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

UCDAVIS

Outline

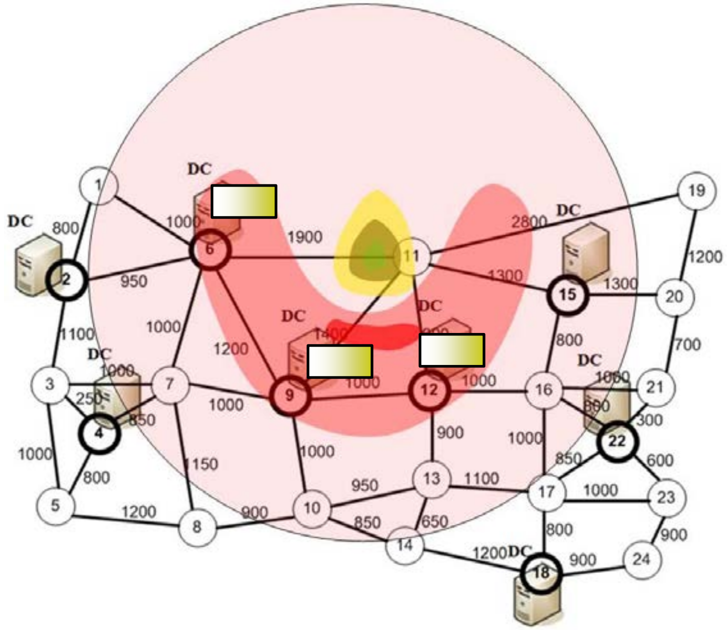
1. Background
2. Rapid data evacuation for large-scale disasters
3. Problem statement
4. Rapid data evacuation schemes
5. Conclusions

Background

- Many famous company have given there *cloud services* for the customers in the world wide
- Cloud services are delivered by *datacenter (DC) networks*, which provides the storages of many important contents
- The datacenter networks may be damaged by natural (i.e., hurricane and earthquake) or man-made (i.e., weapons of mass destruction attacks) disasters, which will lead to the *service disruptions* and *permanent data loss*
- In case of *large-scale disasters*, there is a risk that multiple DCs are damaged and the replicas of a content are within the disaster region and are 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*

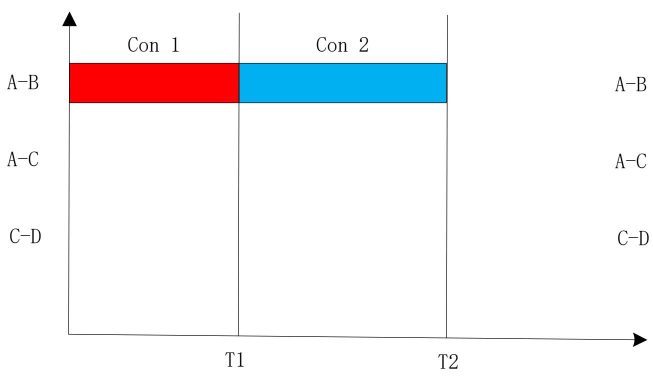
Delay Computation

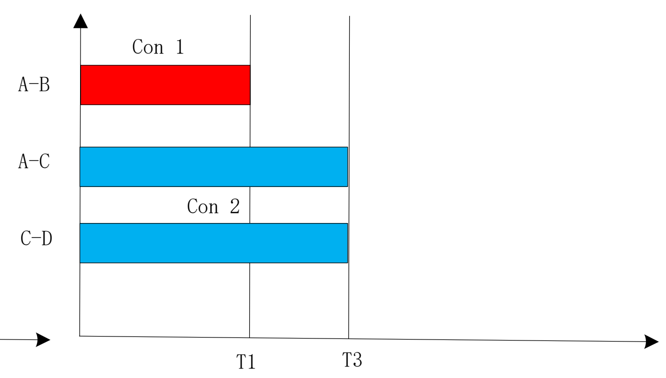
- path-computation delay
  - Time consumption of the path computation
- Connection-setup delay
  - control-message processing delay: $(n + 1) \times \eta$
  - control-message propagation delay: $l \times \mu$
  - switch-configuration delay: $(n + 1) \times \beta$
- Transmission delay: F_C / B_P
- Propagation delay $l \times \mu$

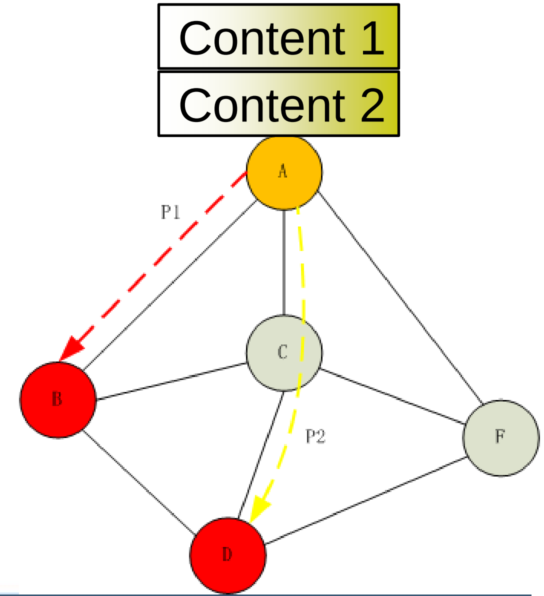
n is the number of hops on path. l is the distance of path. μ is the propagation delay per unit distance. η is processing delay. β is the switch configuration delay. F_C is the size of content C. B_P is the available bandwidth of the path P

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



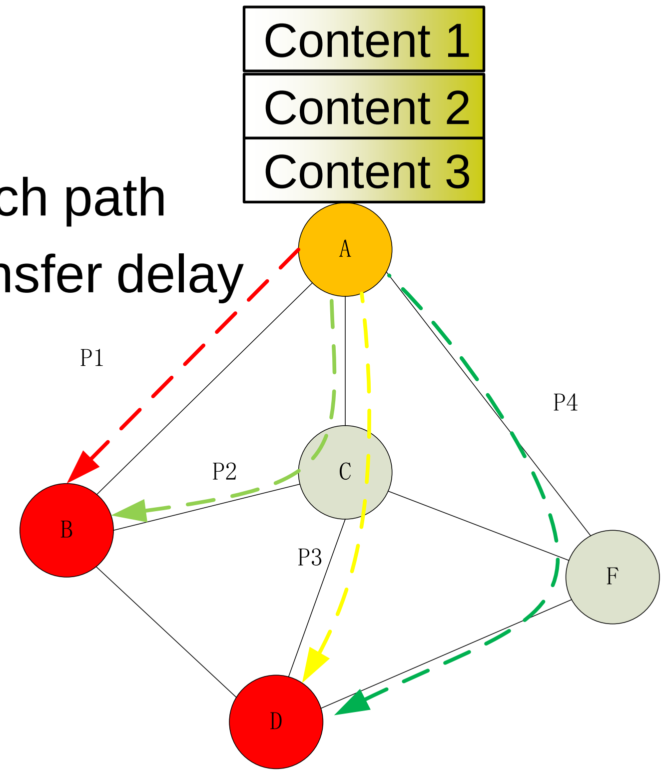


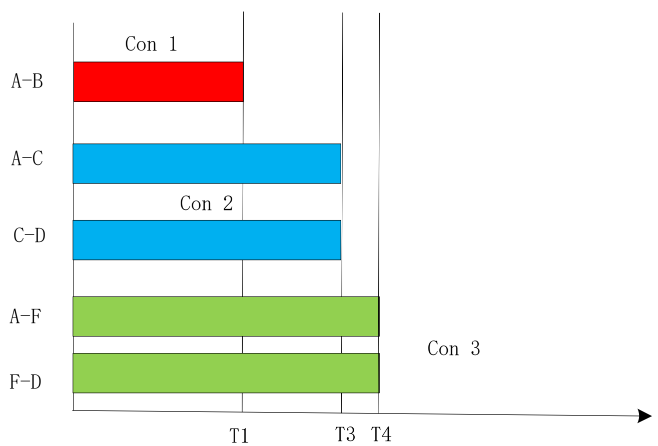


UCDAVIS
UNIVERSITY OF CALIFORNIA

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





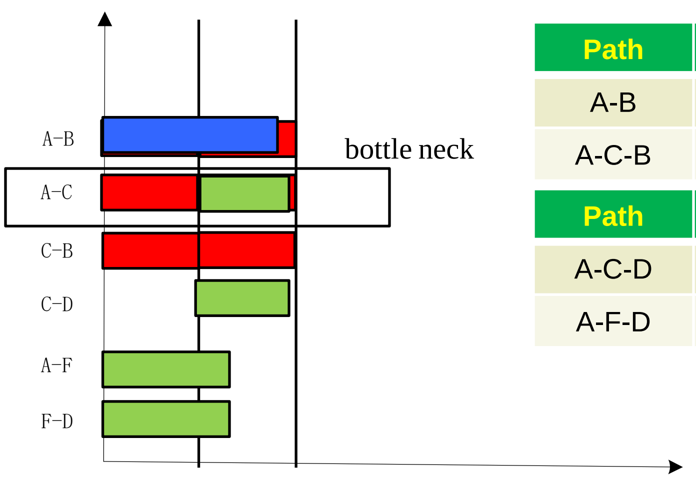
[1] S. Ferdousi, M. Habib, M. Tornatore, and B. Mukherjee, "Rapid data evacuation for large-scale disasters in optical cloud networks," in *Proc. OFC 2015*.

Sub-content

- Example
  - Divide the entire content into serval 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

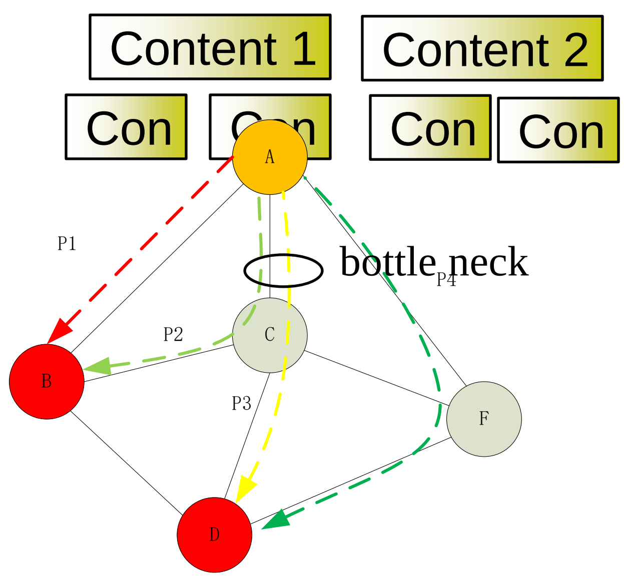
Sub-content

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



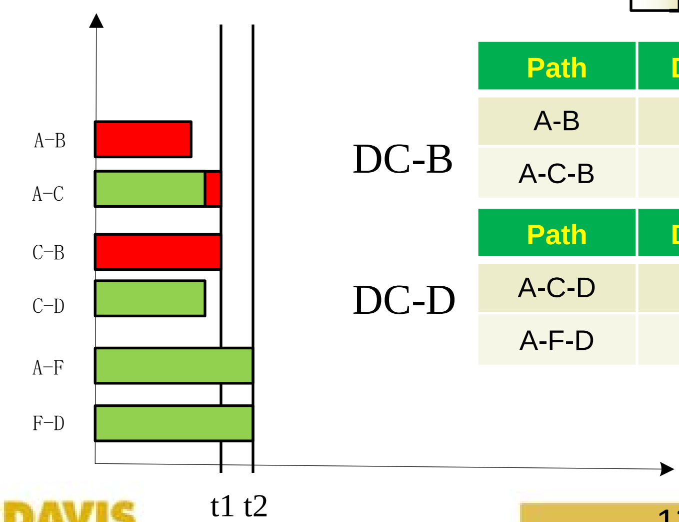
Path	Delay time	Assigned size
A-B	0	½ size
A-C-B	0	½ size

Path	Delay time	Assigned size
A-C-D	0	½ size
A-F-D	0	½ size



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

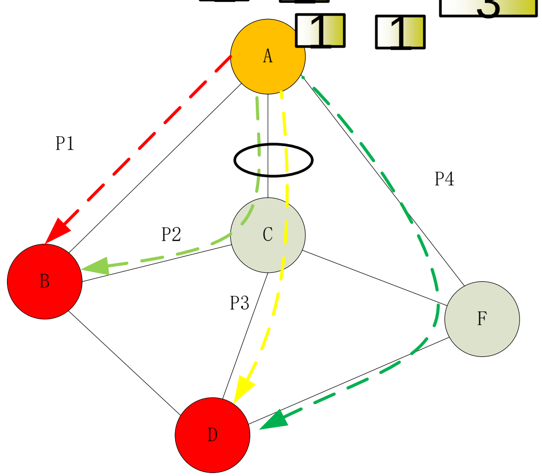


DC-B

Path	Delay time	Assigned size
A-B	0	½ size
A-C-B	0	½ size

DC-D

Path	Delay time	Assigned size
A-C-D	0	½ size
A-F-D	0	½ size

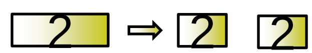


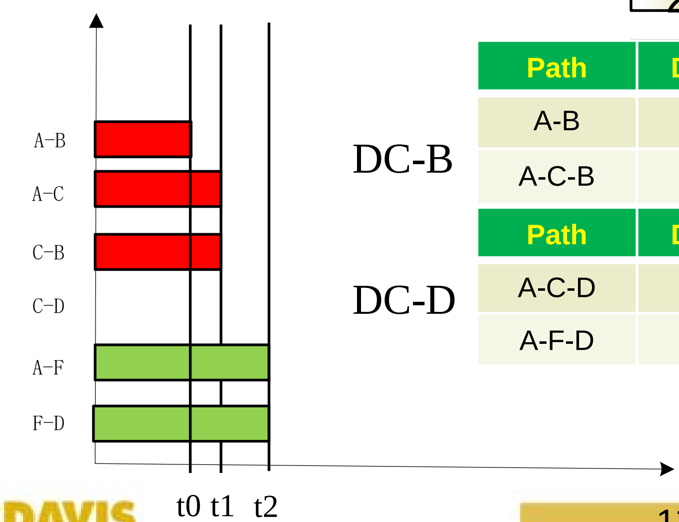
Sub-content

- 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

$$t0/t1 = S_{p1}/S_{P2} \qquad S_{p1} + S_{p2} = S_C$$

$$\Rightarrow \quad S_{p1} = \frac{t0 \cdot S_C}{(t0+t1)} \qquad S_{p2} = \frac{t1 \cdot S_C}{(t0+t1)}$$



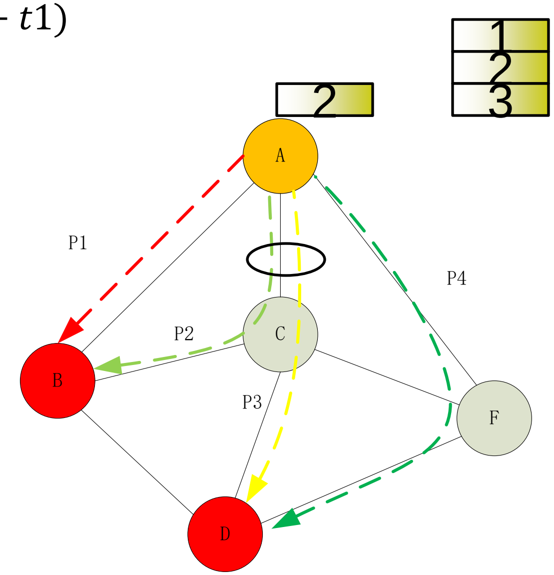


DC-B

Path	Delay time	Assigned size
A-B	t0	S_{p1}
A-C-B	t1	S_{p2}

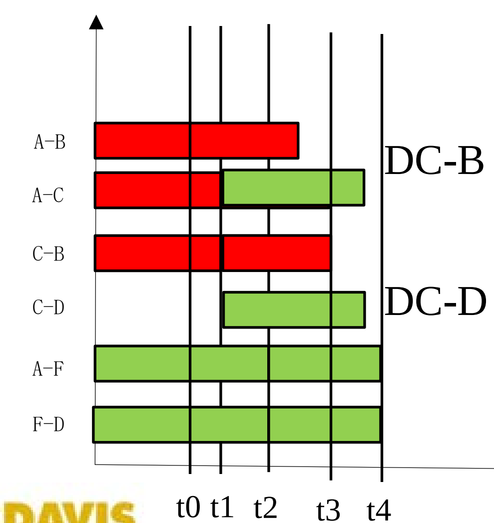
DC-D

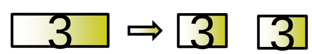
Path	Delay time	Assigned size
A-C-D	t1	0
A-F-D	0	S_C



Sub-content

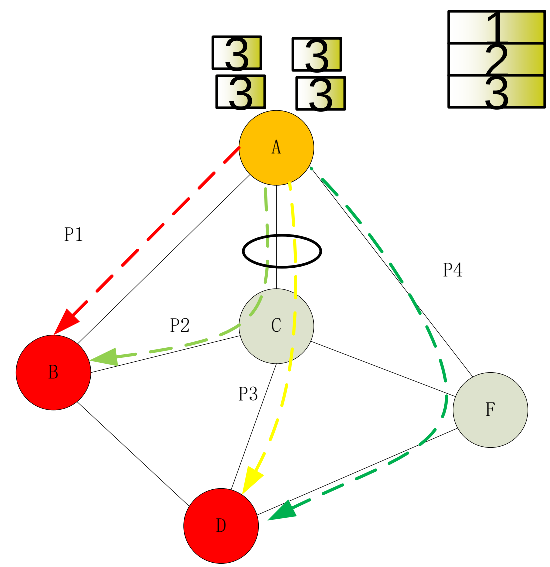
- Time-aware Sub-Content Rapid-Data-Evacuation Algorithm (TA-SCRDE)
  - For each destination node, there is a path-pair
    - B: (P1,P2) D:(P3,P4)



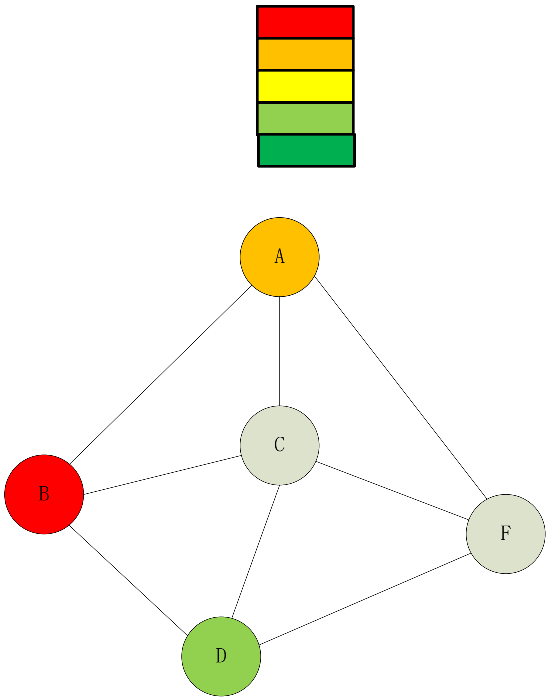


Path	Delay time	Assigned size
A-B	t0	S_{p1}
A-C-B	t1	S_{p2}

Path	Delay time	Assigned size
A-C-D	t1	S_{p3}
A-F-D	t2	S_{p4}



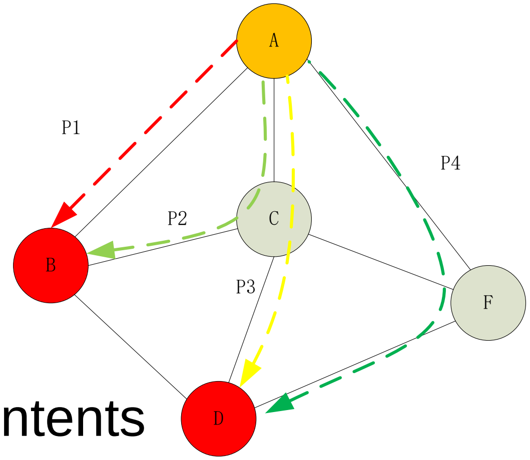
Combination-content



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

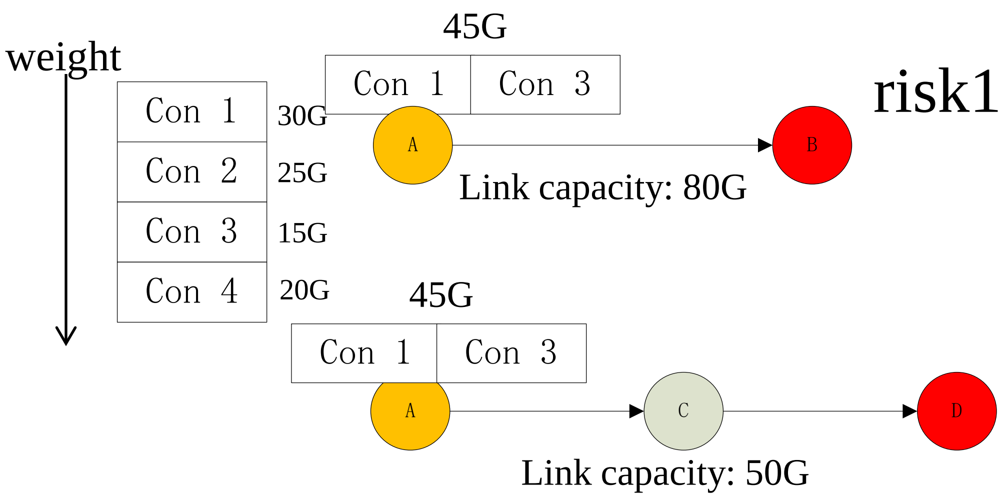
Combination-content

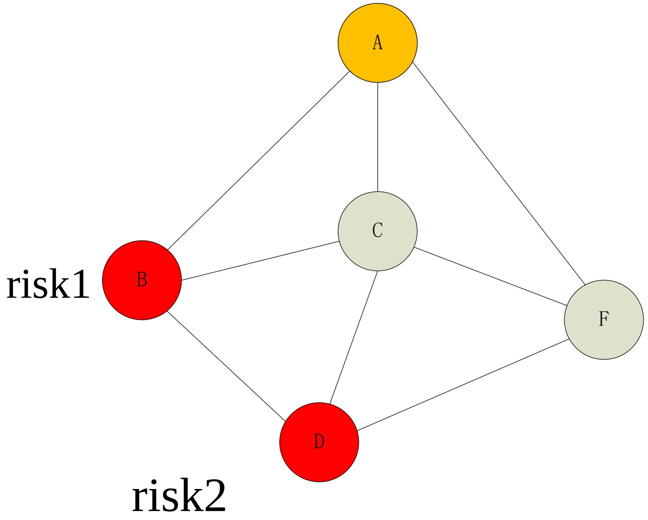
- Motivation
  - The delivered content may be lost due to the poster-disaster
  - We must consider the risk of each datacenter
  - Protect the content by using multi-replicas
- Advantage
  - Multi-replicas which means a lower risk
- Disadvantage
  - Cost a longer time to delivery all the content
- Trade-off 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





Conclusion

Schemes	Consideration					
	Time efficiency	Protection	Multi-path	Sub-content	Combination-content	Poster-disaster
N-RDE	x	x	x	x	x	x
LD-RDE	√	x	√	x	x	x
B-SCRDE/TA-SCRDE	√	x	√	√	x	x
MR-CCRDE	√	√	√	x	√	√

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

UCDAVIS
UNIVERSITY OF CALIFORNIA