A Dynamic Data-Driven Simulation Approach for Preventing Service Level Agreement Violation in Cloud Federation

Funmilade Faniyi[a,*], Rami Bahsoon[a], and Georgios Theodoropoulos[b]

[a]School of Computer Science, University of Birmingham, UK
[b]IBM Research Lab, Dublin, Ireland

*fof861@cs.bham.ac.uk

June, 2012

Motivation

- Cloud computing vision (McCarthy, 1961)
  - appealing economic incentive,
  - reduced administrative burden,
  - provides expansion opportunities for SMEs,
  - reduced under-utilisation or over-utilisation of computing resources, and
  - increase in adoption from education, finance, healthcare, and government sectors.

Cloud Computing: Overview

- Essential Characteristics: (NIST, 2011)
  - self-service, network access, resource pooling (multi-tenancy), rapid elasticity
- Types:
  - SaaS (e.g. Salesforce, Google Docs)
  - PaaS (e.g. Google AppEngine, Microsoft Azure)
  - IaaS (e.g. Amazon EC2, S3; Rackspace)
- Billing Model:
  - pay-per-use (on-demand, reserved, spot market)
  - subscription-based (similar to conventional hosting service)

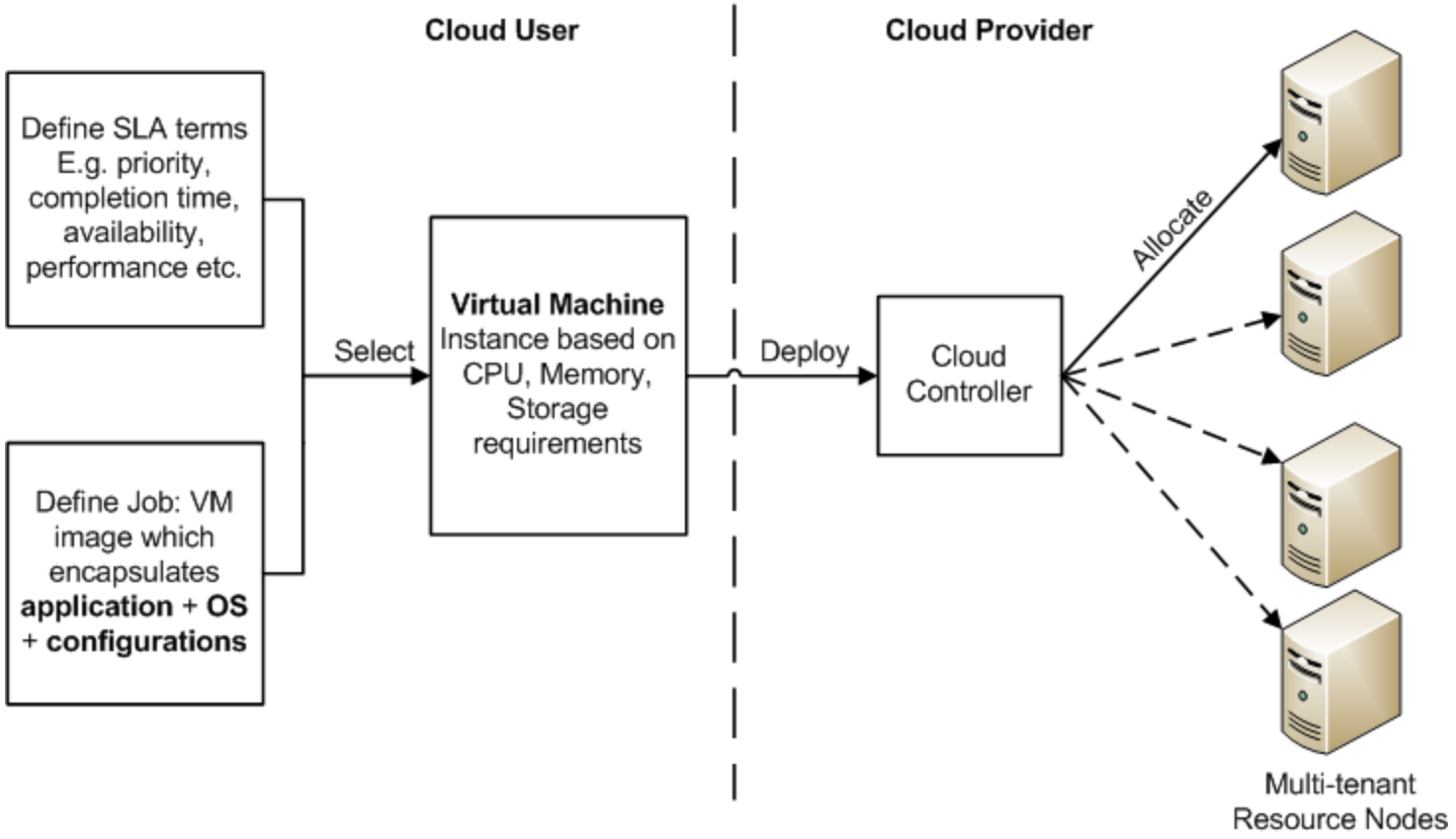
Cloud Computing: Overview

- Deployment Model: (NIST, 2011)
  - private (one organisation),
  - community (shared among several organisations),
  - public (one organisation, publicly accessible),
  - hybrid, and
  - federation model (utilise multiple clouds)

- Focus of this work:
  - public clouds that are capable of on-demand provisioning of computational resources

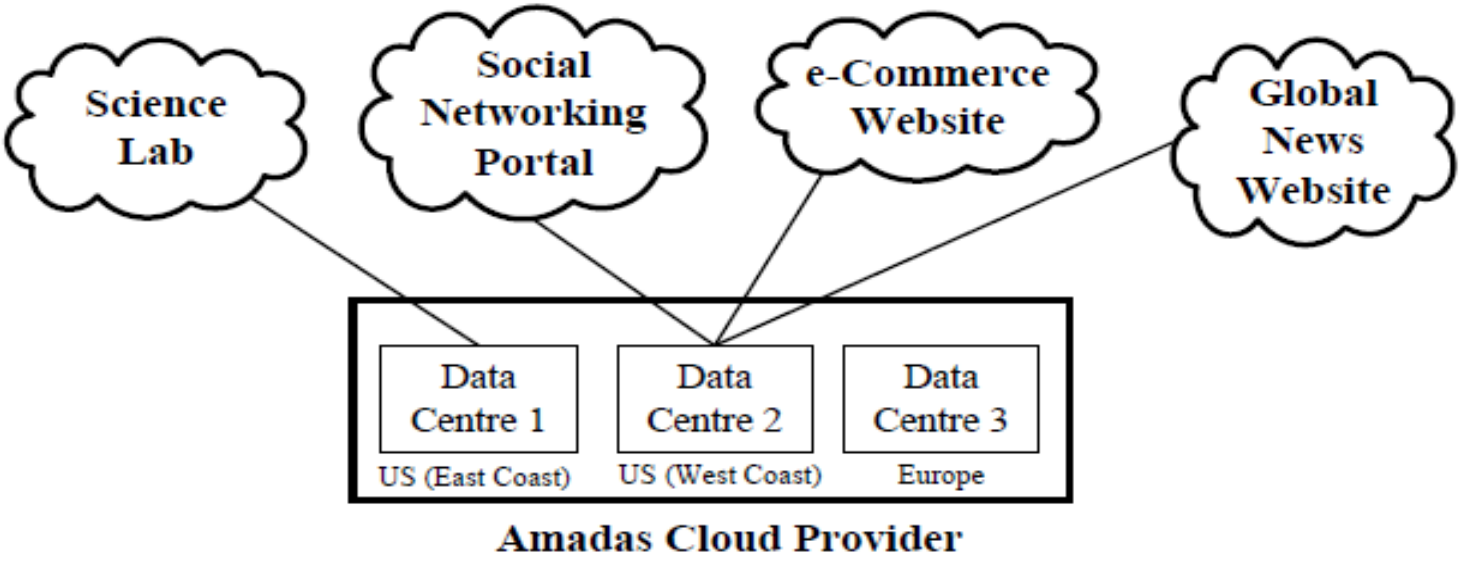
Service Level Agreement

- A typical use case of Single Provider Model



Single Cloud Provider Model

- Unpredictable Spike in Workload Scenario



- Key Properties
  - cloud users have different applications and QoS needs differ
  - cloud providers are not interconnected (no resources sharing)
  - the burden of harnessing multiple cloud providers' resources is on the user

SLA Violation Incidents

- Recent outages of popular cloud service, Gunawi et al. (2011)
  - network misconfiguration, upgrade events,
  - maintenance, power failure,
  - network failure, system overload, and
  - overheated domain controller
- CloudHarmony* surveyed 38 cloud providers for a period of one year (2011)

  Objective: Compare actual availability to those in SLAs
  - vendors that met or exceeded SLA terms: 39%
  - vendors with 100% SLA achieving their SLA: 17%

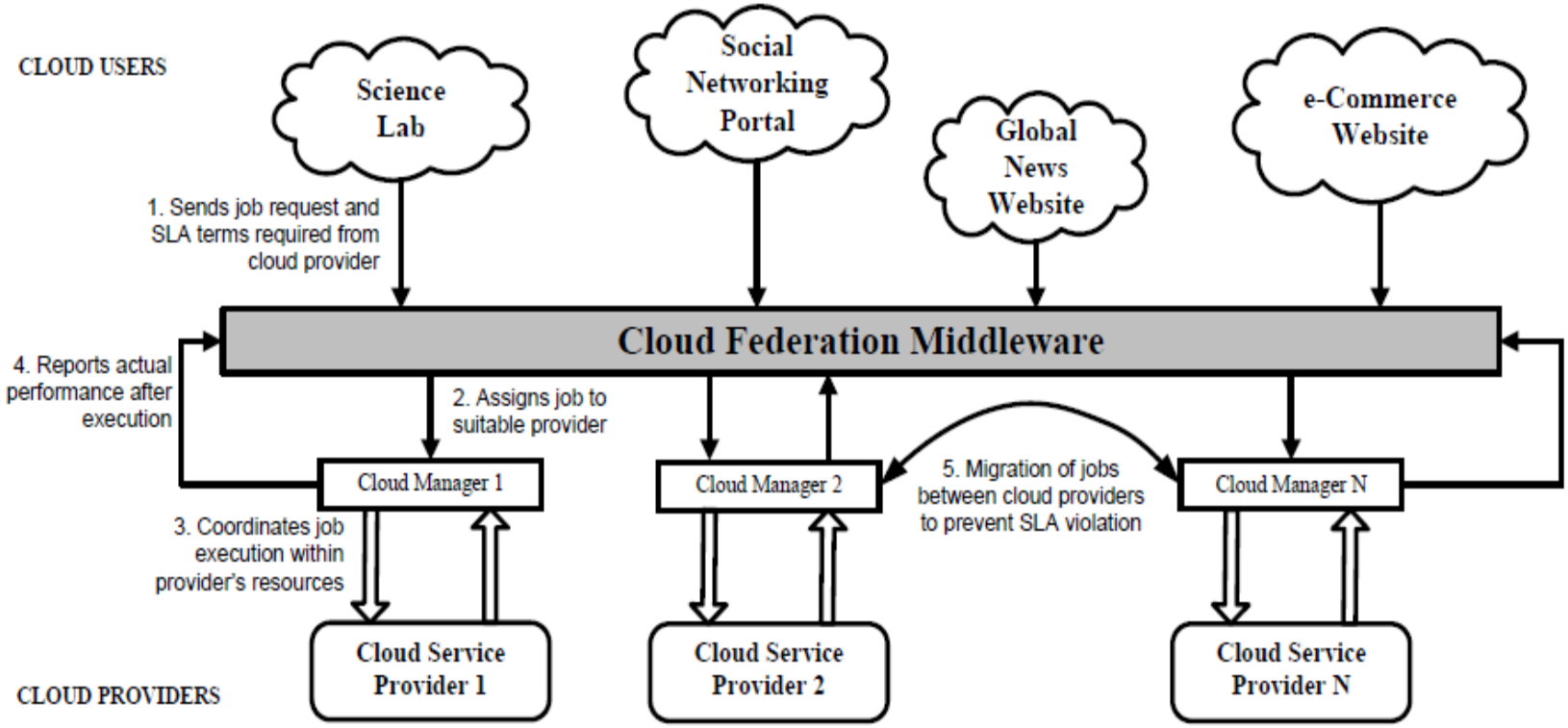
  * http://www.cloudharmony.com

Causes of SLA Violations

- Post-deployment scenarios:
  - changes in application e.g. workload, configuration
  - infrastructural failure e.g., network, power
  - resource node misbehaviour such as:
    - resource contention among multiple VMs,
    - failure of underlying hypervisor or hardware node

- Research Question:
  - Can a Cloud Federation model improve SLA compliance beyond those achievable using single provider model?

Cloud Federation Model

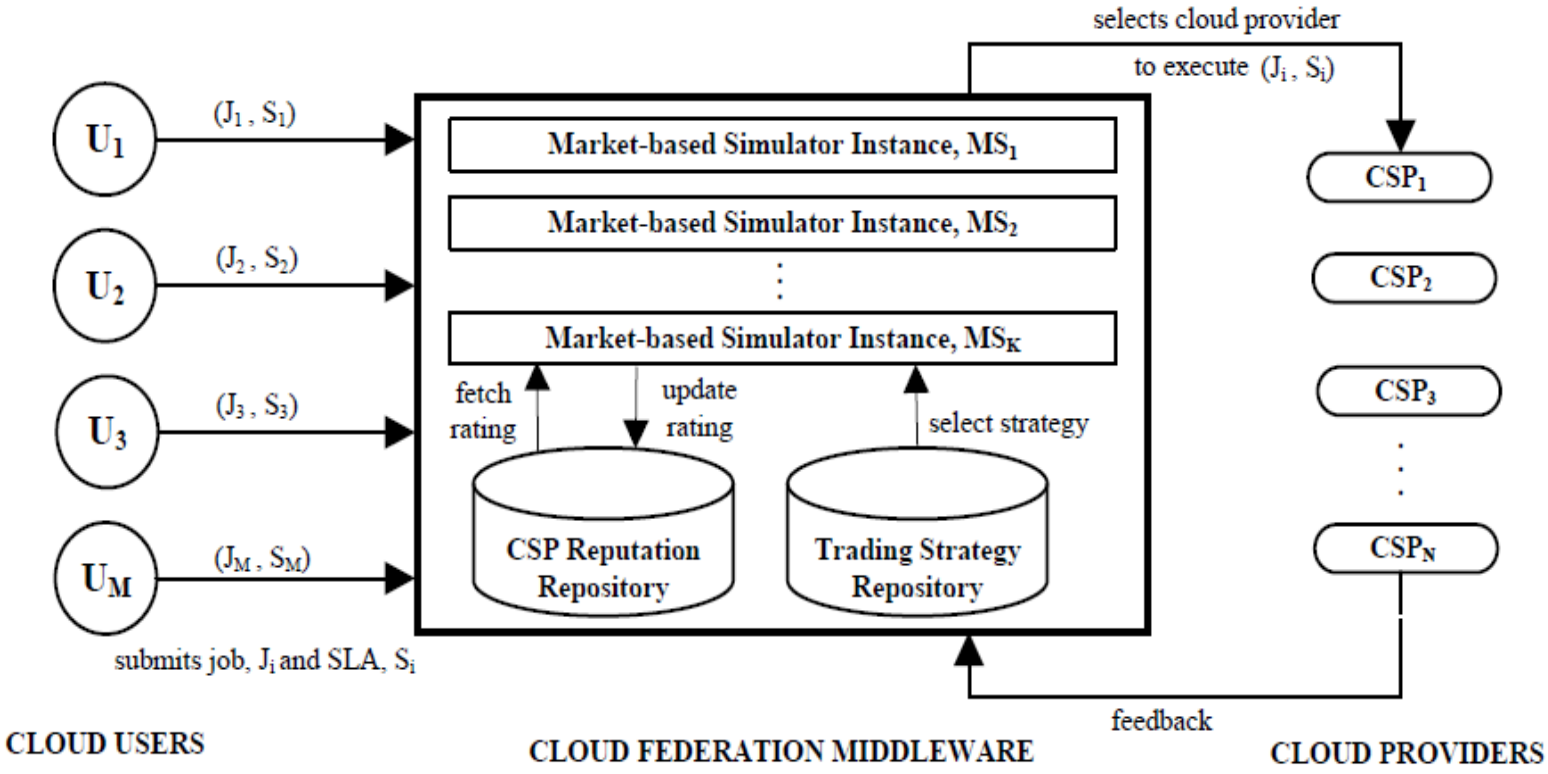


- Key Properties:
  - cloud providers are interconnected via middleware
  - has capability for run-time adaptation and reconfiguration
  - job management is transparent to cloud users

Proposed Solution (DDDAS)

- Strategy:
  - collect data about performance of cloud providers (past behaviour)
  - use these data to make future decision about job allocation to cloud providers (prediction)
  - decision could also involve real-time migration of jobs across cloud providers
  - continuously adapt to changes in cloud providers' behaviour and incoming jobs' SLA terms
- Technique:
  - model the cloud federation as a distributed computational resource market

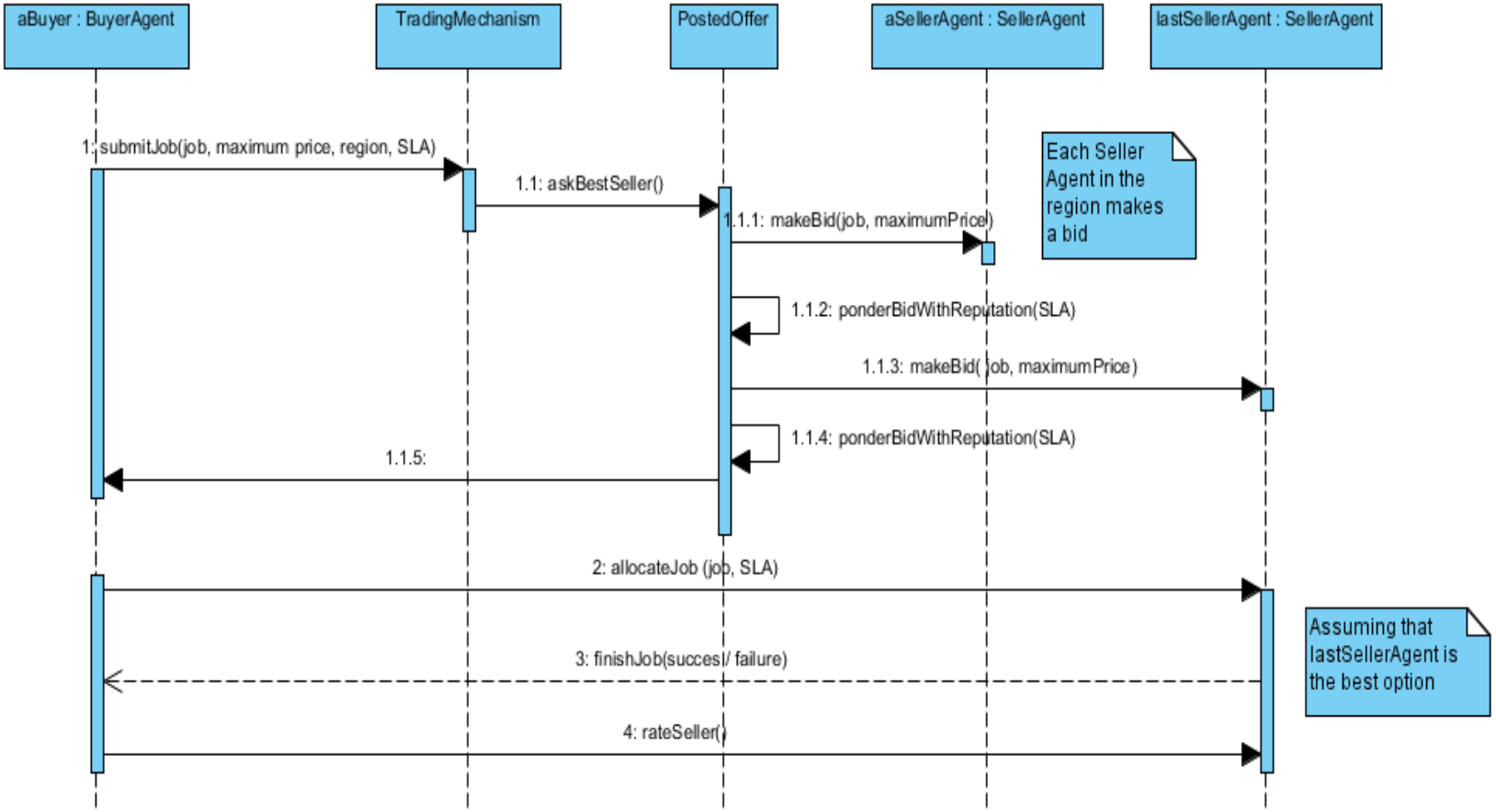
Data-Driven Cloud Federation Middleware



- Distributed simulator instances (market based control techniques)
- Adaptable trading strategy repository
- Assessing cloud providers' via reputation measurement
- Monitoring and selection of cloud providers

Market-based Simulator

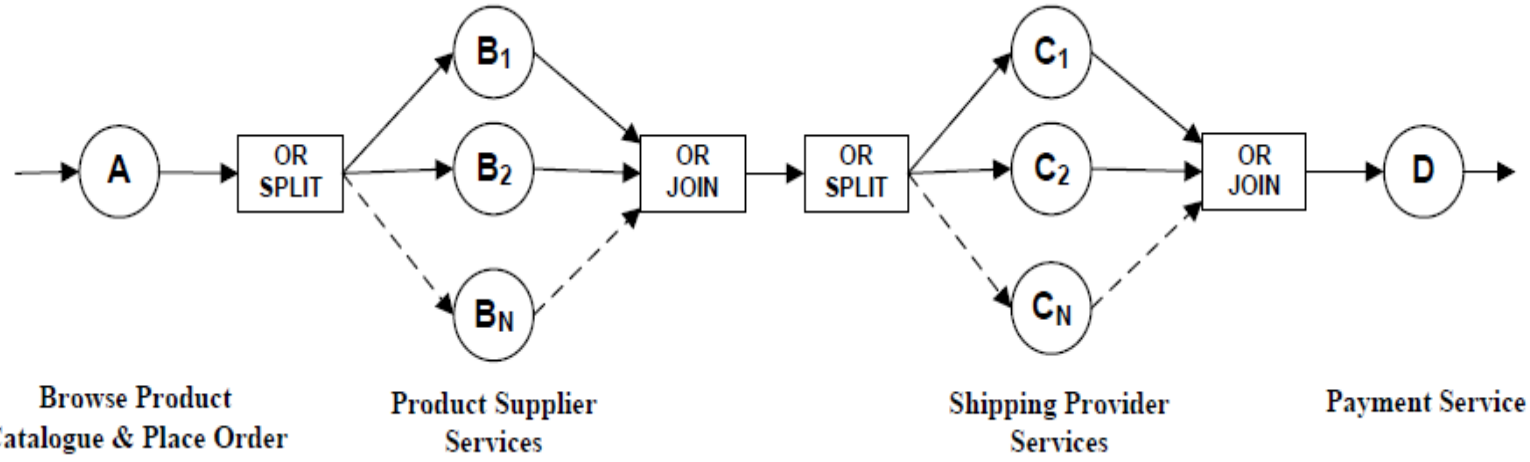
- Mechanism: Posted-Offer model (Lewis et al., 2010)



Sequence diagram of Posted-Offer Model

Sample Application: An Online Shopping Cart

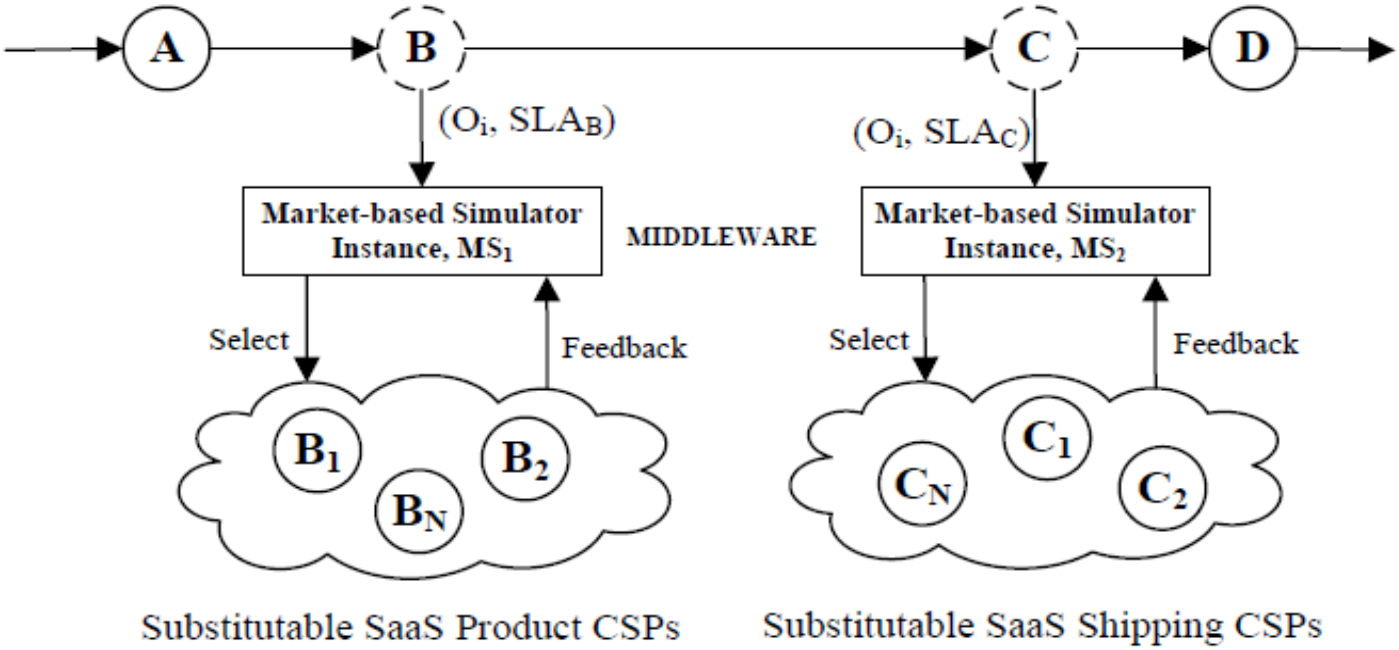
- Workflow Compositional Model



- Objective of Application:
  - *To minimise the cost of meeting customer orders without exceeding the promised delivery time*
- Application Features:
  - Multiple product suppliers with varying quality of service can meet customer order.
  - Multiple shipping providers can also deliver ordered product

Sample Application: An Online Shopping Cart

- Cloud Federation Instantiation



- Simulator, MS_1, monitors and selects from pool of product suppliers
- Simulator, MS_1, monitors and selects from pool of shipper services
- Adaptation actions are effected based on observed states

Reflection and Summary

- The success of cloud computing depends on the ability of cloud providers to improve SLA compliance
- Unlike the single provider model, the cloud federation model harnesses multiple providers' resources
- Using DDDAS for coordinating providers in a cloud federation is a novel exploitation of the approach with promising prospects
- Future Work:
  - Experimental evaluation in real cloud setting
  - Incorporate on-line learning capabilities
  - Reasoning about the tradeoffs between the overhead of the approach and achievable SLA compliance

Questions & Comments

References

- Lewis et al. (2010). Resource Allocation in Decentralised Computational Systems: An Evolutionary Market-Based Approach.
- H. S. Gunawi et al. (2011), Failure as a service (faas): A cloud service for large-scale, online failure drills
- P. Mell and T. Grance (2009), The NIST Definition of Cloud Computing, Tech. rep., NIST, Information Technology Laboratory.
- Wikipedia (McCarthy, 1961)

  http://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist)