Exploration of secured Workflow scheduling models in cloud environment:
Cost, Makespan, Robustness and Security level requirement aware workflow scheduling
∗ ∗ ∗Secured Workflow Scheduling: SWS∗ ∗ ∗

BUGINGO EMMANUEL[1]

[1]Xiamen University
School of information Science and Engineering
Department of Computer Science
email:emmanuelbugingo@stu.xmu.edu.cn

20th September 2018

Summary

Environment

- Considering the virtual machines from **one provider.**
- Considering the virtual machines from **more than one provider.**

Objectives

- **Single objective and Multi-objective.**
- Optimization of (makespan, Cost,Energy,Resource Utilization)

Constraints

- **Single or Multiple constraints.**
- Budget, Deadline, Risk Rate, Reliability, Security level Requirement

Models/approach types

- Meta-heuristic
- Heuristic
- Hybrid

Workload

- Static
- Dynamic

Category of WS

- Single workflow scheduling
- Multiple workflow scheduling

Services

- Processing
- Storage

Factors

- **Environment :** Hybrid, Multiple cloud and within Datacenter.
- **Application requirements :** with sensitive information.

Domains

- **Most of the domain with sensitive data**(Defense, Bio-informatics, business, etc)

Security

- **Common attacks :**snooping and Change
- **Common solutions :** Confidentiality, Integrity and Authentication
- **Source of threat :** Providers side, User Side and Application side

Security strength and time overhead differ

- **Confidentiality :** RC4, IDEA
- **Integrity :** HMAC-SHA-1, HMAC-MDS
- **Authentication :** MDA, SHA-x, DES

Secured workflow scheduling

Alto stratus project

- Considered executing workflow tasks in Hybrid cloud(public and private).
- To fulfill **Security Level Requirements :** decided tasks (less sensitive) to be executed in public cloud and the most sensitive tasks to be executed in private.
- **Objective : Cost,under Deadline constraint and fulfill Security requirement.**
  D.S.Marcon et al "Workflow specification and scheduling with security constraints in hybrid clouds," 2nd IEEE Latin American Conference on Cloud Computing and Communications.

SOLID

- To reduce the data dependency encryption overhead they have used **task duplication technique. Problem :** generate useless data.
- **Objective : Makespan , cost and fulfill Security level requirement.**
  H.Chen et al "Scheduling for workflows with security-sensitive intermediate data by selective tasks duplication in clouds," IEEE Transactions on Parallel and Distributed Systems.

Secured workflow scheduling

SAWS

- The same as **SOLID** but with main focus on the minimization of start time of each task.
- **Objective : Makespan , cost , Resource utilization and fulfill Security level requirement.**
  X. Zhu et al "Security-aware workflow scheduling with selective task duplication in clouds," in Proceedings of the 24th High Performance Computing Symposium

MOPA

- Uses **divide and conquer technique :** Divide tasks instance into multiple group and then execute each group separately.
- **Objective : Makespan , monetary cost and privacy.**
  Y.Wen et al "Scheduling workflows with privacy protection constraints for big data applications on cloud," Future Generation Computer Systems

Secured workflow scheduling

SCAS

- In other to be able to optimize Multidimensional and Multi constraint problem **SCAS** has employed **PSO.**
- **Objective : Makespan and cost under deadline and budget, fulfill Security level requirement.**

  Z. Li et al A security and cost aware scheduling algorithm for heterogeneous tasks of scientific workflow in clouds," Future Generation Computer Systems.

SABA

- Introduced the concept of immovable dataset. This concept affect the performance and restrict the movement of certain dataset due to their cost and security.
- The concept helps in providing short makespan and security service.
- **Objective : Security level requirement, budget and makespan**

  L. Zeng et al Saba : A security-aware and budget-aware workflow scheduling strategy in clouds," Journal of Parallel and Distributed Computing.

Secured workflow scheduling

CEDP

- Provide awareness over big data application in Hybrid cloud.
- Use cost and Energy aware data placement to reduce cost, access time and energy consumption.
- It is better reducing energy and saving renting monetary cost.
- **Objective : Reduce energy , save monetary cost for renting and fulfill Security level requirement.**

  X. Xiaolong et al Data placement for privacy-aware applications over big data in hybrid clouds," Security and Communication Networks.

Common consideration

- **Alto Stratus Project, FFBAT and CEDP :** Considered Hybrid environment. They deal with deciding which tasks to be scheduled in private and which to be scheduled in public environment.
- **FFBAT and SCAS :** considered Risk Rate as the additional constraint.
- **SOLID and SAWS :** both considered task duplication technique.

workflow scheduling

HEFT

- Selects the task with the highest upward rank value at each step and assigns the selected task to the processor, which minimizes its earliest finish time with an insertion-based approach.
- Schedules the critical tasks onto the processor that minimizes the total execution time of the critical tasks
- **Objective : High performance and Makespan reduction**
  Topcuoglu Haluk, Hariri Salim, Wu Min-you. Performance-effective and low-complexity task scheduling for heterogeneous computing. IEEE transactions on parallel and distributed systems. 2002.

Robustness Policies

- Robust and fault tolerant scheduling algorithm with three multi-objective resource selection policies.(Deal with 3 type of failures : Task failure, workflow Level failure and VM failure)
- **Objective : Minimize makespan and cost under deadline and budget constraints by maximizing the robustness.**
  D. Poola, S. K. Garg, R. Buyya, Y. Yang and K. Ramamohanarao, "Robust Scheduling of Scientific Workflows with Deadline and Budget Constraints in Clouds," 2014 IEEE 28th International Conference on Advanced Information Networking and Applications, Victoria, BC, 2014, pp. 858-865. doi : 10.1109/AINA.2014.105

workflow scheduling

Cost aware models

- Considered the emergence of new pricing(Combination of CPU frequency and Cost) schema and its impact to the User CPU frequency Selection.
- **Objective : Optimize cost under deadline constraint**
  Pietri I., Sakellariou R. (2016) Cost-Efficient CPU Provisioning for Scientific Workflows on Clouds. In : Altmann J., Silaghi G., Rana O. (eds) Economics of Grids, Clouds, Systems, and Services. GECON 2015. Lecture Notes in Computer Science, vol 9512. Springer, Cham

Comprehensive comparison of WS models

- Study the performance of different workflow scheduling models.
- Their studies shows that many WS model have the same performance for the same workload.
- **Objective : Minimise cost under deadline**
  Bugingo E., Qin. Y.,Wang J., Zhang D. and Zheng W. (2018), Cost optimization heuristics for deadline constrained workflow scheduling on clouds and their comparative evaluation, CCPE., 2018 ;00 :1–20 .

Objectives, Relationship and Difference

Multi Objectives

- Minimize cost and Makespan under Deadline and budget.
- Maximize Robustness probability.
- Ensure Security Level Requirement.

Relationship with other works

- Like SCAS and Robustness policies we consider cost and makespan as our main objectives to optimize.
- Like Cost aware models and Comprehensive comparison of WS models we consider CPU frequency as the only one fact to charge VM and VM performance variation.
- Like Robustness policies we consider to use maximum Robustness.
- Like all SWS models we want to ensure Security Level Requirement.

Difference with other Multi objectives

- Optimization of multiple objectives based on the parameters defined by the users.

Cost, Uncertainties,Execution time

Cost

- Cost will depend on the CPU frequency allocated to the VM during the execution of each task.
- In other papers, they had a rounding policy for the cost(5.1 hours:6 hours)

Uncertainties

- **VM performance Variation**: Depends on the CPU frequency allocated.
- **Task failure**: Depends on the performance of the VM. This may cause the failure of the whole workflow.
- Diversity of users' requirements.

Execution time, Makespan

- **Execution time**: will depend on the CPU frequency allocated to the machine.
- **Makespan** : will depend on the **Execution time** and **Encryption time.**

Priorities:Cost, Makespan, Security Level Requirement

Cost

- Ensure Low cost high makespan by running the model using Minimal CPU frequency at each VM.
- Change CPU Frequency to reduce makespan according to the Deadline.
- Check robustness for each security Level Requirement(Choose maximum).

Makespan

- Ensure Low Makespan high cost by running the model using Maximum CPU frequency at each VM.
- Change CPU Frequency to reduce cost according to the Budget.
- Check robustness for each security Level Requirement(Choose maximum).

Priorities:Cost, Makespan, Security Level Requirement

Security Level Requirement:Low and High

- Ensure Low Makespan, Low Security Level Requirement high cost by running the model using Minimum CPU frequency at each VM.
- Change CPU Frequency to reduce cost according to the Deadline.
- Check robustness for each security Level Requirement(Choose maximum).

Security Level Requirement:Low and High

- Ensure Low Makespan, High Security Level Requirement high cost by running the model using Minimum CPU frequency at each VM.
- Change CPU Frequency to reduce cost according to the Deadline.
- Check robustness for each security Level Requirement(Choose maximum).

Conclusion and Future Works

Recall

- we have reviewed WS and SWS model and determine their objective.
- we have identified Relationship and difference of our model and the state of the art.

Future works

- Experiments

Q and A