Cost Reduction in Hybrid Clouds for Enterprise Computing

Biyu Zhou, Fa Zhang, Zhiyong Liu (ICT, CAS)
Jie Wu (Temple University)
Outline

• Background and motivation
• Problem
• Framework
• Evaluation
• Discussion
# The Rapid Development Of Cloud Computing

## The Rapid Growth of Cloud Computing, 2015-2020

<table>
<thead>
<tr>
<th>Year</th>
<th>Public Cloud Spending ($B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>$67</td>
</tr>
<tr>
<td>2016</td>
<td>$82</td>
</tr>
<tr>
<td>2017</td>
<td>$99</td>
</tr>
<tr>
<td>2018</td>
<td>$117</td>
</tr>
<tr>
<td>2019</td>
<td>$138</td>
</tr>
<tr>
<td>2020</td>
<td>$162</td>
</tr>
</tbody>
</table>

Worldwide Spending on Public Cloud Computing, 2015-2020 ($B)

- Average Compound Growth Rate, YE2015-YE2020, 19%
- IT Spending Average Compound Growth Rate, YE2015-YE2020, 3%

*Source: IDC, 2016*
Concerns with Public Cloud

- Cheap cost
- Scalability
- SLA
- Privacy
Call For Hybrid Cloud!

Enterprise Cloud Strategy
1000+ employees

- Multi-Cloud: 85%
- Single private: 5%
- Single public: 9%
- No plans: 1%
- Multiple private: 7%
- Multiple public: 20%
- Hybrid cloud: 58%

Source: RightScale 2017 State of the Cloud Report
Our Focus: Planning Hybrid Cloud Layouts

- Cost savings, application response times, communication costs
Model of Enterprise application
Abstracting the planning problem

- **Objective:** Maximize cost savings on migration
  - Benefits due to hosting servers in the cloud
  - Costs change related to wide area Internet communication (simple but practical linear model)

- **Time Constraints:**
  - The completion time of the application is defined as the maximum completion time of all workflows
Motivation Example

Fig. 1. The application and four migration plans with the corresponding cost reduction and completion time.
Analysis

**Complexity:**
① Solving the general application deploy problem is NP-Complete.

**Key observations:**
① Most of the multi-tier enterprise application can be easily divided into multiple DAGs.
② Solving the DAG deploy problem is much easier!
An overview of the Framework

- Partition
  - The characteristics of typical multi-tier applications, e.g., traffic

- Solving sub-problems in a parallel manner
  - A two-stage algorithm based on dynamic programming

- Constructing final solution
  - Feasible and efficient
A two-stage algorithm

- Step 1: transforming a DAG into a sequence

- Step 2: dynamic programming-based algorithm

\[ c[i][j][k]: \text{the max cost reduction of the subgraph rooted at node } v_i \text{ when node } v_i \text{ is assigned to cloud } h_k \text{ and when the total delay is no larger than } j. \]

- Case 1: all the children of node \( v_i \) has only one parent node
- Case 2: at least one of the children of node \( v_i \) has multiple parent nodes
The evaluation setting

Applications:
① Six randomly generated DAGs as application architectures.
   ① Each DAG involves a number of nodes between 500 to 2K.
   ② Each communication pair is associated with unit traffic.
   ③ Simulation results are

Hybrid cloud:
① A local cloud.
② Two public cloud from amazon:
   ① one in Northern Virginia (NOVA).
   ② One in Tokyo.
The performance-cost reduction

- (a) The average cost reduction of different strategies under different time constraints.
- (b) our framework can bring up to 79.15% cost reduction to enterprises. Besides, the cost reduction obtained by our algorithm is close to that of the optimal solution solved by COMBSPO.
- (c) our framework performs better in reducing enterprise costs leveraging the hybrid cloud architecture under controllable time overhead than the other two strategies.
• (a) by varying the value of time constraint, one can obtain a large cost reduction with large time overheads.
• (b) Choosing a proper value for an application depends on the performance requirement of the application manager.
The performance—the effect of user location.

- (a) migrating applications that have larger percentage of external user to cloud will bring more cost reduction than migrating the ones that have smaller percentage of external user.
- (b) The cost reduction of migrating the applications with users evenly distributed in three regions is the least.
Thanks for your attention!