Risk-Aware Rapid Data Evacuation 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, Friday, August 26, 2016

UCDAVIS

Outline

1. Background
2. Risk-Aware Rapid Data Evacuation For Large-Scale Disasters
3. Heuristic Algorithm
4. Performance Evaluation
5. Conclusion
6. Future Work

Background

- Enterprises deploy their cloud services such as cloud data storage and applications in distributed datacenter (DC) networks.
- Cloud services require *Terabytes* or *Petabytes* of data transfer.
- Optical networks can be used to facilitate data transfers.
  - Advantage: high bandwidth and low latency in inter-DC networks.
  - Disadvantage: services can be disrupted by disasters (such as earthquakes, tornadoes, and intentional attacks).
- A large-scale disaster can lead to high data loss and service disruptions.
  - 2011 Japan Earthquake damaged many cloud providers' data.

Background (contd.)

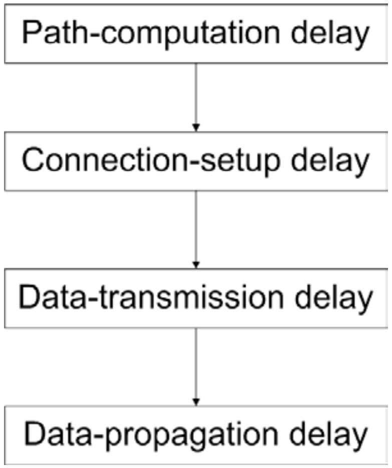
- Provide redundancy (and protection) against data loss.
  - Distributed content/service replicas in different DCs.
- All replicas of a content can be lost in a large-scale disaster.
  - Critical data must be quickly evacuated from DCs in the disaster region to safe DCs prior to the disaster.
- Rapid data evacuation.
  - Receive warning of an oncoming disaster from various sensors and monitors in their network and/or from government or intelligence agencies.
  - Depending on the type of disaster (e.g., earthquake, hurricane or weapons of mass destruction (WMD)), do the following prediction.
    - Disaster zone
    - Evacuation deadline
    - Potential damage in the infrastructure
  - Before deadline, quickly evacuate as much critical data as possible.

Background (contd.)

- Safe data transfers
  - Possible independent disasters may compromise the process of data evacuation.
  - Risk of node/link failures must be considered to ensure safe data transfers.
  - How to objective a tradeoff between evacuation time and risk of data loss during evacuation.

Risk-Aware Rapid Data Evacuation For Large-Scale Disasters (contd.)

- Time delay
  - Path-computation delay.
  - Connection-setup delay.
  - Data-transmission delay.
  - Data-propagation delay.
- Notations
  - Distance of path: l.
  - Number of hops on path: n.
  - Bandwidth of path: B_p
  - Propagation delay per unit distance: μ.
  - Processing delay: η.
  - Switch configuration delay: β.
  - Assume the same propagation delay for data and control messages.



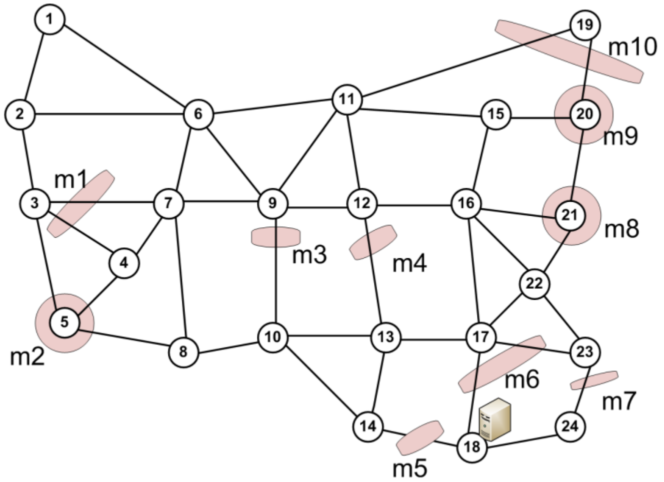
Risk-Aware Rapid Data Evacuation For Large-Scale Disasters (contd.)

- Equations
  - Connection-setup delay
    - Control-message processing delay: $(n + 1) * \eta$
    - Control-message propagation delay: $l * \mu$
    - Switch-configuration delay: $(n + 1) * \beta$
  - Transmission delay: F_c / B_p
  - Propagation delay: $l * \mu$

Risk-Aware Rapid Data Evacuation For Large-Scale Disasters (contd.)

- The risk of node/link failures
  - $risk_l^m/risk_n^m$ are the probabilities of link l or node n being damaged due to disaster $m \in \boldsymbol{M}$.
  - $\prod_{l \in \boldsymbol{L}, m \in \boldsymbol{M}}(1 - risk_l^m) \cdot \prod_{n \in \boldsymbol{N}, m \in \boldsymbol{M}}(1 - risk_n^m)$ computes the probability that all links and nodes are not damaged by any disaster.
  - Path failure probability:

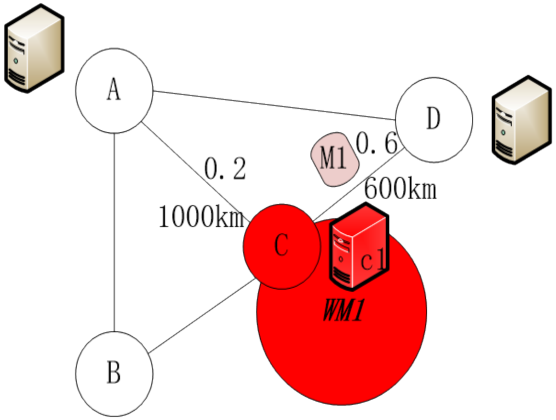
$$risk_p = 1 - \prod_{l \in \boldsymbol{L}, m \in \boldsymbol{M}}(1 - risk_l^m) \cdot \prod_{n \in \boldsymbol{N}, m \in \boldsymbol{M}}(1 - risk_n^m).$$



UCDAVIS
UNIVERSITY OF CALIFORNIA

Risk-Aware Rapid Data Evacuation For Large-Scale Disasters (contd.)

- Example
  - Transfer content c1 from node C to safe DCs A or D as fast as possible.
  - Failure probabilities of links *C-D* and *A-C* is assumed as 0.6 and 0.2.
  - Select destination DC *D*
    - Evacuation time is less but the risk along path *C-D* is higher.
  - Select destination DC *A*
    - Evacuation time is high but the risk along path *C-D* is less.



Risk-Aware Rapid Data Evacuation For Large-Scale Disasters (contd.)

- Problem statement
  - Objective:
    - Achieve an optimal tradeoff between evacuation time and risk of data loss during evacuation.
  - Given inputs:
    - $\boldsymbol{N}$ is the set of nodes and $\boldsymbol{L}$ is the set of links.
    - Physical topology $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{L})$.
    - Set of possible disaster zones $\boldsymbol{M}$.
    - Set of predicted WMD attack zones $\boldsymbol{WM}$.
    - Set of DCs $\boldsymbol{D}$.
    - Storage capacity $\boldsymbol{S_d}$.
    - Set of locally hosted contents $\boldsymbol{C_d}$.
    - Number of replicas of content $\boldsymbol{c}$ $\boldsymbol{R_c}$.
    - Importance metric of content $\boldsymbol{c}$ $\boldsymbol{\alpha_c}$.

UCDAVIS
UNIVERSITY OF CALIFORNIA

Risk-Aware Rapid Data Evacuation For Large-Scale Disasters (contd.)

- Problem statement
  - Input:
    - Size of content $\boldsymbol{c}$ F_c.
    - Residual link capacity $\boldsymbol{B_l}$, $\boldsymbol{l} \in \boldsymbol{L}$.
    - Evacuation deadline $\boldsymbol{T}$.
    - Set of *k*-shortest paths $\boldsymbol{R_{ij}}$ for each node pair $(\boldsymbol{i}, \boldsymbol{j})$, $\boldsymbol{i}, \boldsymbol{j} \in \boldsymbol{N}$.
  - Constraint:
    - Available link capacity is limited.
    - Data-transfer delay to be upper bounded by evacuation deadline.
    - Different paths can be used in parallel for different connections if paths do not overlap.

Heuristic Algorithm

- Disaster mapping:
  - Get a set of DCs $\boldsymbol{D_{wm}} \in \boldsymbol{D}$, $\boldsymbol{wm} \in \boldsymbol{WM}$. *// Datacenters in predicted disaster zone*
  - Get a set of DCs $\boldsymbol{D_{out\_wm}} \in \boldsymbol{D}$, $\boldsymbol{D_{out\_wm}} = \boldsymbol{D} - \boldsymbol{D_{wm}}$. *// Datacenters outside predicted disaster zone*
- Content selection:
  - For each DC $\boldsymbol{d} \in \boldsymbol{D_{wm}}$
    - Get a set of contents, $\boldsymbol{C_d}$, $\boldsymbol{d} \in \boldsymbol{D_{wm}}$, $\boldsymbol{C_{wm}} = \boldsymbol{C_{wm}} \cup \boldsymbol{C_d}$, $\boldsymbol{wm} \in \boldsymbol{WM}$. *// Contents in predicted disaster zone*
  - For each content $\boldsymbol{c} \in \boldsymbol{C_{wm}}$
    - If all replicas of content $\boldsymbol{c}$ are in the disaster zone $\boldsymbol{WM}$ then
      - Put *c* in a content list $\boldsymbol{C_{Eva}}$ and get set of DCs $\boldsymbol{D_c}$, which host the replicas of *c*. *// $\boldsymbol{C_{Eva}}$ is a set of contents to be evacuated*
    - Sort $\boldsymbol{C_{Eva}}$ based on α_c in a descending order.

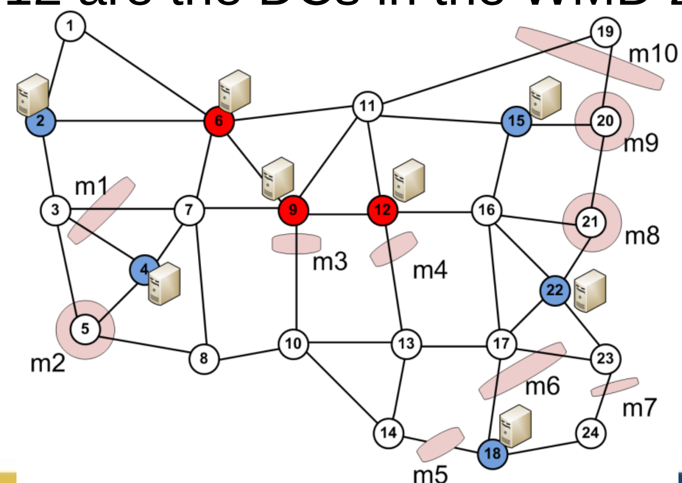
Heuristic Algorithm

- Destination DC selection and path delay and risk computation:
  - For each content $\boldsymbol{c} \in \boldsymbol{C_{Eva}}$
    - For each $\boldsymbol{d} \in \boldsymbol{D_{out\_wm}}$, if $\boldsymbol{F_c} \leq \boldsymbol{S_d}$ then put $\boldsymbol{d}$ into list of available DCs $\boldsymbol{D_{Ava}}$
      // $\boldsymbol{D_{Ava}}$ is a set of safe DCs with available storage
    - Get set of *k*-shortest paths $\boldsymbol{R_{ij}}$, $\boldsymbol{i} \in \boldsymbol{D_c}$,$\boldsymbol{j} \in \boldsymbol{D_{Ava}}$
    - For each path $\boldsymbol{p} \in \boldsymbol{R_{ij}}$
      - Obtain total delay $\boldsymbol{delay_p}$ and risk $\boldsymbol{risk_p}$ to calculate “general cost” $\boldsymbol{Cost_p}$
      - $Cost_p = delay_p + \varphi \cdot risk_p$
    - Set path $\boldsymbol{p}^*$ which has minimum cost $\boldsymbol{Cost_{p^*}}$ as final solution

UCDAVIS
UNIVERSITY OF CALIFORNIA

Performance Evaluation

- Simulation conditions
  - Network topology: 24-node USNET topology.
  - Disaster type: WMD attacks.
  - A predicted WMD zone: ***WM***
  - A set of 10 possible independent disaster zones: ***M.***
  - Blue nodes represent safe DCs.
  - Red nodes 6, 9, and 12 are the DCs in the WMD zone ***WM***.

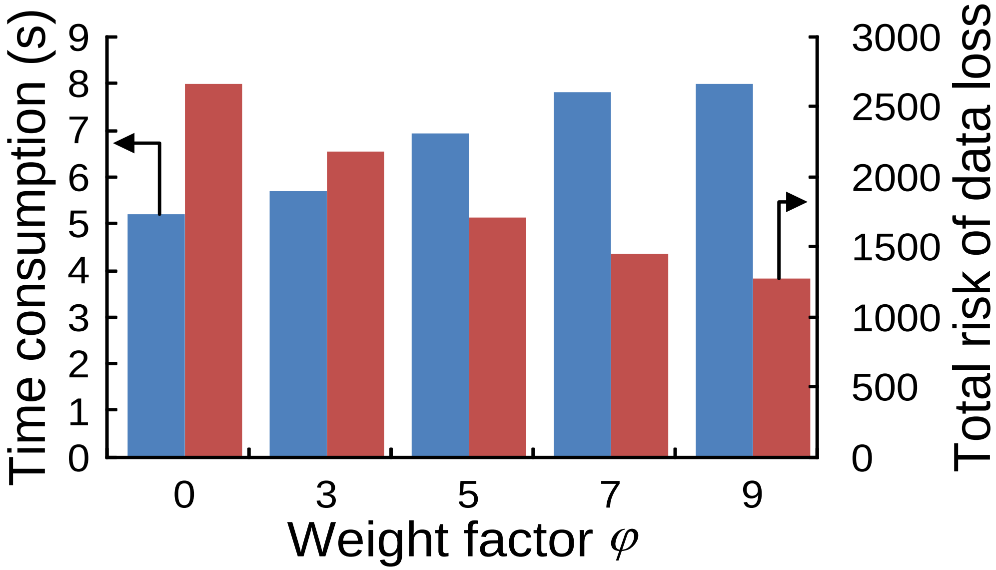


Performance Evaluation (contd.)

- Large storage capacity ranging from 10TB to 100TB (average occupation is assumed to be 40%).
- Residual link capacity is assumed to range from 500Gbps to 1Tbps (network is assumed to have 30% utilization).
- Number of contents is assumed to be 300.
- Size of each content is randomly generated within the range of [100GB, 200GB].
- The contents are uniformly distributed among different DCs with the number of replicas ranging from 2 to 4.
- Contents are randomly assigned the importance metric α_c on a scale from 1 to 10.
- Processing delay, propagation delay, and switch configuration delay to be 10 μs, 5 μs/km, 15 ms.

Performance Evaluation (contd.)

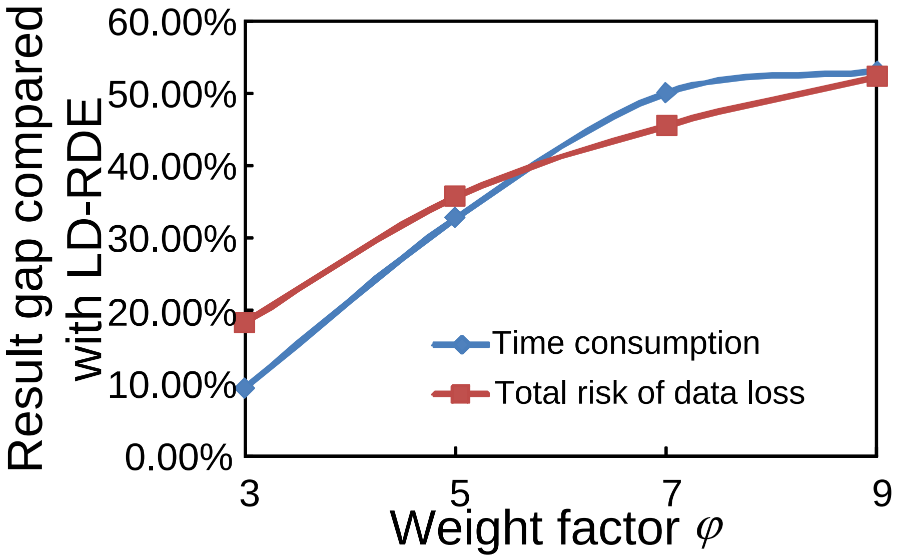
- Total risk of data loss and evacuation time



- **Shows the time consumption and the total risk of data loss for the RA-RDE algorithm with an increasing weight factor φ.**
- **We see that RA-RDE can significantly reduce the total risk with an increasing φ.**
- **This is reasonable since a larger weight factor can lead to higher risk reduction in the "general cost".**

Performance Evaluation (contd.)

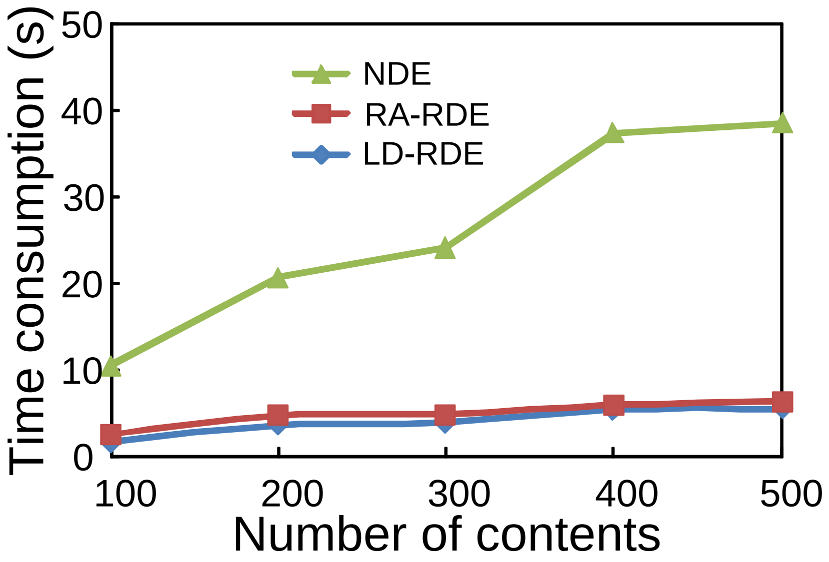
- Performance comparison



- **Compare the performance of RA-RDE with the LD-RDE algorithm which selects the least-delay paths for data evacuation without considering risk.**
- **It should be noted that the RA-RDE algorithm is equivalent to the LD-RDE scheme when $\varphi = 0$.**
- **We see that, when $\varphi = 3$, our approach reduces the total risk by 20% and needs less than 10% additional time consumption.**

Performance Evaluation (contd.)

- Time consumption with an increasing number of contents



- **Compare time consumption of the two rapid evacuation approaches with the nearest data evacuation (NDE) approach which evacuates data only to the nearest DC.**
- **With an increasing number of contents from 100 to 500 and $\varphi = 3$, we can see that RA-RDE performs close to LD-RDE and is much better than NDE, which verifies its time efficiency.**

Conclusion

- To balance performance between time consumption and total risk of data loss, we defined a “general cost” considering path delay and path risk using a weight factor.

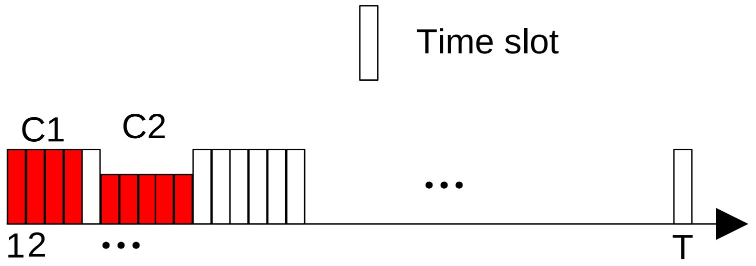
- We develop a risk-aware rapid data evacuation scheme for large-scale disasters in optical cloud networks.

- Results show that proposed approach significantly reduces total risk with minimal addition time consumption.

- Time consumption close to LD-RDE under different number of contents.

Future Work

- Try to propose an MILP model to solve the rapid data evacuation problem by using time slot
  - Time slot contiguity
  - Time slot continuity



- Investigate an new heuristic algorithms with the ICC paper of Wu Yu

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, Friday, August 26, 2016