A novel time-aware rapid data evacuation heuristic algorithm for large-scale disasters in optical cloud networks

Presenter: Yongcheng(jeremy) Li
PhD student, School of Electronic and Information Engineering, Soochow University, China
Email: liyongcheng621@163.com

Group Meeting, Apr. 08 2016
Outline

1. Background
2. Rapid data evacuation for large-scale disasters
3. Problem statement
4. Rapid data evacuation schemes
5. Conclusions
Background

- Many famous companies have given their cloud services for customers worldwide.
- Cloud services are delivered by datacenter (DC) networks, which provide storage for many important contents.
- Datacenter networks may be damaged by natural (e.g., hurricanes and earthquakes) or man-made (e.g., weapons of mass destruction attacks) disasters, leading to service disruptions and permanent data loss.
- In case of large-scale disasters, there is a risk that multiple DCs may be damaged, and replicas of a content within the disaster region may be vulnerable to loss.
Rapid Data evacuation (RDE) for large-scale disaster

- Many techniques can be used to forecast an upcoming disasters like the example of an earthquake
- According to a received warning and the predicted time, we can transfer the vital data to the other safe datacenter
- The key problems are *where* and *how* to evacuate vulnerable contents before the disaster occurs
Problem statement

- **Objective:** Minimize the total data transfer delay of the data evacuation

- **Inputs:**
  - Topology
  - Disaster region
  - Damaged and safe datacenters
  - Storage capacity of each datacenter
  - Contents in different datacenters
  - Link capacity

- **Constraint**
  - Evacuation deadline
Problem statement

- Disaster Mapping phase
  - The predicted span of the disaster is mapped over the physical network topology
  - Determine the disaster zone and identify the set of damaged and safe DCs

- Content Selection phase
  - Obtain the set of vulnerable contents, as candidate for evacuation

- Source/Destination DC Selection, Path Selection, and Scheduling phase

- Delay Computation phase
  - path-computation delay + connection-setup delay + data-transmission delay + data-propagation delay

- Scheme of Path Selection and Scheduling phase
Delay Computation

- **path-computation delay**
  - Time consumption of the path computation

- **Connection-setup delay**
  - control-message processing delay: \((n + 1) \times \eta\)
  - control-message propagation delay: \(l \times \mu\)
  - switch-configuration delay: \((n + 1) \times \beta\)

- **Transmission delay**: \(F_C / B_P\)

- **Propagation delay**: \(l \times \mu\)

\(n\) is the number of hops on path. \(l\) is the distance of path. \(\mu\) is the propagation delay per unit distance. \(\eta\) is processing delay. \(\beta\) is the switch configuration delay. \(F_C\) is the size of content C. \(B_P\) is the available bandwidth of the path P
Rapid data evacuation schemes

- Nearest Rapid-Data-Evacuation Algorithm (NRDE)
  - Evacuate data only to nearest DC with the shortest path
  - Advantage: simplicity
  - Disadvantage: contents in the same DC will be delivered on the same shortest path to the same DC, which is high time consumption
Rapid data evacuation schemes

- Least Delay Rapid-Data-Evacuation Algorithm (HRDE)[1]
  - Multi-path vs Single-destination node
  - Calculate the data transfer delay of each path
  - Select the path with minimum data transfer delay

Sub-content

- Example
  - Divide the entire content into several pieces

- Advantage
  - Delivery the smaller size sub content by different path to reduce the data transfer delay

- Problem
  - How to divide the content?

- Solution
  - Bisection algorithm
  - Time-aware algorithm
Sub-content

- Bisection Sub-Content Rapid-Data-Evacuation Algorithm (B-SCRDE)(2-disjoint shortest path)
  - Divide one content into two sub-content with equivalent size

<table>
<thead>
<tr>
<th>Path</th>
<th>Delay time</th>
<th>Assigned size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>0</td>
<td>(\frac{1}{2}) size</td>
</tr>
<tr>
<td>A-C-B</td>
<td>0</td>
<td>(\frac{1}{2}) size</td>
</tr>
<tr>
<td>A-C-D</td>
<td>0</td>
<td>(\frac{1}{2}) size</td>
</tr>
<tr>
<td>A-F-D</td>
<td>0</td>
<td>(\frac{1}{2}) size</td>
</tr>
</tbody>
</table>

Content 1: Con Con
Content 2: Con Con

bottle neck
Sub-content

- **Time-aware Sub-Content Rapid-Data-Evacuation Algorithm (TA-SCRDE)**
  - For each destination DC, there is a path-pair
    - B: (P1,P2) D:(P3,P4)
  - Divide one content into two sub-content with different size according to the maximum time delay of each path
  - Principle 1: if the delay time of each path is zero, the content is divided into two equal-size sub-contents

<table>
<thead>
<tr>
<th>Path</th>
<th>Delay time</th>
<th>Assigned size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>0</td>
<td>½ size</td>
</tr>
<tr>
<td>A-C-B</td>
<td>0</td>
<td>½ size</td>
</tr>
<tr>
<td>A-C-D</td>
<td>0</td>
<td>½ size</td>
</tr>
<tr>
<td>A-F-D</td>
<td>0</td>
<td>½ size</td>
</tr>
</tbody>
</table>

![Diagram showing path delays and sizes](image)
Sub-content

- **Time-Aware Sub-Content Rapid-Data-Evacuation Algorithm (TA-SCRDE)**
  - Principle Two: if the delay time of one path is 0 and the other isn’t, we will evacuate the entire content by the path with 0 delay time
  - Principle Three: if both of the paths have a delay time, we will divide the content by the following formulation
    \[
    \frac{t_0}{t_0 + t_1} = \frac{S_{p_1}}{S_{p_2}}
    \]
    \[
    S_{p_1} = \frac{t_0 \cdot S_c}{(t_0 + t_1)} \quad \quad S_{p_2} = \frac{t_1 \cdot S_c}{(t_0 + t_1)}
    \]

---

<table>
<thead>
<tr>
<th>Path</th>
<th>Delay time</th>
<th>Assigned size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>t0</td>
<td>(S_{p_1})</td>
</tr>
<tr>
<td>A-C-B</td>
<td>t1</td>
<td>(S_{p_2})</td>
</tr>
<tr>
<td>A-C-D</td>
<td>t1</td>
<td>0</td>
</tr>
<tr>
<td>A-F-D</td>
<td>0</td>
<td>(S_c)</td>
</tr>
</tbody>
</table>

---

[Diagram showing paths and delay times]
Sub-content

- Time-aware Sub-Content Rapid-Data-Evacuation Algorithm (TA-SCRDE)
  - For each destination node, there is a path-pair
  - B: (P1,P2) D:(P3,P4)
Combination-content

1. Risk of data loss in the evacuation
2. Risk of poster-disaster
3. Waster of link capacity
Combination-content

- **Motivation**
  - The delivered content may be lost due to the poster-disaster
  - We must consider the risk of each datacenter
  - Protect the content by using multi-replicas

- **Advantage**
  - Multi-replicas which means a lower risk

- **Disadvantage**
  - Cost a longer time to delivery all the content

- **Trade-off between data transfer delay and risk**

- **Open Problems:**
  - The number of replicas
  - The combination of different weight and size contents
Combination-content

- Multi-Replicas Combination-Content Rapid-Data-Evacuation Algorithm (MR-CCRDE)
  - Combine the different weight and different size contents
  - Delivery them to the different risk DCs
  - Different from the RDE without risk, it is complex due to addition degree of risk and the combination of different contents
  - It is an multi-objectives optimization
  - Considering multi-cast/multi-path

<table>
<thead>
<tr>
<th>Con 1</th>
<th>Con 2</th>
<th>Con 3</th>
<th>Con 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>30G</td>
<td>25G</td>
<td>15G</td>
<td>20G</td>
</tr>
</tbody>
</table>

Link capacity: 80G

Link capacity: 50G
## Conclusion

<table>
<thead>
<tr>
<th>Schemes</th>
<th>Consideration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time efficiency</td>
</tr>
<tr>
<td>N-RDE</td>
<td>x</td>
</tr>
<tr>
<td>LD-RDE</td>
<td>√</td>
</tr>
<tr>
<td>B-SCRDE/TA-SCRDE</td>
<td>√</td>
</tr>
<tr>
<td>MR-CCRDE</td>
<td>√</td>
</tr>
</tbody>
</table>
Thank you for your attention!

Presenter: Yongcheng(jeremy) Li
PhD student, School of Electronic and Information Engineering, Soochow University, China
Email: liyongcheng621@163.com

Group Meeting, Apr. 08 2016