at the network packet level as well. This allows updating the levels of energy consumption whenever a new packet leaves or arrives from the link, or whenever a new task execution is started or completed at the server. Servers are modeled as single core nodes that are responsible for task execution and may contain different scheduling strategies. The server power consumption model implemented in GreenCloud depends on the server state and its utilization and allows capturing the effects of both of the Dynamic Voltage and Frequency Scaling (DVFS) and Dynamic Power Management (DPM) schemes. At the links and switches level, GreenCloud supports Dynamic Voltage Scaling (DVS) and Dynamic Network Shutdown (DNS) techniques. The DVS method introduced a control element at each port of the switch that - depending on the traffic pattern and current levels of link utilization - could downgrade the transmission rate. DNS approach allows putting some network equipment into sleep mode.

To cover the vast majority of cloud computing applications, GreenCloud defines three types of workloads: computationally intensive workloads that load computing servers considerably, data-intensive workloads that require heavy data transfers, and finally balanced workloads which aim to model the

applications having both computing and data transfer requirements. GreenCloud describes application with a number of computational requirements. Moreover, it specifies communication requirements of the applications in terms of the amount of data to be transferred before and after a task completion. The execution of each application requires a successful completion of its two main components: computing and communicational. In addition time constraints can be taken into account during the simulation by adding a predefined execution deadline, which aims at introducing Quality of Service constraints specified in a Service Level Agreement. Nevertheless, GreenCloud does not support application performance modeling. Aforementioned capabilities allow only incorporating simple requirements that need to be satisfied before and during task execution.

Contrary to what the GreenCloud name may suggest, it does not allow testing the impact of a virtualization-based approach on the resource management. GreenCloud simulator is released under the General Public License Agreement.

2.2. *CloudSim*

CloudSim [1] is an event-based simulation tool written in Java. Initially CloudSim was based on the well-known GridSim framework, however since the last few releases it is an independent simulator and does not benefit from most of the GridSim functionality.

CloudSim allows creating a simple resource hierarchy containing computing resources that consist of machines and processors. Additionally, it may simulate the behavior of other components including storage and network resources. However, it focuses on computational resources and provides an extra virtualization layer that acts as an execution, management, and hosting environment for application services. It is responsible for the VM provisioning process as well as managing the VM life cycle such as: VM creation, VM destruction, and VM migration. It also enables evaluation of different economic policies by modeling the cost metrics related to the SaaS and IaaS models.

The CloudSim framework provides basic models and entities to validate and evaluate energy-conscious provisioning of techniques and algorithms. Each computing node can be extended with a power model that simulates the power consumption. CloudSim offers example implementations of this component that characterize some popular server models. Needless to say, it

can be easily extended for simulating user-defined power consumption models. That allows estimating the current power usage according to the current utilization level or the host model. This capability enables the creation of energy-conscious provisioning policies that require real-time knowledge of power consumption by Cloud system components. Furthermore, it allows an accounting of the total energy consumed by the system during the simulation period. CloudSim comes with a set of predefined and extendable policies that manage the process of VM migrations in order to optimize the power consumption. However, the proposed solution is not appropriate for more sophisticated power management policies. In particular, CloudSim is not sufficient for modeling frequency scaling techniques and managing resource power states.

Similar to GreenCloud, CloudSim defines a simple application model that includes computational and data requirements. Although all these constraints are taken into account during scheduling, they do not affect the application execution. Thereby, a researcher is required to put a lot of effort to incorporate an application performance model into his experiments. On the other hand CloudSim offers modeling of utilization models that are used to estimate the current load of processor, bandwidth and memory and can be taken into account during the task allocation process. Concerning workloads, simulator is able to partially support SWF [10] files and read data in a user-defined file format. Moreover, it can handle a wide variety of workload types, including parallel, and pre-emptive jobs

CloudSim is available as Open Source under GPL license.

2.3. DCSG Simulator

DCSG Simulator [2] is a Data Centre Cost and Energy Simulator that has been developed under the Carbon Trust Low Carbon Collaborations program in conjunction with the BCS and Romonet Ltd. The simulator works at both a data center infrastructure level where analysis of the achieved efficiency of the data center mechanical and electrical plant can be performed but also at the IT level. The simulator implements a set of basic rules that have been developed, based on a detailed understanding of the data center as a system, to allow cost and energy use to be usefully allocated to IT devices within the data center.

As far as data center infrastructure level is concerned, DCSG Simulator is calculates the power and cooling schema of data center equipment with

respect to their performance. User is able to take into account a wide variety of mechanical and electrical devices like: transformers, power distribution units, power supply, cabling, computer room air conditioning units and chiller plant. For each of them a numerous factors can be defined, including device capacity and efficiency, load operating points. These data can be derived from a generic list as well as from the information given by particular manufacturers. There is a wide range of pre-defined models, but user can easily extend them or create a new ones.

To perform an IT simulation, it is possible to extend the data center infrastructure by putting IT devices into that data center. That enables detailed simulation of the energy efficiency of devices across a specified time period. In this case performance of each piece of equipment (facility and IT) within a data center is determined by a combination of factors, including workload, data center conditions, the manufacturer's specifications of the machine's components and the way in which the machine is utilized based on its provisioned IT load. IT is possible to bind the operational characteristics, proper to the particular geographic locations, with the simulation process. These characteristics may include temperature profile as well as the power cost that vary depending on the time and place. The output of this simulation is a set of energy and cost data representing the IT device and data center energy consumption, capital and operational costs.

According to the tool evaluation presented in [3] an accuracy of models delivered by Romonet is at the level of 95% when compared with metered data. The simulator is available under an OSL V3.0 open-source license, however it can be only accessed by the DCSG Members.

3. DCWoRMS

The following picture (Figure 1) presents the overall architecture of the simulation tool.

Data Center workload and resource management simulator (DCWoRMS) is a simulation tool based on the GSSIM framework [6] developed by Poznan Supercomputing and Networking Center (PSNC). GSSIM has been proposed to provide an automated tool for experimental studies of various resource management and scheduling strategies in distributed computing systems. DCWoRMS extends its basic functionality and add some additional features related to the energy efficiency issues in data centers. In this section we

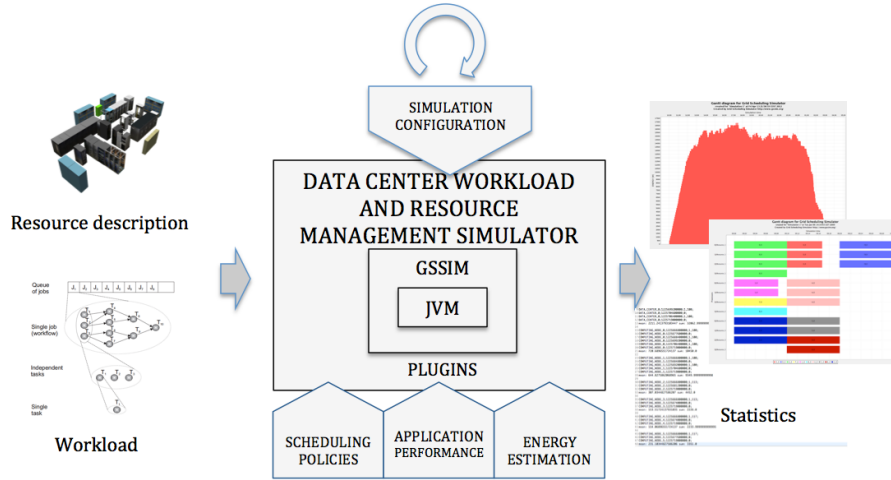


Figure 1: DCWoRMS architecture

will introduce the functionality of the simulator, in terms of modeling and simulation of large scale distributed systems like Grids and Clouds.

3.1. Architecture

DCWoRMS is an event-driven simulation tool written in Java. In general, input data for the DCWoRMS consist of workload and resources descriptions. They can be provided by the user, read from real traces or generated using the generator module. However, the key elements of the presented architecture are plugins. They allow the researchers to configure and adapt the simulation environment to the peculiarities of their studies, starting from modeling job performance, through energy estimations up to implementation of resource management and scheduling policies. Each plugin can be implemented independently and plugged into a specific experiment. Results of experiments are collected, aggregated, and visualized using the statistics module. Due to a modular and plug-able architecture DCWoRMS can be applied to specific resource management problems and addressing different users requirements.

3.2. Workload modeling

As it was said, experiments performed in DCWoRMS require a description of applications that will be scheduled during the simulation. As a primary

definition, DCWoRMS uses files in the Standard Workload Format (SWF) or its extension the Grid Workload Format (GWF) [8]. In addition to the SWF file, some more detailed specification of a job and tasks can be included in an auxiliary XML file. This form of description provides the scheduler with more detailed information about application profile, task requirements, user preferences and execution time constraints, which are unavailable in SWF/GWF files. To facilitate the process of adapting the traces from real resource management systems, DCWoRMS supports reading those delivered from the most common ones like SLURM [9] and Torque [11]. Since the applications may vary depending on their nature in terms of their requirements and structure, DCWoRMS provides user flexibility in defining the application model. Thus, considered workloads may have various shapes and levels of complexity that range from multiple independent jobs, through large-scale parallel applications, up to whole workflows containing time dependencies and preceding constraints between jobs and tasks. Each job may consist of one or more tasks and these can be seen as a group of processes. Moreover, DCWoRMS is able to handle rigid and moldable jobs, as well as pre-emptive ones. To model the application profile in more detail, DCWoRMS follows the DNA approach proposed in [4]. Accordingly, each task can be presented as a sequence of phases, which shows the impact of this task on the resources that run it. Phases are then periods of time where the system is stable (load, network, memory) given a certain threshold and. Each phase is linked to values of the system that represent a resource consumption profile. Such a stage could be for example described as follows: 60% CPU, 30%net, 10%mem

Levels of information about incoming jobs are presented in Figure 2.

This form of representation allows users to define a wide range of workloads: HPC (long jobs, computational-intensive, hard to migrate) or virtualization (short requests) typical for cloud computing environments. Further, the DCWoRMS benefits from the GSSIM workload generator tool and extends it with that allows creating synthetic workloads.

3.3. Resource modeling

The main goal of DCWoRMS is to enable researchers evaluation of various resource management policies in diverse computing environments. To this end, it supports flexible definition of simulated resources both on physical (computing resources) as well as on logical (scheduling entities) level. This flexible approach allows modeling various computing entities consisting of compute nodes, processors and cores. In addition, detailed location

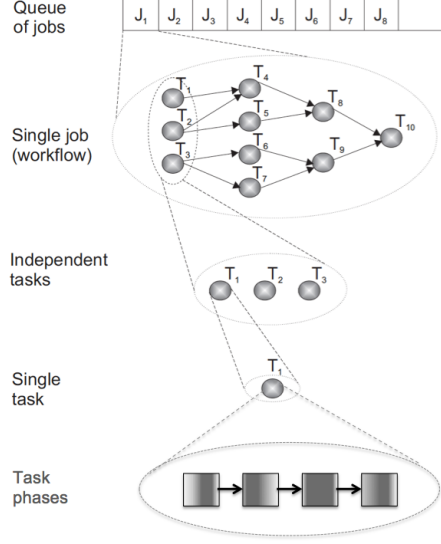


Figure 2: Levels of information about jobs

of the given resources can be provided in order to group them and arrange into physical structures such as racks and containers. Each of the components may be described by different parameters specifying available memory, storage capabilities, processor speed etc. In this way, it is possible to describe power distribution system and cooling devices. Due to an extensible description, users are able to define a number of experiment-specific and visionary characteristics. Moreover, with every component, dedicated profiles can be associated that determines, among others, power, thermal and air throughput properties. The energy estimation plugin can be bundled with each resource. This allows defining various power models that can be then followed by different computing system components. Details concerning the approach to energy-efficiency modeling in DCWoRMS can be found in the next sections.

Scheduling entities allow providing data related to the brokering or queuing system characteristics. Thus, information about available queues, resources associated with them and their parameters like priority, availability of AR mechanism etc. can be defined. Moreover, allocation policy and task scheduling strategy for each scheduling entity can be introduced in form of the reference to an appropriate plugin. DCWoRMS allows building a hier-

archy of schedulers corresponding to the hierarchy of resource components over which the task may be distributed.

In this way, the DCWoRMS supports simulation of a wide scope of physical and logical architectural patterns that may span from a single computing resource up to whole data centers or geographically distributed grids and clouds. In particular, it supports simulating complex distributed architectures containing models of the whole data centers, containers, racks, nodes, etc. In addition, new resources and distributed computing entities can easily be added to the DCWoRMS environment in order to enhance the functionality of the tool and address more sophisticated requirements. Granularity of such topologies may also differ from coarse-grained to very fine-grained modeling single cores, memory hierarchies and other hardware details.

3.4. Energy management concept in DCWoRMS

The DCWoRMS allows researchers to take into account energy efficiency and thermal issues in distributed computing experiments. That can be achieved by the means of appropriate models and profiles. In general, the main goal of the models is to emulate the behavior of the real computing resources, while profiles support models by providing data essential for the power consumption calculations. Introducing particular models into the simulation environment is possible through choosing or implementation of dedicated energy plugins that contain methods to calculate power usage of resources, their temperature and system air throughput values. Presence of detailed resource usage information, current resource energy and thermal state description and a functional energy management interface enables an implementation of energy-aware scheduling algorithms. Resource energy consumption and thermal metrics become in this context an additional criterion in the resource management process. Scheduling plugins are provided with dedicated interfaces, which allow them to collect detailed information about computing resource components and to affect their behavior. The following subsections present the general idea behind the energy-efficiency simulations.

3.4.1. Power management

The motivation behind introducing a power management concept in DCWoRMS is providing researchers with the means to define the energy efficiency of resources, dependency of energy consumption on resource load and specific applications, and to manage power modes of resources. Proposed solution extends the power management concept presented in GSSIM [7] by

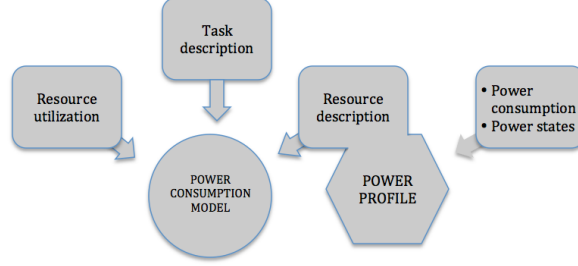


Figure 3: Energy consumption modeling

offering a much more granular approach with the possibility of plugging energy consumption models and power profiles into each resource level.

Power profile. In general, power profiles allow specifying the power usage of resources. Depending on the accuracy of the model, users may provide additional information about power states which are supported by the resources, amounts of energy consumed in these states, and other information essential to calculate the total energy consumed by the resource during runtime. In such a way each component of IT infrastructure may be described, including computing resources, system components and data center facilities. Moreover, It is possible to define any number of new, resource specific, states, for example so called P-states, in which processor can operate.

Energy consumption model. The main aim of these models is to emulate the behavior of the real computing resource and the way it consumes energy. Due to a rich functionality and flexible environment description, DCWoRMS can be used to verify a number of theoretical assumptions and develop new energy consumption models. Modeling of energy consumption is realized by the energy estimation plugin that calculates energy usage based on information about the resource power profile, resource utilization, and the application profile including energy consumption and heat production metrics. Relation between model and power profile is illustrated in Figure 3.

Power management interface. DCWoRMS is complemented with an interface that allows scheduling plugins to collect detailed power information about computing resource components and to change their power states. It enables performing various operations on the given resources, including dynamically changing the frequency level of a single processor, turning off/on

computing resources etc. The activities performed with this interface find a reflection in total amount of energy consumed by the resource during simulation.

Presence of detailed resource usage information, current resource energy state description and functional energy management interface enables an implementation of energy-aware scheduling algorithms. Resource energy consumption becomes in this context an additional criterion in the scheduling process, which use various techniques to decrease energy consumption, e.g. workload consolidation, moving tasks between resources to reach full load on one resource and zero load on the other or to balance the load, dynamic power management, cutting down CPU frequency, and others.

3.4.2. Air throughput management concept

The presence of an air throughput concept addresses the issue of resource air-cooling facilities provisioning. Using the air throughput profiles and models allows anticipating the air flow level on output of the computing system component, resulting from air-cooling equipment management.

Air throughput profile. The air throughput profile, analogously to the power profile, allows specifying supported air flow states. Each air throughput state definition consists of an air flow value and a corresponding power draw. It can represent, for instance, a fan working state. An air throughput value can also express a fan rotation speed. In this way, associating the air throughput profile with the given computing resource, it is possible to describe mounted air-cooling devices. Possibility of introducing additional parameters makes the air throughput description extensible for new specific characteristics.

Air throughput model. Similar to energy consumption models, the user is provided with a dedicated interface that allows him to describe the resulting air throughput of the computing system components like cabinets or server fans. The general idea of the air throughput modeling is shown in Figure 4. Accordingly, air flow estimations are based on detailed information about the involved resources, including their air throughput states.

Air throughput management interface. The DCWoRMS delivers interfaces that provide access to the air throughput profile data, allows acquiring detailed information concerning current air flow conditions and changes in air

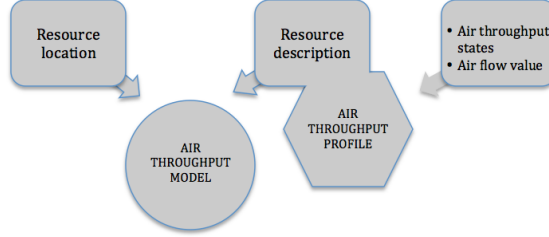


Figure 4: Air throughput modeling

flow states. The availability of these interfaces support evaluation of different cooling strategies.

3.4.3. Thermal management concept

The primary motivation behind the incorporation of thermal aspects in the DCWoRMS is to exceed the commonly adopted energy use-cases and apply more sophisticated scenarios. By the means of dedicated profiles and interfaces, it is possible to perform experimental studies involving temperature-aware workload placement.

Thermal profile. Thermal profile expresses the thermal specification of resources. It consists of the definition of the thermal design power (TDP), thermal resistance and thermal states that describe how the temperature depends on dissipated heat. For the purposes of more complex experiments, introducing of new, user-defined characteristics is supported. The aforementioned values may be provided for all computing system components distinguishing them, for instance, according to their material parameters and models.

Temperature estimation model. Thermal profile, complemented with the temperature measurement model implementation may introduce temperature sensors simulation. In this way, users have means to approximately predict the temperature of the simulated objects. The proposed approach assumes some simplifications that ignore heating and cooling processes.

Figure 5 summarizes relation between model and profile and input data.

Thermal resource management interface. As the temperature is highly dependent on the dissipated heat and cooling capacity, thermal resource management is performed via a power and air throughput interface. Nevertheless,

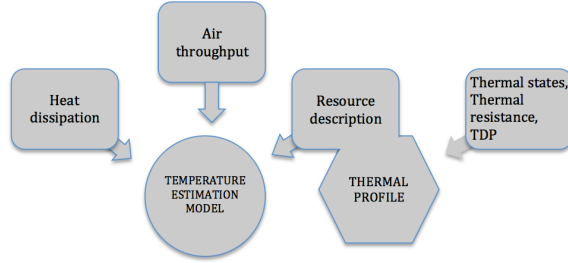


Figure 5: Temperature estimation modeling

the interface provides access to the thermal resource characteristics and the current temperature values

3.5. Application performance modeling

In general, DCWoRMS implement user application models as objects describing computational, communicational and energy requirements and profiles of the task to be scheduled. Additionally, simulator provides means to include complex and specific application performance models during simulations. They allow researchers to introduce specific ways of calculating task execution time. These models can be plugged into the simulation environment through a dedicated API and implementation of an appropriate plugin. To specify the execution time of a task user can apply a number of parameters, including:

- task length (number of CPU instructions)
- task requirements
- detailed description of allocated resources (processor type and parameters, available memory)
- input data size
- network parameters

Using these parameters developers can for instance take into account the architectures of the underlying systems, such as multi-core processors, or virtualization overheads, and their impact on the final performance of applications.

4. Modelling of energy efficiency in DCWoRMS

To facilitate the simulation process, DCWoRMS provides some basic implementation of power consumption and air throughput models.

4.1. Power consumption models

The energy consumption models provided by default can be classified into the following groups, starting from the simplest model up to the more complex ones. Users can easily switch between the given models and incorporate new, visionary scenarios.

Static approach. is based on a static definition of resource power usage. This model calculates the total amount of energy consumed by the computing resource system as a sum of energy, consumed by all its components (processors, disks, power adapters, etc.). More advanced versions of this approach assume definition of resource states along with corresponding power usage. This model follows changes of resource power states and sums up the amounts of energy defined for each state.

Resource load. model extends the static power state description and enhances it with real-time resource usage, most often simply the processor load. In this way it enables a dynamic estimation of power usage based on resource basic power usage and state (defined by the static resource description) as well as resource load. For instance, it allows distinguishing between the amount of energy used by idle processors and processors at full load. In this manner, energy consumption is directly connected with power state and describes average power usage by the resource working in a current state.

Application specific. model allows expressing differences in the amount of energy required for executing various types of applications at diverse computing resources. It considers all defined system elements (processors, memory, disk, etc.), which are significant in total energy consumption. Moreover, it also assumes that each of these components can be utilized in a different way during the experiment and thus have different impact on total energy consumption. To this end, specific characteristics of resources and applications are taken into consideration. Various approaches are possible including making the estimated power usage dependent on defined classes of applications, ratio between CPU-bound and IO-bound operations, etc.

4.2. Air throughput models

The DCWoRMS comes with the following predefined models. By default, air throughput estimations are performed according to the first one.

Static. model refers to a static definition of air throughput states. According to this approach, output air flow depends only on the present air cooling working state and the corresponding air throughput value. Each state change triggers the calculations and updates the current air throughput value. This strategy requires only a basic air throughput profile definition.

Space. model allows taking into account a duct associated with the investigated air flow. On the basis of the given fan rotation speed and the obstacles before/behind the fans, the output air throughput can be roughly estimated. Thus, it is possible to estimate the air flow level not only referring to the current fan operating state but also with respect to the resource and its sub-component placement. More advanced scenario may consider mutual impact of several air flows.

4.3. Thermal models

The following models are supported natively. By default, the static strategy is applied.

Static. approach follows the changes in heat, generated by the computing system components and matches the corresponding temperature according to the specified profile. Since it tracks the power consumption variations, corresponding values must be delivered, either from power consumption model or on the basis of user data. Replacing the appropriate temperature values with function based on the defined material properties and/o experimentally measured values can easily extend this model.

Ambient. model allows taking into account the surrounding cooling infrastructure. It calculates the device temperature as a function adopted from the static approach and extends it with the influence of cooling method. The efficiency of cooling system may be derived from the current air throughput value.

5. Experiments and evaluation

Results + RECS and MOP description

....

In this section, we present computational analysis that were conducted to emphasize the role of modelling and simulation in studying computing systems performance. We carried out two types of experiments. The former one aimed at demonstrating the capabilities of the simulator in terms of verifying the research hypotheses. The latter set of experiments was performed CoolEmAll testbed and then repeated using DCWoRMS tool. The comparative analysis of obtained results shows the reproducibility of experiments and prove the correctness of .

5.1. Testbed description

The RECS Cluster System is an 18 node computer system that has an monitoring and controlling mechanism integrated. Through the integrated novel monitoring approach of the RECS Cluster System the network load can be reduced, the dependency of polling every single compute node at operation system layer can be avoided. Furthermore this concept build up a basis on which new monitoring- and controlling-concepts can be developed. Therefore, each compute node of the RECS Cluster Server is connected to an Operation System independent microcontroller that collects the most important sensor data like temperature, power consumption and the status (on/off) from every single node.

5.2. Computattional analysis

6. DCWoRMS application

DCWoRMS in CoolEmAll, integration with CFD

7. Conclusions and future work

References

- [1] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, Cesar A. F. De Rose, and Rajkumar Buyya, CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms, Software: Practice and Experience (SPE), Volume 41, Number 1, Pages: 23-50, ISSN: 0038-0644, Wiley Press, New York, USA, January, 2011.

- [2] <http://dcs.bcs.org/welcome-dcs-simulator>
- [3] <http://www.datacenterdynamics.com/blogs/ian-bitterlin/it-does-more-it-says-tin%E2%80%A6>
- [4] Ghislain Landry Tsafack Chetsa, Laurent Lefvre, Jean-Marc Pierson, Patricia Stolf, Georges Da Costa. DNA-inspired Scheme for Building the Energy Profile of HPC Systems. In: International Workshop on Energy-Efficient Data Centres, Madrid, Springer, 2012
- [5] D. Kliazovich, P. Bouvry, and S. U. Khan, A Packet-level Simulator of Energy-aware Cloud Computing Data Centers, Journal of Supercomputing, vol. 62, no. 3, pp. 1263-1283, 2012
- [6] S. Bak, M. Krystek, K. Kurowski, A. Oleksiak, W. Piatek and J. Weglarz, GSSIM - a Tool for Distributed Computing Experiments, Scientific Programming Journal, vol. 19, no. 4, pp. 231-251, 2011.
- [7] M. Krystek, K. Kurowski, A. Oleksiak, W. Piatek, Energy-aware simulations with GSSIM. Proceedings of the COST Action IC0804 on Energy Efficiency in Large Scale Distributed Systems, 2010, pp. 55-58.
- [8] <http://gwa.ewi.tudelft.nl/>
- [9] <https://computing.llnl.gov/linux/slurm/>
- [10] Parallel Workload Archive, <http://www.cs.huji.ac.il/labs/parallel/workload/>
- [11] <http://www.adaptivecomputing.com/products/open-source/torque/>