Preserving Privacy for Workflow Executions in Hybrid Clouds

Shaghayegh Sharif Nabavi (*Professor Albert Y.Zomaya* , Dr Javid Taheri)
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

Introduction

- Distributed applications require the execution of complex ***workflows*** consisting of computing modules with inter-module communications in large network environments.
- These workflows constitute a common pattern for describing a wide range of applications such as ***business processes***, ***scientific applications***, and ***web services***.
- In this work, we study the workflow execution in Hybrid clouds . A ***hybrid cloud*** (**HC**) is a composition of two or more heterogeneous resource infrastructures Figure 1. (private, community, and/or public)

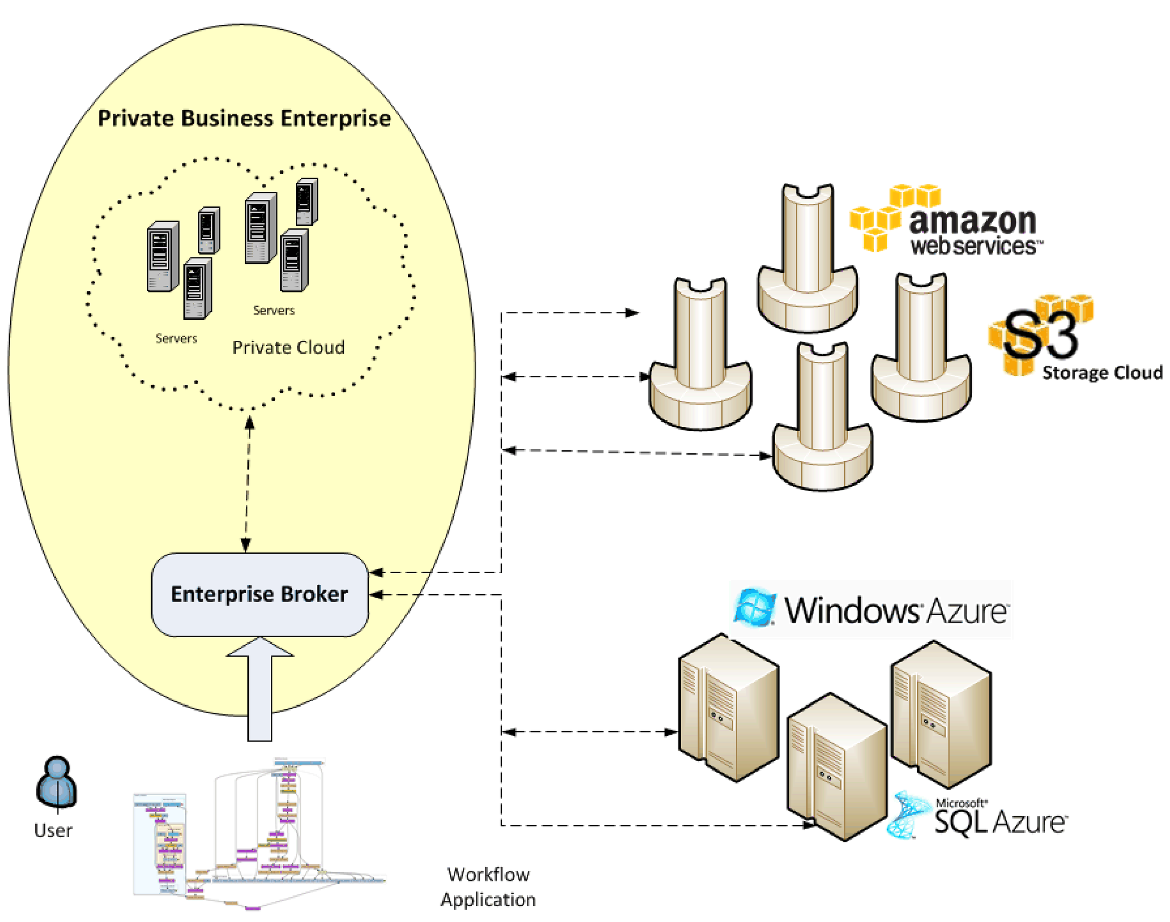

Fig. 1. Hybrid Cloud

Challenges

Hybrid clouds encounter the following two main obstacles in reaching their full potential:

- Customers dissatisfaction due to the conflicting nature of the constraints (budget and deadline
- Exposure of customers' private data in hybrid cloud infrastructures.

Contribution

1. Optimising the cost and improving the efficiency of cloud resources while preserving privacy for workflow scheduling under customers requirements such as deadline and budget.
2. An improved cost model for hybrid cloud environments.
3. Evaluation of a privacy preservation scheduling algorithm using the defined cost model for various de facto standard experimental workflows.

Sample Scenario

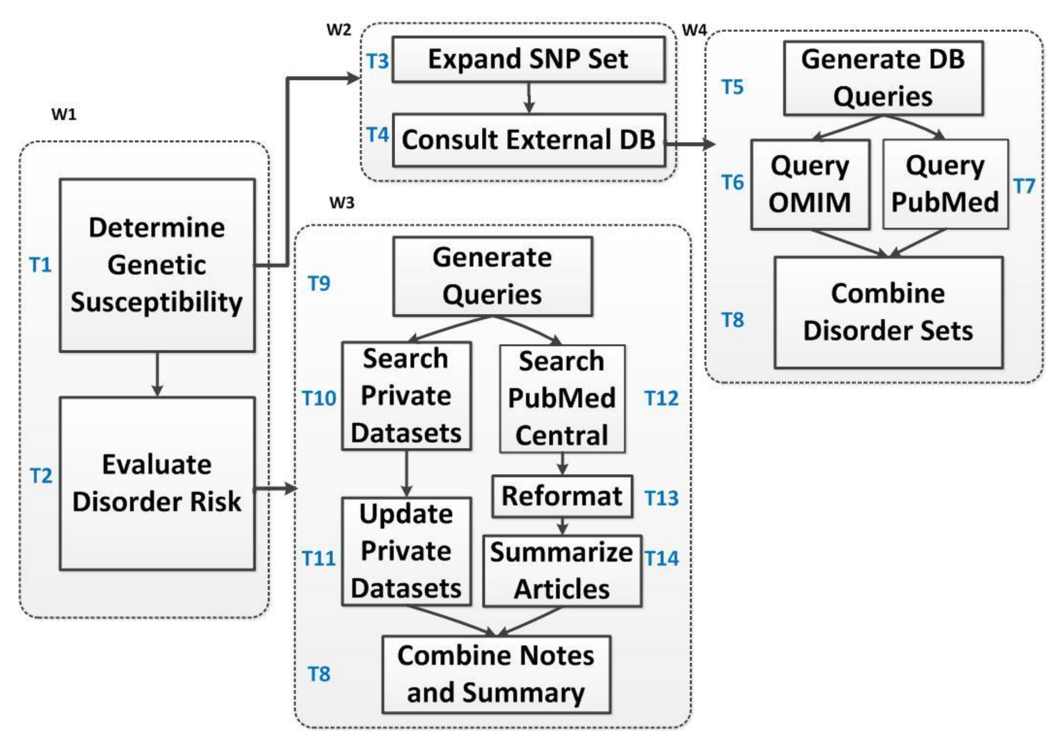

Fig. 2. Disease susceptibility workflow specification

Privacy in workflows includes three types;

- Data Privacy
- Task Privacy
- Structural Privacy

Solution

We designed, implemented and evaluated a multi-criteria workflow scheduler in hybrid cloud environment, ***Multiterminal Cut for Privacy in Hybrid Clouds*** (**MPHC**), to optimise cost under privacy and deadline constraints. We solve the privacy issue by applying multiterminal cut algorithm to partition the workflow based on tasks privacy privilege.

Result

We compared MPHC with two efficient algorithms each considering one constraint only:

- ***IC-PCP[1]*** scheduler considers the workflow's deadline but ignores its privacy requirements.
- ***DAGMan/HTCondor[2]*** scheduler assigned tasks according to their privacy requirements to machines; however, ignored the workflow deadline.

MPHC vs. ICPCP I:

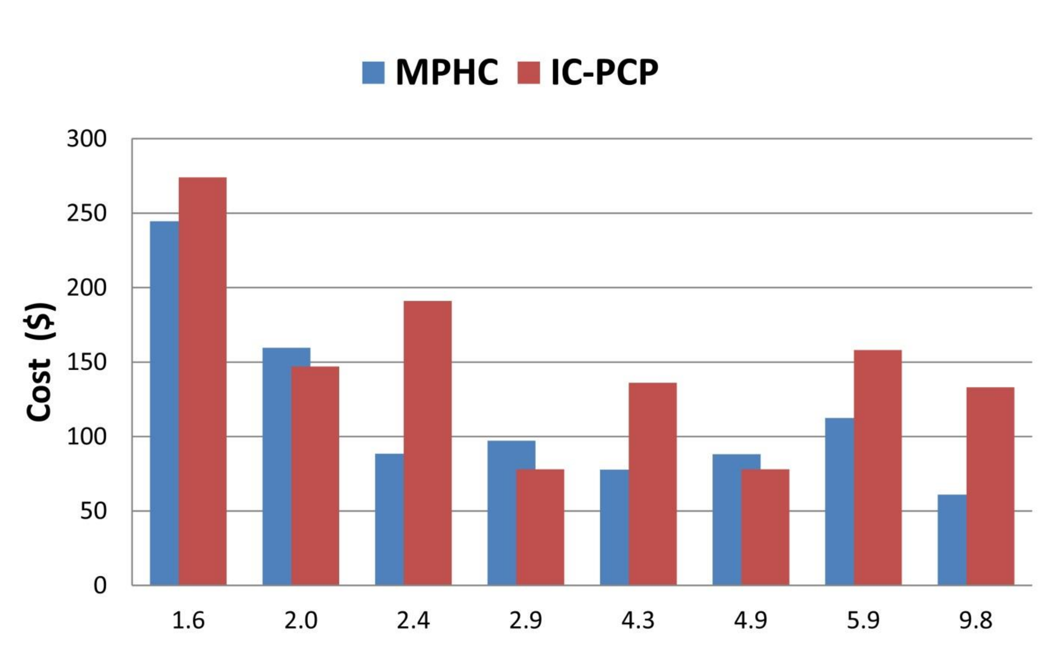

Fig. 3. MPHC vs. IC-PCP behaviour

MPHC vs. ICPCP II:

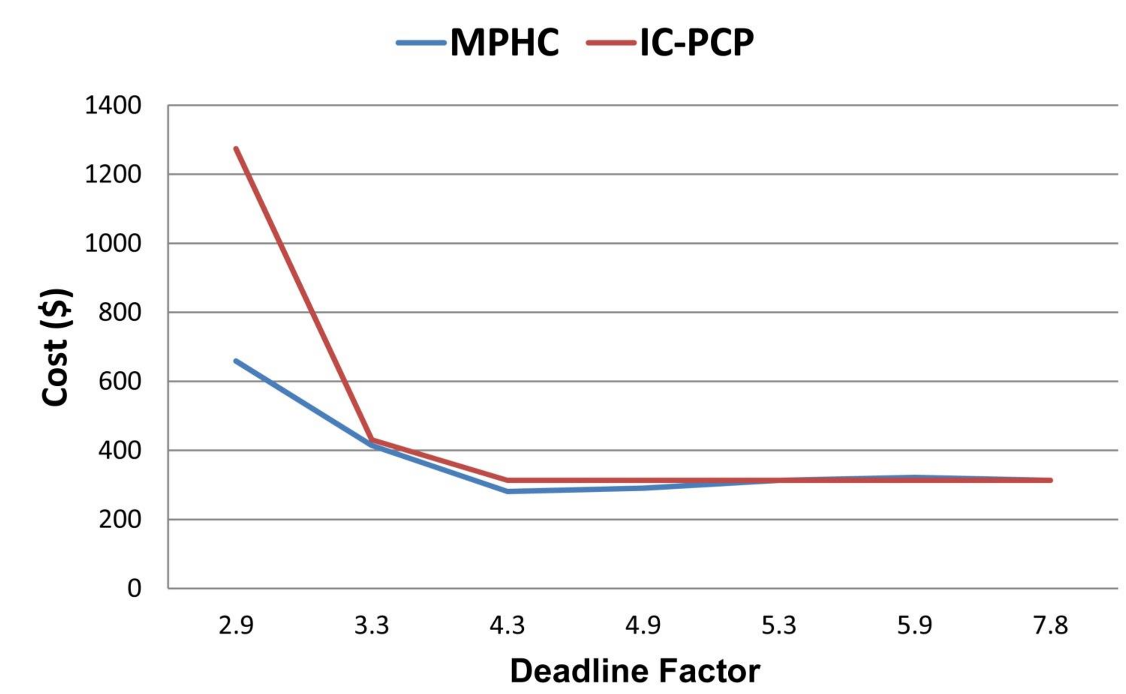

Fig. 4. Effect of extending the deadline

MPHC vs. HTCondor: MPHC algorithm scheduled all the workflows tasks before the deadline in contrast to HTCondor.

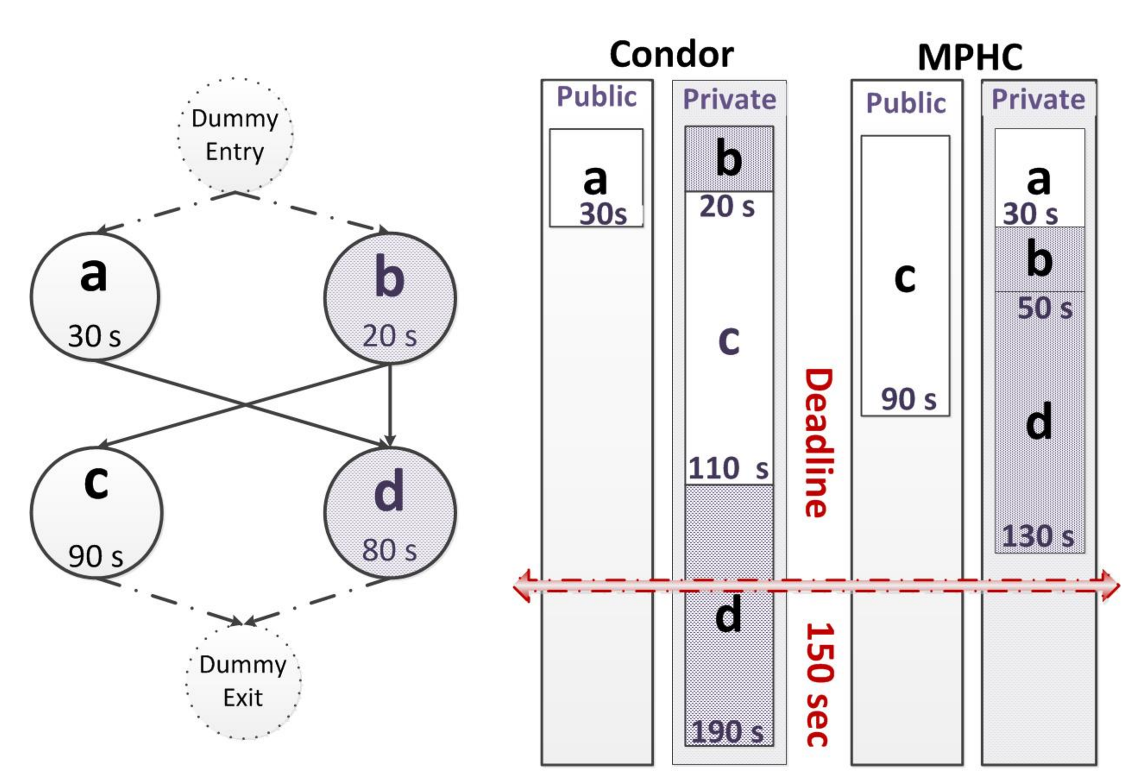

Fig. 5. MPHC vs. HTCondor

Data Size: In MPHC, the minimum possible data were transferred, while DAGMan did not consider the data size.

Data Size	1 G		0.5 G	
WF Size	25 Node	50 Node	25 Node	50 Node
MPHC	1.5	5.1	0.26	0.4
HTCondor	2.5	10.3	0.2	0.5

Table. 1. Performance regarding to data size

Conclusion

We introduce ***MPHC*** algorithm to minimise the cost of executing workflows, while satisfying both task/data privacy and deadline constraints in HCs**.** MPHC outperformed IC-PCP and HTCondor algorithms not only by reducing the cost of running a workflow by 7-31%, but also executing all tasks of a workflow before their deadlines.

References

1. S. Abrishami, M. Naghibzadeh, and D. H. Epema, "Deadline constrained workflow scheduling algorithms for infrastructure as a service clouds," Future Generation Computer Systems, vol. 29, no. 1, pp. 158–169, 2013.
2. M. J. Litzkow, M. Livny, and M. W. Mutka, "Condor-a hunter of idle workstations," in Distributed Computing Systems, 1988., 8th International Conference on. IEEE, 1988, pp. 104–111.

THIS RESEARCH SPONSORED BY CSIRO