TF4: Fair demands distribution for Cloud Federation

Wojciech Burakowski and Maciek Sosnowski
Warsaw University of Technology, Poland
Agenda:

• Traffic management scheme for Cloud Federation

• Strategy for handling request
  – Some numerical examples
  – Conclusions

• Further work

• Summary
Cloud Federation (2)

New elements comparing to the separated Cloud:
- CF network
- Resources in many Clouds

Potentiality of Cloud Federations vs. Separated Clouds:
- Better resource utilization
- More offered services

Problem:
- More complicated traffic management

Exemplary Cloud Federation with 5 Clouds
Multi-level traffic model for Cloud Federation

Level 5
- Strategies for handing demands
  - Distribution of demands between Clouds

Level 4
- Cloud Federation Setting
  - Resources in Cloud Federation
    - types of resources
    - number of resources
  - Network connecting Clouds
    - network topology
    - network provision

Level 3
- Service of IS requests
  - Service specification
    - specification of task sequence
  - Service provision
    - allocating resources to execute particular task

Level 2
- Service of ES requests
  - Service of appropriate task sequence
  - Allocating tasks to the resources (if more than one resource can serve the task)

Level 1
- Task service in cloud resources
  - Task scheduling
  - Fairness

Installing Service (IS) Requests

Execution Service (ES) Requests
Level 5: Strategies for handling requests

Considered systems:
- Separated cloud: SC
- Cloud Federation with centralized management: CF-C
- Cloud Federation with Home Clouds: CF-HC
- Advanced Cloud Federation with Home Clouds: CF-A-HC

System description (simple assumptions just for comparison):
- N – number of Clouds in CF
- Ci – number of resources in the i-th Cloud (ass. – all resources are the same)
- αi - arrival rate of requests for each cloud acc. to Poisson distribution (ass. each request demands one resource, resource is not shared)
- hi – service rate of requests in each resource (ass. negative exponential distribution)
- Pure loss system

Cloud performances:
- Ai : carried load by i-th cloud
- Ploss(i) – request lost rate for i-th cloud
Level 5: Strategies for handling requests

Scenario 1: separated clouds

Request rate $\lambda_1$

Cloud#1: number of resources = $C_1$

Request rate $\lambda_2$

Cloud#1: number of resources = $C_2$
Level 5: Strategies for handling requests

Scenario 2: full federation

Request rate $\lambda_1 + \lambda_2$

Cloud#1: number of resources = $C_1$

Cloud Federation: number of resources = $C_1 + C_2$

Cloud#1: number of resources = $C_2$
Level 5: Strategies for handling requests

Scenario 3: Home Clouds + Cloud Federation

Request rate $\lambda_1$
- If no available resources
  - Cloud#1: number of resources = $C_1$
  - Number of resources available for Cloud1 users = $C_1 + C_2$

Request rate $\lambda_2$
- If no available resources
  - Cloud#1: number of resources = $C_2$
  - Number of resources available for Cloud2 users = $C_1 + C_2$
### Level 5: Conclusions about strategies 1, 2, and 3

#### Exemplary results

<table>
<thead>
<tr>
<th>Single cloud</th>
<th>Lambda</th>
<th>Load</th>
<th>Loss ratio</th>
<th>WORK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.398</td>
<td>5.30E-03</td>
<td>3.98</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.575</td>
<td>4.30E-02</td>
<td>5.75</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.703</td>
<td>1.22E-01</td>
<td>7.03</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.749</td>
<td>1.70E-01</td>
<td>7.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Full FED</th>
<th>Lambda</th>
<th>Load</th>
<th>Loss ratio</th>
<th>WORK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18</td>
<td>0.599</td>
<td>2.70E-03</td>
<td>17.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Home Cloud</th>
<th>Lambda</th>
<th>Load</th>
<th>Loss ratio</th>
<th>WORK</th>
<th>SC GAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8</td>
<td>0.709</td>
<td>2.70E-03</td>
<td>7.09</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.48</td>
<td>2.70E-03</td>
<td>4.8</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.605</td>
<td>2.70E-03</td>
<td>6.05</td>
<td>0.3</td>
</tr>
</tbody>
</table>

#### Conclusions

- Single Cloud is not preferable solution for small clouds.

- Full Cloud is not preferable solutions for clouds with different volume of resources as well as number of own users.

- Home Cloud have better performances comparing to Single Cloud but such solution does not guarantee fair sharing of new profit.
The proposed solution: to split the resources belonging to a given cloud into two groups:
* private resources (only used by Home Cloud users)
* common resources (used by Cloud Federation)
* the profit from the carried load by common resources is equally distributed between two Home Clouds

Load for i-th Home Cloud =
load carried by private resources +
+ ½ of the load carried by common resources
Level 5: Strategies for handling requests in Cloud Federation

Profit - volume of carried traffic

Calculation of the profit for the case of 2 clouds

Load for i-th Home Cloud =
load carried by private resources +
+ ½ of the load carried by common resources

By choosing appropriate parameters for c12 and c22 values we receive almost equal profit (0.340 and 0.474)
C1=C2=10 -

<table>
<thead>
<tr>
<th></th>
<th>C12</th>
<th>C22</th>
<th>C11+C21</th>
<th>load C1</th>
<th>load C2</th>
<th>load Fed</th>
<th>Work C1</th>
<th>Work C2</th>
<th>Work total</th>
<th>Loss C1</th>
<th>Loss C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>0.765</td>
<td>0.641</td>
<td>0.357</td>
<td>6.120</td>
<td>3.205</td>
<td>11,824</td>
<td>1,80E-02</td>
<td>1,20E-02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Work Fed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7,370</td>
<td></td>
<td></td>
<td>11,824</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4,455</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC GAIN</td>
<td>0.340</td>
<td>0.474</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OFFER=12
Case with 3 clouds

Assumptions:
- if we add next cloud we should guarantee additional profit for existing consortium and profit for new consortiant

Calculation of the profit for the case of 3 clouds

Load for i-th (i=1,2) Home Cloud =
- load carried by private resources +
- \( \frac{1}{2} \) of the load carried by common resources (in the case of 2 consortiants)
- \( \frac{1}{2} \) (1/2 of the additional load caried by the common pool of resources)

Load for i-th (i=3) Home Cloud =
- load carried by private resources
- \( \frac{1}{2} \) of the additional load caried by the common pool of resources
### Level 5: Strategies for handling requests in Cloud Federation

#### Some numerical results

<table>
<thead>
<tr>
<th></th>
<th>lambda1</th>
<th>Work Fed</th>
<th>OFFER=16</th>
<th></th>
<th></th>
<th>Loss C1</th>
<th>Loss C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda2</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C22</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C11+C21</td>
<td>4</td>
<td>0,765</td>
<td>0,765</td>
<td>0,58</td>
<td>6,120</td>
<td>6,120</td>
<td>14,560</td>
</tr>
<tr>
<td>Fed1 GAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,160</td>
<td>1,160</td>
<td></td>
</tr>
<tr>
<td>Work+Fed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7,280</td>
<td>7,280</td>
<td></td>
</tr>
<tr>
<td>SC GAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,250</td>
<td>0,250</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>lambda3</th>
<th>Work Fed</th>
<th>OFFER=20</th>
<th></th>
<th></th>
<th>Loss C1</th>
<th>Loss C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C22</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C32</td>
<td>7</td>
<td>0,765</td>
<td>0,765</td>
<td>0,536</td>
<td>0,4864</td>
<td>6,120</td>
<td>3,752</td>
</tr>
<tr>
<td>Fed1 GAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,160</td>
<td>1,160</td>
<td></td>
</tr>
<tr>
<td>Fed2 GAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,275</td>
<td>0,275</td>
<td>0,549</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7,555</td>
<td>4,301</td>
<td></td>
</tr>
<tr>
<td>SC GAIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,525</td>
<td>0,525</td>
<td>0,321</td>
</tr>
</tbody>
</table>
- We have presented the main idea for new model to share resources in Cloud Federation
- We need more time to specify some details and to derive supporting formulas
Level 5: Strategies for handling requests in Cloud Federation

Centralized approach

- All demands are submitted to Cloud Federation
- Cloud Federation is managed as one Cloud
- All resources belong to Cloud Federation

With Home Clouds

- Home Clouds have its own users
- The requests are submitted to Home Clouds
- If no resources in Home Cloud then they are submitted to Cloud Federation
Considered system (cloud sizes and request arrival rate to each Cloud are the same):
$N=5,10; \; Ci=10,20; \; \alpha_i=5(10)-9.5(19); \; hi=1$

Conclusions:
- If capacity of the CF is growing (greater number of resources) then loss rate is going less
- CF has more resources than SC so lost rate
- From the point of loss rate the FC-HC has the same characteristics as CF-C
- Final conclusion – CF has better performances than SC
Level 5: Strategies for handling Demands

Considered system (Cloud sizes are identical but arrival rates to each Cloud are different): \( N=5 \); \( C_i=10; h_i=1; \alpha_1=2, \alpha_2=4, \alpha_3=6, \alpha_4=8, \alpha_5=10 \)

<table>
<thead>
<tr>
<th>SC</th>
<th>lambda</th>
<th>load</th>
<th>loss rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0,2</td>
<td>4,00E-05</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0,398</td>
<td>5,40E-03</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0,575</td>
<td>1,22E-02</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0,703</td>
<td>2,15E-01</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0,786</td>
<td>2,15E-01</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions:

• All Clouds in FC-HC have better resource utilization comparing to SC
• FC-C vs. FC-HC:
  • Loss rate for each Home Cloud is the same
  • Home Cloud with greater request rate in the FC-HC scheme utilizes more resources comparing to FC-C scheme
  • Home Cloud with less request rate utilizes less resources comparing to FC-C scheme

Final conclusion: choice between FC-HC and FC-C depends on business model for FC
Network level: new element in CF comparing to SC

Exemplary two approaches

CF using independent SLAs

CF using Virtual Network Infrastructure

Network design aspects:

- Profile and volume of traffic submitted to the network
- Requirements for QoS guarantees, Which Classes of Services are needed, if needed
- Methods for network dimension and adaptation to current needs (when traffic is changed in time)
- Which technology: SDN?

Remark: if some clouds will have worst connections when their resources can be not choosen
Requests for installing and provision of a new service, named IS requests that are send by service providers to Cloud Federation owners

- Service specification
- Service provision
Level 3: Handling Service Installation requests

Service specification (acc. to SoA- Software Oriented Architecture):
• The specification of a service has a form of specification (in the simplest case) of task sequence, that should be executed in *a priori* defined order
• More complicated task sequence – with tasks executed in parallel
• Task description: service number, sequence number, ...

Different task sequences corresponding to different services

Example of task sequence with tasks executed in parallel
Service provision

- responsible for choosing a number of resources to execute particular tasks
- Execution of specific task could be possible done by more than one resource (not necessarily belonging to the same cloud)
- Service provision process should be updated depending on service requests rate of given service

Remark: For service provision we dedicate a subset of all resources
Requests for service execution are send by users to Cloud Federation (CF-C) or Home Clouds (CF-HC)

- Service execution demands execution of task sequence
- The issue is (possible dynamic) task allocation taking into account load conditions in resources
## Comparing requests for service setting and service execution

<table>
<thead>
<tr>
<th></th>
<th>IS requests</th>
<th>ES requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements for time execution</td>
<td>off line</td>
<td>on line</td>
</tr>
<tr>
<td>Arrival rate</td>
<td>hours, days</td>
<td>seconds, minutes</td>
</tr>
<tr>
<td>Issues to be solved</td>
<td>How to allocate resources, how to provision how to do re-provision</td>
<td>How to allocate execution of tasks?</td>
</tr>
</tbody>
</table>

**Remark:** Provision and execution processes are dependent: the quality of service execution depends on service provision
Level 4: Cloud Federation setting

Cloud Federation resources (compte resources)
- Detail knowledge about resources in all Clouds
- We should take into account that compute resources
  - are packed in racks and organized in clusters (high bandwidth access)
  - are often connected to shared storage infrastructure (databases)
- Data centers consume energy
Planed evaluation strategy

• To define use cases
  • The best solution - to take use cases currently studied by European projects
• To use CloudSim or to develop simulation tool
  • It looks that CloudSim is not suited
• Try to do some experiments
  • in one of European testbeds
    • Under study
  • in PL-LAB
    • that will be connected to European testbed infrastructure by Fed4Fire at the end of this year
PL-LAB

- finished at the end of this year
- we are thinking about designing a testbed for CF
  - SDN net (about 30 switches)
  - Many servers
  - generator/analyser
The presented multi-level approach for traffic management in Cloud Federation environment looks to be completed

- We distinguish between 5 levels that are essentially different
- For level 5, 3, 2 and 1 we have some approaches
- Only few for level 4 (CF network)

So, the problem is how to merge all levels (to take into account „all pieces together“)

- Close-levels traffic management?

The evaluation approaches

- by simulation
- In testbed environment