# Progressive Recovery of Virtual Infrastructure Services After Large Disasters

Mahsa Pourvali, Kaile Liang, Feng Gu, Hao Bai, Khaled Shaban, Samee Khan, Nasir Ghani
*University of South Florida, University of New Mexico, VM Ware, Qatar University, ND State University*

**Sifat Ferdousi**

December 2, 2016

UCDAVIS

*OFC 2016/ ICNC 2016

# Introduction

- Organizations are starting to outsource their physical infrastructure needs to third-party cloud infrastructure service providers, e.g., via *infrastructure as a service* (IaaS) offerings.
- These services use advanced network virtualization techniques to provision separate customized *virtual networks* (VN) for clients over common physical networks.

- A typical VN consists of VN nodes (datacenter computing/storage resources) interconnected by VN links (optical bandwidth connections).
- The VN nodes are *mapped* onto dedicated resources at datacenter sites whereas VN links are *mapped* onto underlying connections across substrate networks interconnecting datacenter sites [VN embedding (VNE) problem].

# Disaster resiliency

- Given the stringency of many business/mission critical applications, VN survivability is a major concern. Datacenter outages or network link failures can easily disrupt many VN mappings, causing service degradation and even service level agreement (SLA) violations.

- Large-scale disasters (natural disasters, malicious weapons of mass destruction (WMD) attacks, cascading power outages) are most challenging here as they can cause multiple correlated physical node and link failures and result in widespread VN service outages.

- Research on survivable VNE schemes have looked into single node/link failure recovery and multi-failure disaster mitigation.

- Nevertheless, these strategies focus on *pre-provisioning* backup VN node/link resources prior to faults and cannot guarantee recovery from all random multi-failure disasters/scenarios.

# Post-disaster recovery

- ***Post-disaster* VN service restoration**

- Damaged networks will likely be repaired over longer time spans (days, weeks) in a *progressive* staged manner as backup resources arrive, e.g., datacenter computing/ storage racks, optical link transponders, switching units, etc.

- At the same time, service providers will still want to maintain some partial *degraded* level of service for their affected customers. Hence, the careful placement/scheduling of repair resources during recovery stages is critical - even partial recovery of VN services can provide critical business continuity and minimize revenue carrier losses/penalties.

UCDAVIS
UNIVERSITY OF CALIFORNIA

## Post-disaster progressive recovery

- The authors propose post-disaster repair resource placement (scheduling) framework for progressive VN recovery.

- Clearly the sequencing of datacenter node and network link recoveries will have a direct impact on VN services, i.e., scheduling of repair resources. Other studies have only looked at progressive recovery for point-to-point connections and not complex VN demands.

UCDAVIS
UNIVERSITY OF CALIFORNIA

# Notations

- Network topology $G_s(V_s,E_s)$ [$V_s$ is the set of datacenter nodes and $E_s$ is the set of network links]
- Maximum resource capacity of $R_n^{max}$ for each node $n \epsilon V_s$, maximum bandwidth $B_e^{max}$ for each link $e \epsilon Es$
- VN request (IaaS demand) is given by an undirected graph, $G_a(V_a,E_a)$, where $V_a$ is the set of VN nodes and $E_a$ is the set of VN links
- Here each VN node $p \epsilon V_a$ requires $r(a)$ in node resources and each VN link $(p,q) \epsilon E_a$ requires $b(a)$ in bandwidth capacity

UCDAVIS
UNIVERSITY OF CALIFORNIA

- A disaster occurs at time $T_0$ and is followed by multiple recovery stages at times $T_k$, $k=1, 2; \ldots$ ($T_{k+1} > T_k$)

- After a disaster, all damaged nodes and links after a disaster event fail completely, i.e., available resource levels fall to zero.

- During these recovery stages, all failed nodes transition through several states, i.e., *fully-failed* to *partially-recovered* to *fully-recovered*.

- As repair resources are subsequently installed, these resource levels start to rise and eventually recover to their pre-fault maximums by the final recovery stage, i.e., $R_n^{max}$ or $B_e^{max}$ units.

- All failed and partially-recovered nodes and links in stage $k$ are eligible to receive repair resources: $F_v^k$ (eligible nodes) and $F_e^k$ (eligible links). $F_v^0$ and $F_e^0$ represents the initial set of failed nodes and links.

- Set of all affected VN demands are given by $A$. $V'_a$ and $E'_a$ represent the set of affected virtual nodes and virtual links in VN demand $a \epsilon A$.

UCDAVIS
UNIVERSITY OF CALIFORNIA

- Aggregate datacenter node resource loss:

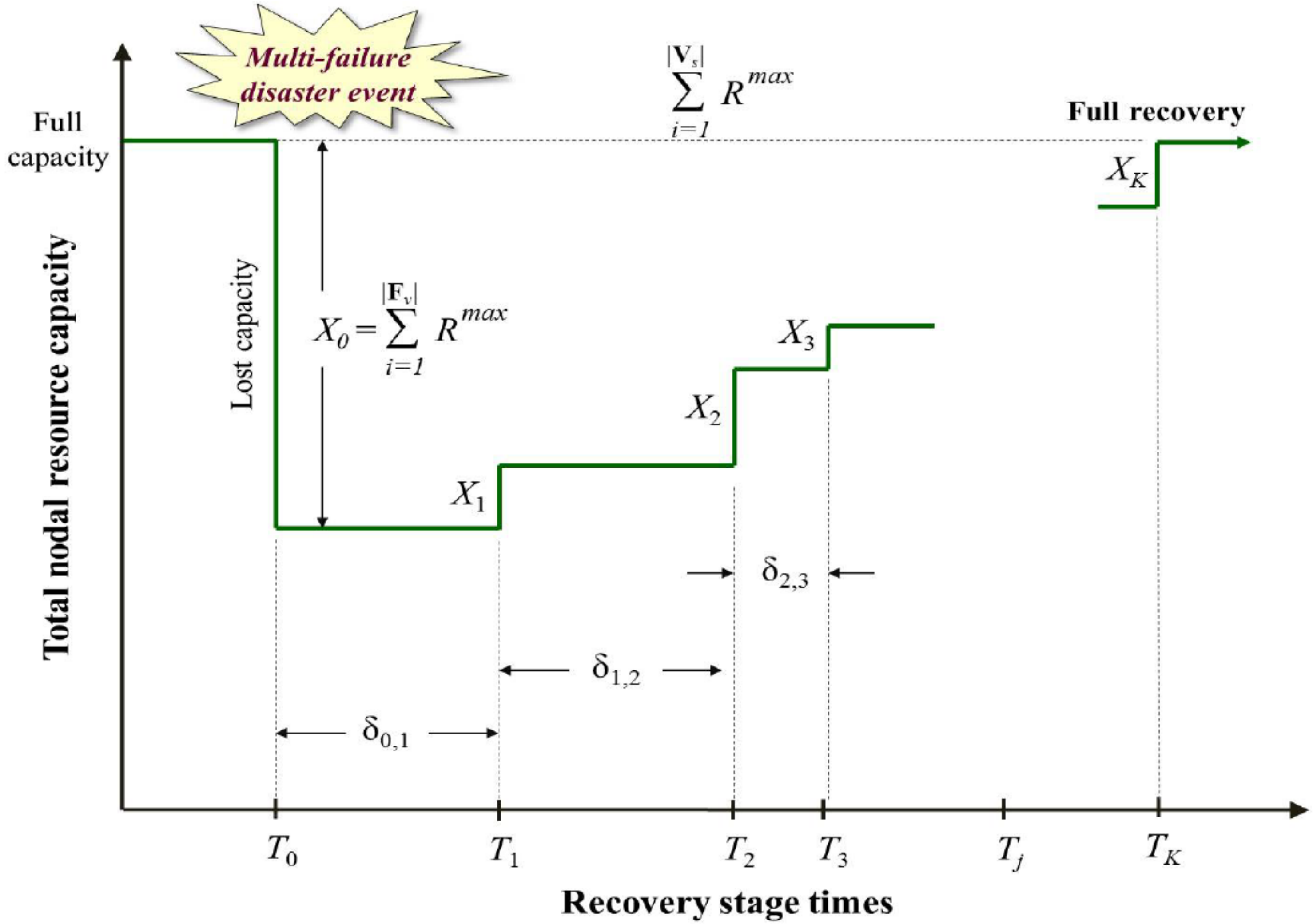
$$X_0 = \sum_{n \in \mathbf{F}_v^0} R_n^{max}$$

- Aggregate network link capacity loss

$$Y_0 = \sum_{(m,n) \in \mathbf{F}_e^0} B_{(m,n)}^{max}$$

- $X_k$ datacenter repair resources arrive in the *k*-th stage (computing racks, storage disks, etc) and $Y_k$ link repair resources arrive in the *k*-th stage (optical link transponders, switching units).

UCDAVIS
UNIVERSITY OF CALIFORNIA


Multi-failure disaster event
$\sum_{i=1}^{|\mathbf{V}_s|} R^{max}$
Full recovery
Full capacity
$X_K$
Total nodal resource capacity
Lost capacity
$X_0 = \sum_{i=1}^{|\mathbf{F}_v|} R^{max}$
$X_3$
$X_2$
$X_1$
$\delta_{2,3}$
$\delta_{1,2}$
$\delta_{0,1}$
$T_0$
$T_1$
$T_2$
$T_3$
$T_j$
$T_K$
Recovery stage times

# Progressive recovery schemes

- The progressive recovery schemes first distribute available repair resources among failed nodes/links and then use any standard VNE algorithm to remap failed demands. The main objective here is to allocate repair resources to minimize disruption, i.e., maximize the number of restored VN requests.

- Only eligible nodes with *at least one non-failed neighboring node* in $G_s(V_s,E_s)$ are considered, i.e., *candidate node* set $\boldsymbol{F'^{k}_{v}}$. (Here, the non-failed neighbors can be partially or fully-recovered nodes.)
-
- Similarly, only eligible links with *non-failed end-points* are considered, i.e., *candidate link* set $\boldsymbol{F'^{k}_{e}}$.

- Physical nodes and links can only receive repair resources up to the minimum of their allocated share or the amounts needed to achieve maximum capacity, i.e., full repair.

UCDAVIS
UNIVERSITY OF CALIFORNIA

# Repair Resource Placement Strategies

- Baseline Random Placement (RD)
- Uniform Placement (UP)
- Physical Degree (PD)
  - Distributed-PD
  - Selective-PD
- Virtual Load (VL)
  - Distributed-VL
  - Selective-VL

## Random Placement (RP)

- Assigns repair resources to failed physical datacenter nodes and network links in a random manner.

- Incoming node repair resources in stage $k$ ($X_k$) are placed at a random failed or partially-recovered node in $F'^{k}_{v}$.
- Similarly, incoming link repair resources ($Y_k$) are assigned to a random failed or partially-recovered link in $F'^{k}_{e}$.

- If the repair resources exceed the amount needed to fully recover the selected entity, then the leftover amounts are iteratively assigned to other randomly-selected nodes or link entities.

UCDAVIS
UNIVERSITY OF CALIFORNIA

## Uniform Placement (UP)

- Distributes repair resources in an even manner across damaged nodes and links.

- Each candidate datacenter node receives an even portion of the $X_k$ node repair resources in stage $k$. Each eligible candidate link also receives an even portion of the $Y_k$ link repair resources in stage $k$.

UCDAVIS
UNIVERSITY OF CALIFORNIA

# Physical Degree (PD)

- Assigns repair resources based upon the static physical node degree in the working pre-fault network topology.
- Repair resources are assigned to failed nodes with the highest number of physical links prior to failure, i.e., node degree in $G_s$.
- Similarly, link resources are assigned to links with the most connectivity before failure, i.e., endpoints with the highest node degree.
- The aim here is to place resources at datacenter nodes and network links that are more likely to support/ embed more VN nodes/links.

UCDAVIS
UNIVERSITY OF CALIFORNIA

# Virtual Load (VL)

- Assigns resources to candidate nodes/links based upon their dynamic VN load levels prior to failure.
- Repair resources are assigned to failed nodes (links) with higher numbers of embedded virtual nodes (virtual links) prior to failure.
- The goal here is to place resources at physical entities carrying more load in order to accelerate VN recovery. Therefore the candidate node and link sets are first sorted in descending order of their initial carried virtual node/link loads. The most loaded entities are then selected first

# Simulation setup

24 node/86 link topology.

All nodes have 100 units of capacity and all links have 10,000 units of bandwidth.

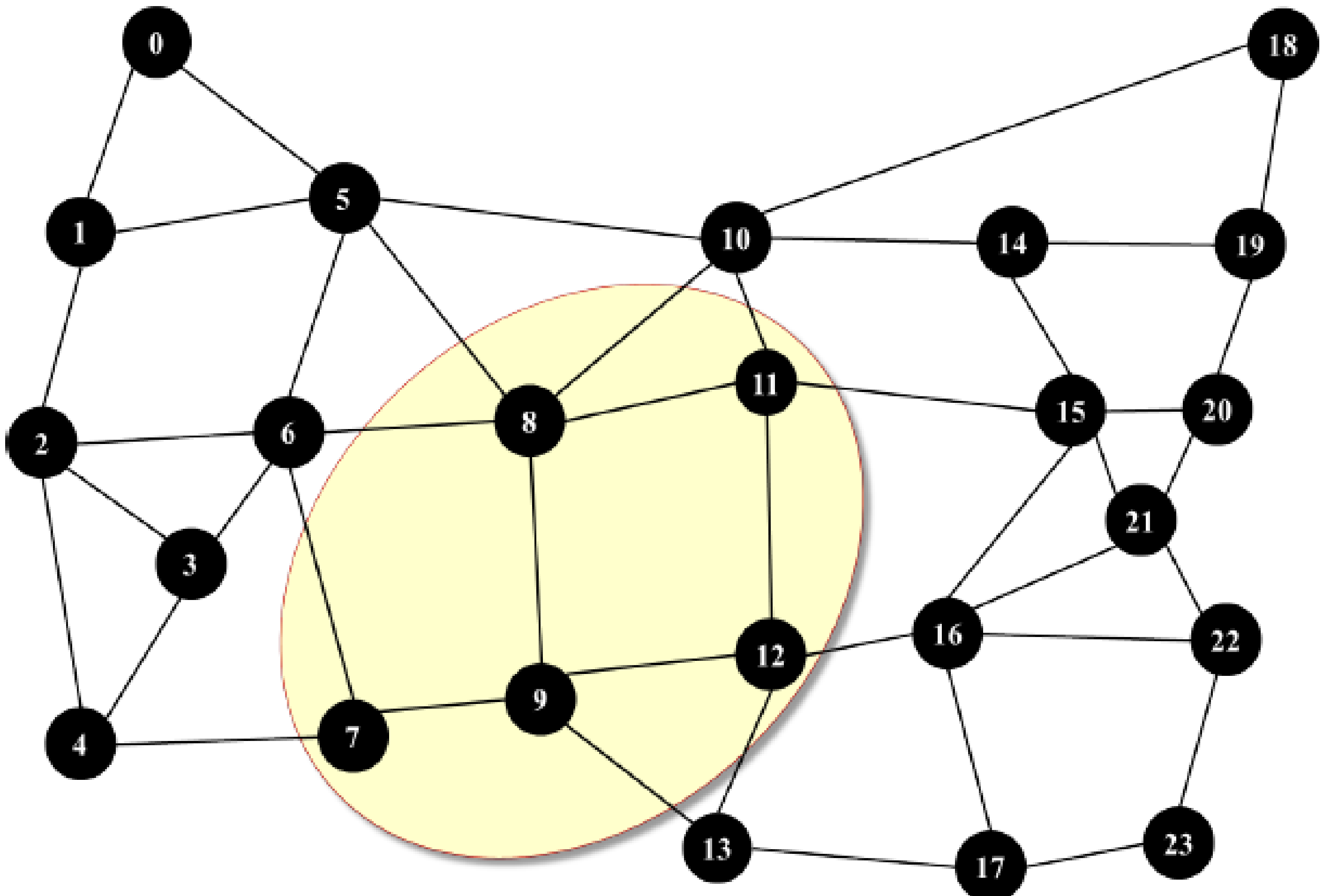
VN requests are specified as random graphs with an average of 3-5 VN nodes and 4-7 VN links each.

VN nodes request between 10-20 units of capacity and VN links request between 50-1,000 units of bandwidth (uniform).

All requests arrive in a random exponential manner and have infinite durations, chosen to reflect long-standing services.

The average amounts of node and link repair resources in each stage (X , Y) are set to 200 and 30,000 units.

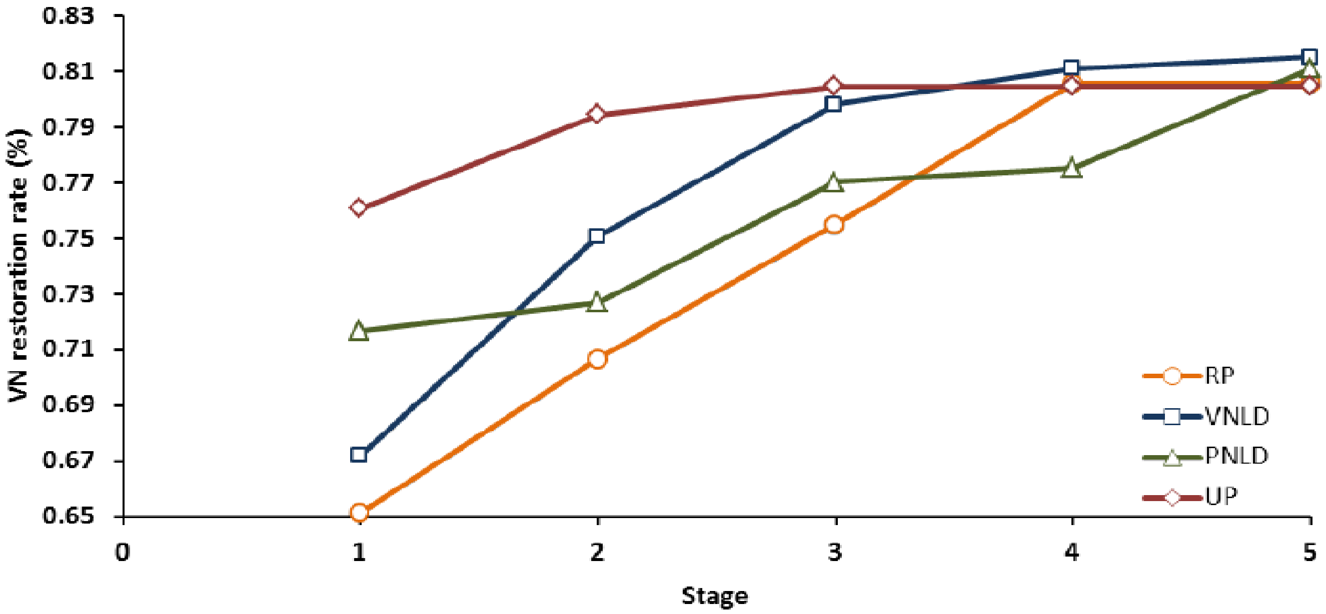
UCDAVIS
UNIVERSITY OF CALIFORNIA




Failure region

# Results

## VN restoration rate




UCDAVIS
UNIVERSITY OF CALIFORNIA

# Average VN path length