Joint Progressive Recovery of Optical Network and Datacenters After Large-Scale Disasters

Sifat Ferdousi

October 14, 2016

UCDAVIS

Introduction

- Content providers offer variety of cloud services ranging from search, social networks, video sharing platforms to tools for online collaboration and more.

- To support such services, they build (or lease from other providers and operators) *datacenters (DCs)* and *optical networks*, both to interconnect their DCs and to achieve customer proximity.

Cloud networks

Storage

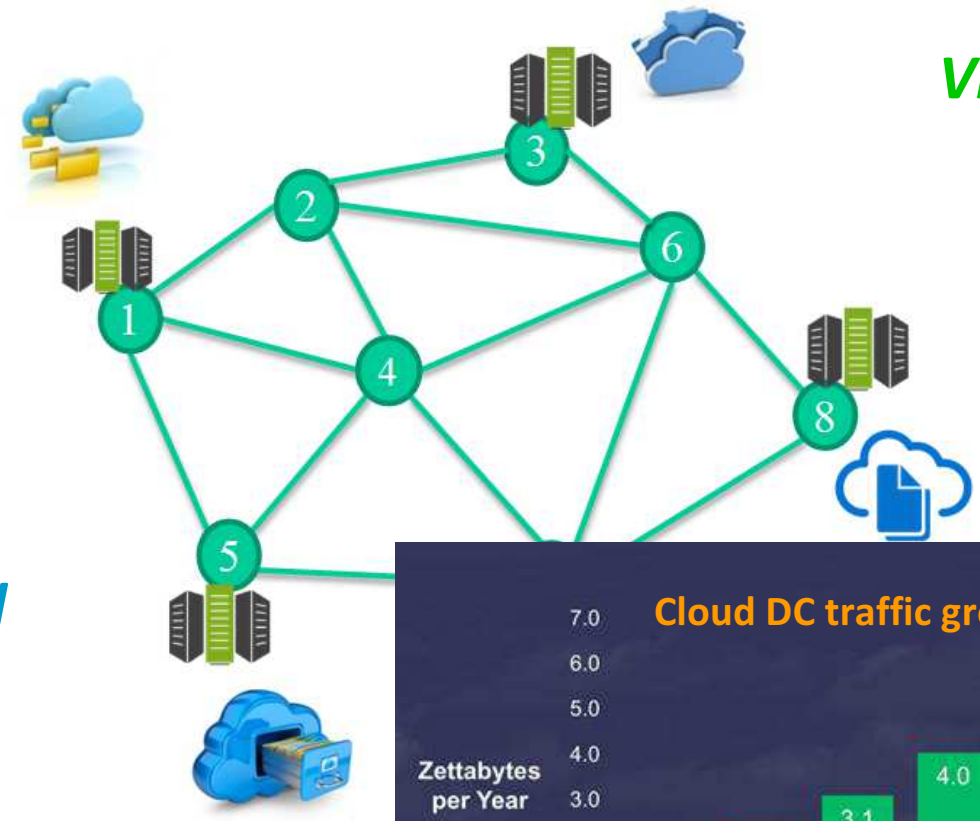
Videos

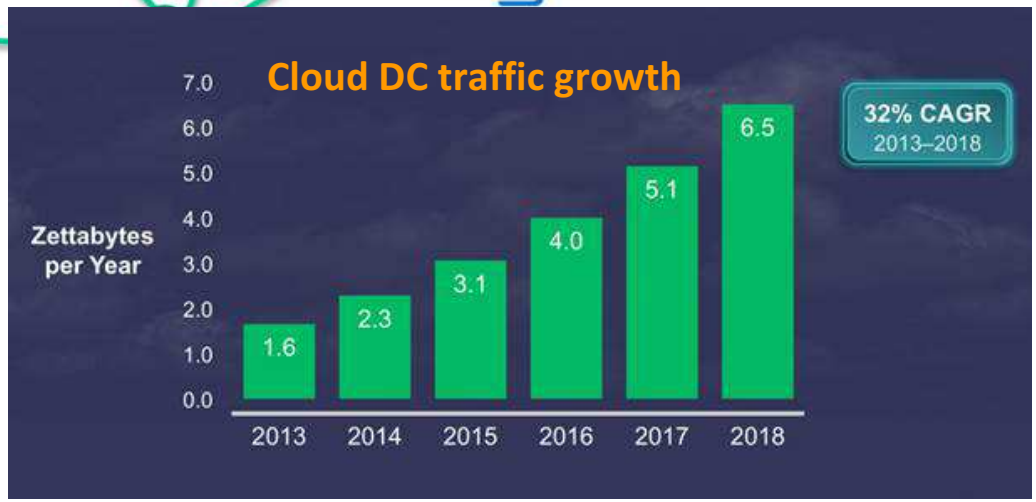
Data

Content

E-mail

Social networking





*Source: Cisco Cloud_Index_White_Paper

UCDAVIS
UNIVERSITY OF CALIFORNIA

Post-disaster recovery

- As more and more customers are migrating to cloud, it is crucial to guarantee survivability of these services.

- Disaster survivability in cloud networks usually focus on pre-disaster preparedness through pre-provision of backup resources - cannot guarantee full recovery at reasonable costs under multiple random/correlated failures due to disasters.

- ***Post-disaster recovery***

UCDAVIS
UNIVERSITY OF CALIFORNIA

Post-disaster recovery

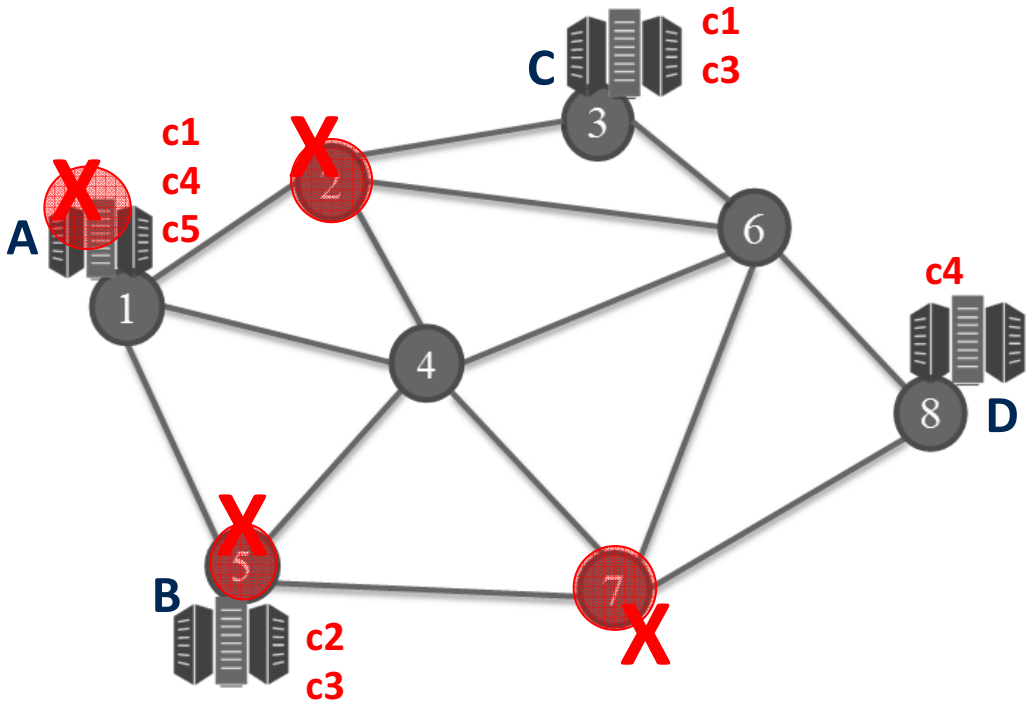
- After a disaster, *infrastructure repairs* are usually carried out in multiple stages (progressive)
  - only a subset of failed components may be repaired at a time due to limited availability of repair resources (equipment, repair crew, vehicles etc.).

- In post-disaster scenarios - both DC and underlying optical network infrastructures can be affected together,
  - a coordinated recovery plan can lead to more efficient repair resource allocation and restoration of services, *especially, for cloud providers that jointly own/manage DCs and networks*.

Joint progressive recovery of network and DCs

- We investigate joint progressive recovery of network and DCs after a large-scale disaster considering an optical network infrastructure that connects DCs and end users.
  - Generally, network recovery problems considers network connectivity in the failed portion of the network.
  - In our work, to ensure delivery of cloud services, we employ content connectivity (reachability of content from any point of a network).

- In post-disaster recovery, the sequence of optical network nodes/links and DCs to be repaired has an impact on users' reachability to important contents/services at a given time.



Considering all possibilities, it is intuitive that network recovery and DC recovery are inter-dependent to provide content reachability to users.

A *disjoint* recovery scheme cannot reflect inter-dependency because network repair phase will not be aware of the state of the DCs and the impact of content reachability.

Joint Progressive Net-DC Recovery

- We propose a Joint Progressive Net-DC Recovery (JR) algorithm which
  - schedules repair of optical network nodes/links and DCs (determines the sequence of network nodes/links and DCs to be repaired)
  - with the objective to attain maximum possible *cumulative weighted content reachability* to users at each stage of the repair process.

- The metric is a measure of reachability of all contents from a set of active DCs to a set of active users *within their latency bound* in the network based on importance of the contents.
  - (a content is reachable, if the content can be reached by users *within their latency bound* from at least 1 DC in the network.)

$$R = \sum_{s \in V} \sum_{c \in C} \{ R_{(c,s)} * \alpha_c \}$$

UCDAVIS
UNIVERSITY OF CALIFORNIA

Constraints

- In post-disaster scenarios, availability of resources can be severely constrained at each stage.
  - DC resources (servers, computing racks, storage disks, etc.)
  - Optical network resources (switching units, fibers, optical link transponders, regenerators, etc.)
  - Repair crew, vehicles, etc.

- We assume that at each stage, **one DC** ***and*** **one network node with its adjacent links with active end-nodes** can be repaired.

Joint Progressive Net-DC Recovery Algorithm

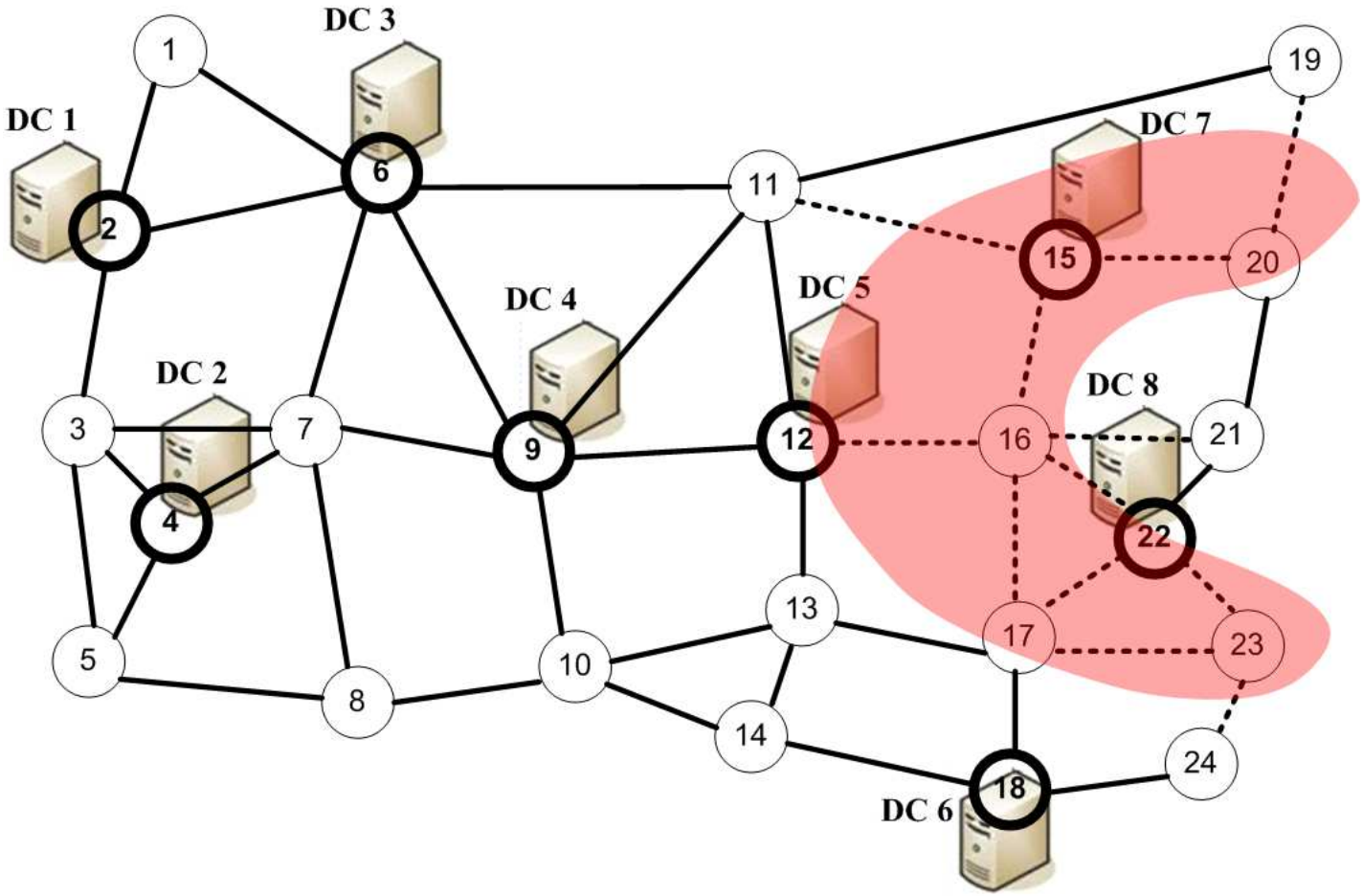
- **Input:**
  - Network topology $G(V;E)$
  - set of failed nodes and links $Gf\,(V_f, E_f)$
  - set of DCs D, set of failed DCs D_f
  - DC placement matrix P_{sd} indicating if DC d is supported by network node s
  - set of contents C
  - importance factor α_c of content c
  - content placement matrix H_{dc} indicating if content c is hosted in DC d
  - latency bound L
  - repair stage counter k.

- **Output:**
  - Set of repaired nodes V_{rep}
  - set of repaired links E_{rep}
  - set of repaired DCs D_{rep}.

UCDAVIS
UNIVERSITY OF CALIFORNIA

Initialization at stage $k = 1$, $V_f^k \leftarrow V_f, E_f^k \leftarrow E_f, D_f^k \leftarrow D_f$.

- While $\{V_f^k$ *and* E_f^k *and* $D_f^k\}$ *!empty*
  1. **Network Node Repair Phase:**
     (a) If any node $s \in V_f^k$ supports an active DC $d \in D_{rep}$, select node s for repair.
     (b) Else, for each node $s \in V_f^k$, calculate content reachability, R_s^k by considering network connectivity of the failed node s to the operational part of the network.
        i. Select node $s \in V_f^k$ with max R_s^k value for repair. If multiple nodes have same max R_s^k value, select node s which support a failed DC $d \in D_f^k$ for repair. If all/none of the nodes with same highest R_s^k value host a failed DC, select node s with the highest nodal degree for repair.
     (c) Set $V_{rep} \leftarrow \{V_{rep} \cup s\}$. Set $V_f^{k+1} \leftarrow \{V_f^k - s\}$.
  2. **Network Link Repair Phase:**
     (a) For the repaired node s, get the set of active neighboring nodes $N_s \in V_{rep}$.
        i. Select set of links $e_{s,j}, j \in N_s$ for repair.
     (b) Set $E_{rep} \leftarrow \{E_{rep} \cup e_{s,j}\}$. Set $E_f^{k+1} \leftarrow \{E_f^k - e_{s,j}\}$.
  3. **DC Repair Phase:**
     (a) If the repaired node s supports a failed DC $d \in D_f^k$, select DC d for repair.
     (b) Else, for each DC $d \in D_f^k$, if it has active supporting node $s \in V_{rep}$, calculate content reachability, R_d^k by considering DC d active in the operational part of the network.
        i. Select DC $d \in D_f^k$ with max R_d^k value for repair. If multiple DCs have same max R_d^k value, select DC d with the highest nodal degree for repair.
     (c) Else, for each DC $d \in D_f^k$, if it has failed supporting node $s \in V_f^k$, calculate cumulative content importance, W_d.
        i. Select DC $d \in D_f^k$ with max W_d value for repair. If multiple DCs have same max W_d value, select DC d with the highest nodal degree for repair.
     (d) Set $Drep \leftarrow \{D_{rep} \cup d\}$, set $D_f^{k+1} \leftarrow \{D_f^k - d\}$.
- Set $k = k + 1$.



UCDAVIS
UNIVERSITY OF CALIFORNIA

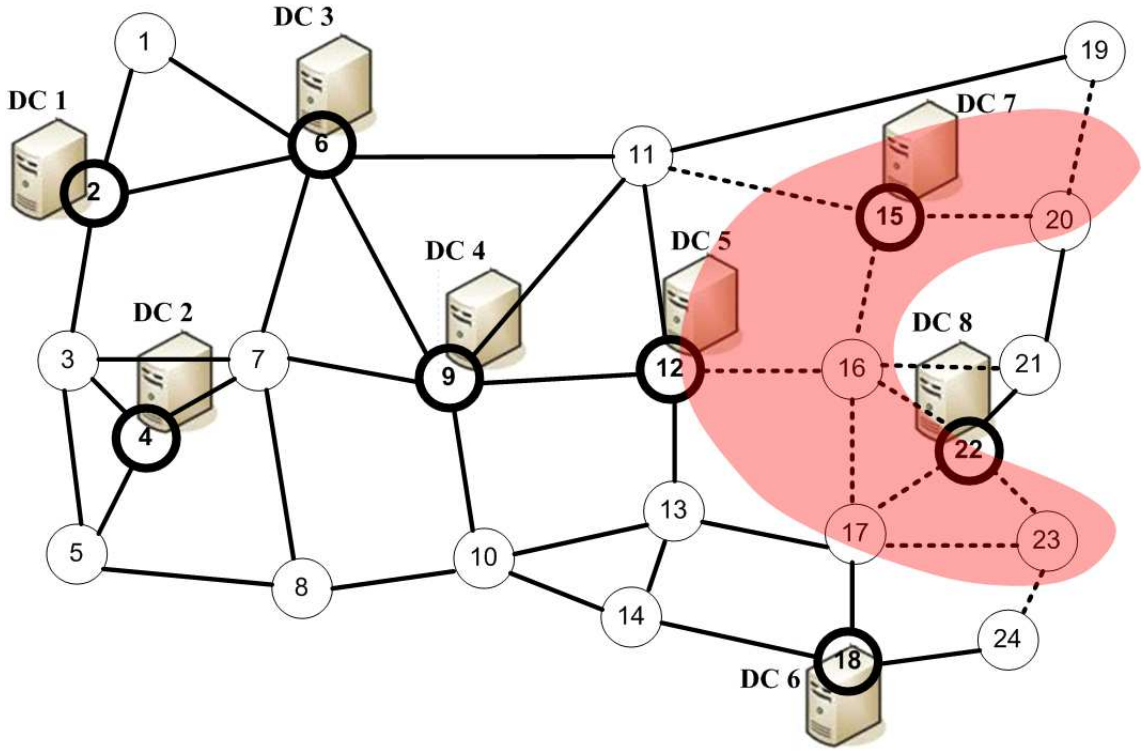
Joint vs. disjoint

- For our numerical analysis – we compare our *joint* network and DC recovery approach with a *disjoint* network and DC recovery approach
  - in which the network repair phase is independent from DC repair phase, i.e., network repair phase does not consider content reachability from users since it is not aware of the state of DCs.
  - In our joint recovery approach, nodes are repaired based on content reachability; in disjoint recovery approach nodes are repaired based on: (1) node's nodal degree in the original physical graph *G*, (2) random selection.
  - In disjoint recovery approach, after a network node is repaired, network link repair phase and DC repair phase are as in our joint recovery approach.

UCDAVIS
UNIVERSITY OF CALIFORNIA

Simulation settings

- 8 DCs mapped on 24-node USnet topology.
- Latency bound = 3 (hop count).
- Number of contents = 50.
  - Number of replicas/content, R_c: 2 ~ 4.
  - Content importance, α_c: on a scale of 1-10.

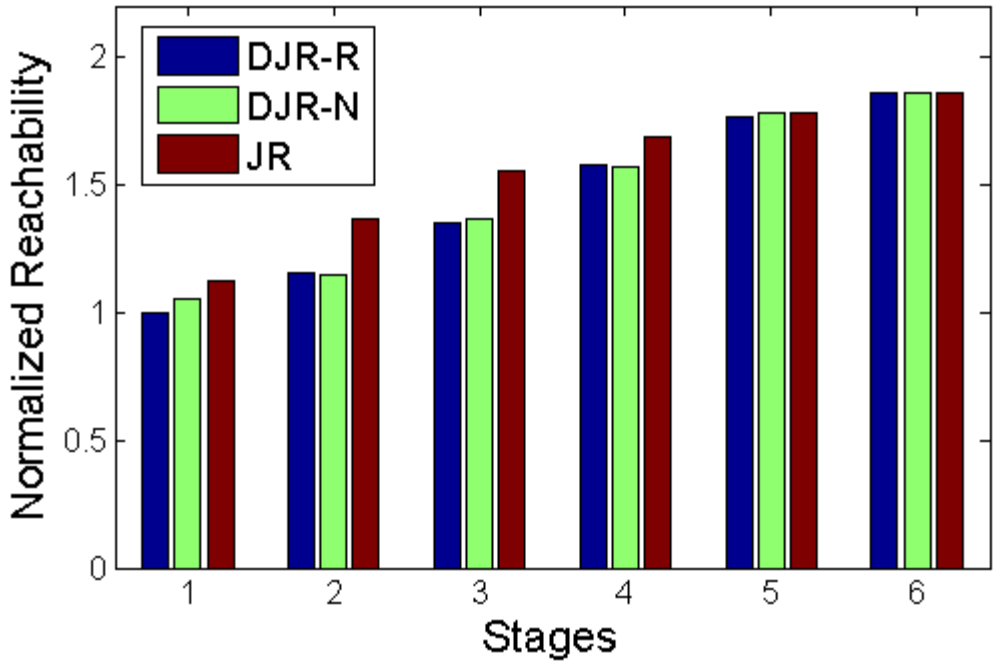


Sequence of node repair over the stages is:
nodes 12 > 16 > 22 > 17 > 15 > 20 > 23
Sequence of DC repair is: *DC5 > DC8 > DC7*

UCDAVIS
UNIVERSITY OF CALIFORNIA

Results

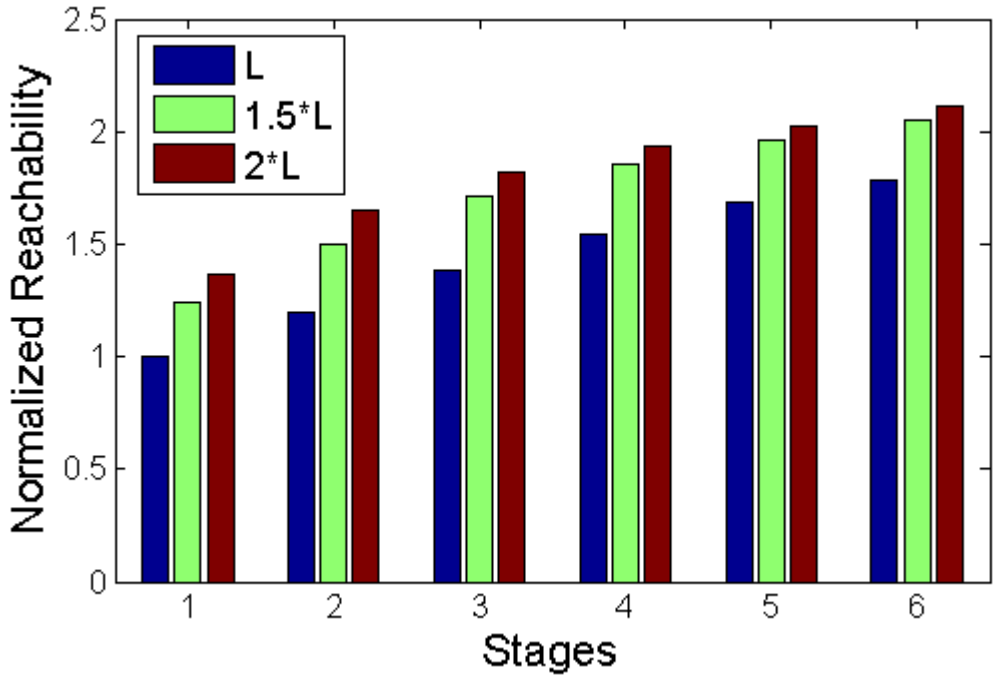
Comparing joint recovery and disjoint recovery approaches.



Our approach provides higher (about 7% to 20%) cumulative content reachability in the first 4 repair stages.

UCDAVIS
UNIVERSITY OF CALIFORNIA

Cumulative weighted content reachability with different latency bounds



UCDAVIS
UNIVERSITY OF CALIFORNIA

Future work

- **Constraints**
  - **Amount of DC repair resources**
  - **Amount network node repair resources, Length of fiber to be repaired**

- **Degraded service**