FUJITSU
shaping tomorrow with you

Survivable Resource Orchestration for Inter-Data Center Networks

Qiong Zhang, Xi Wang, Paparao Palacharla, Motoyoshi Sekiya

Fujitsu Laboratories of America, Inc.
Richardson, TX. USA

What is Cloud?

FUJITSU

- **Cloud must satisfy 5 axioms*:**
  - **C**ommon
    - Resource sharing in common pools
  - **L**ocation independent
    - At least one feasible resource allocation
  - **O**nline
    - No isolated customers, suppliers, traders, components
  - **U**tility
    - Price is proportional to quantity
  - on-**D**emand
    - Respond to meet instantaneously changed demands

- **By 2014, >50% of all workloads will be processed in the cloud**
  - From Cisco Analysis

* Joe Weinman, “Axiomatic Cloud Theory,” http://www.joeweinman.com/Resources/Joe_Weinman_Axiomatic_Cloud_Theory.pdf

Big Data

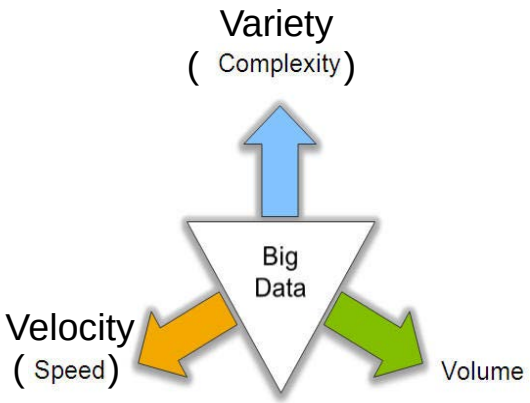
FUJITSU

- **The model of generating/consuming data has changed**

Old model: few companies are generating data, all others are consuming data

New model: all of us are generating data, and all of us are consuming data

- **Big data has 3V characteristics**
- **Stored in distributed databases**

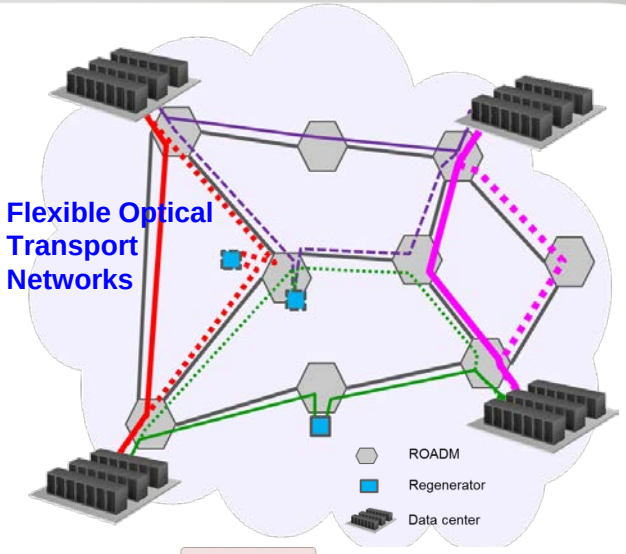


Flexible Optical Transport Networks for Inter-Datacenter Communications

FUJITSU

Optical Transport Networks for Inter-DC

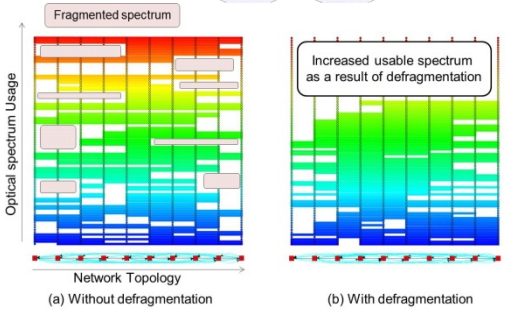
- Software defined optical transport networks to provide **flexible** and **dynamic** transport services between datacenters
- Flexible Reconfigurable Optical Add/Drop Multiplexers (ROADMs) and universal transceivers capable of 100 and 400 Gb/s transmission



Key Features

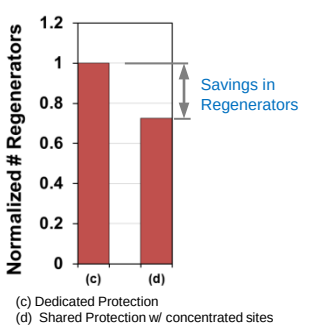
Flexible and dynamic transport

Rapid delivery of on-demand bandwidth service between DCs;
Low latency service for critical data backup and financial applications



Efficient resource sharing

Effective sharing of regenerators for dynamic traffic and restoration;
Reduce CAPEX/OPEX by concentrating regeneration sites



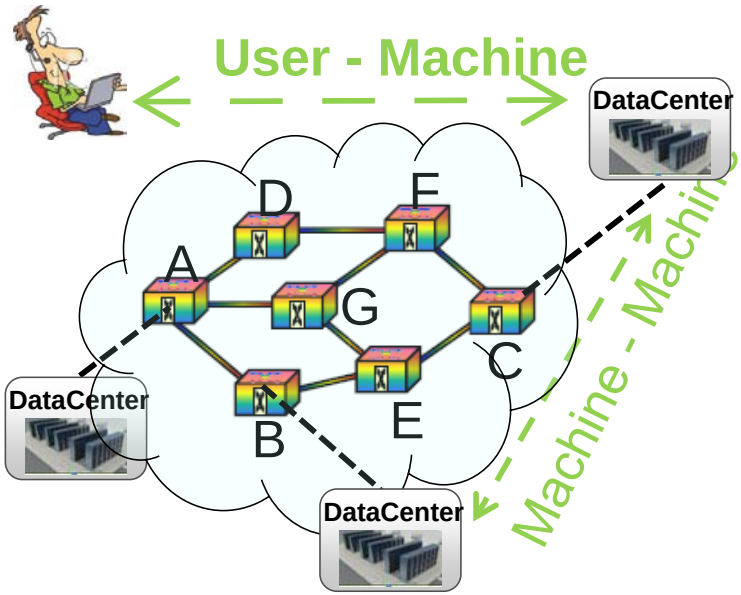
Network optimization

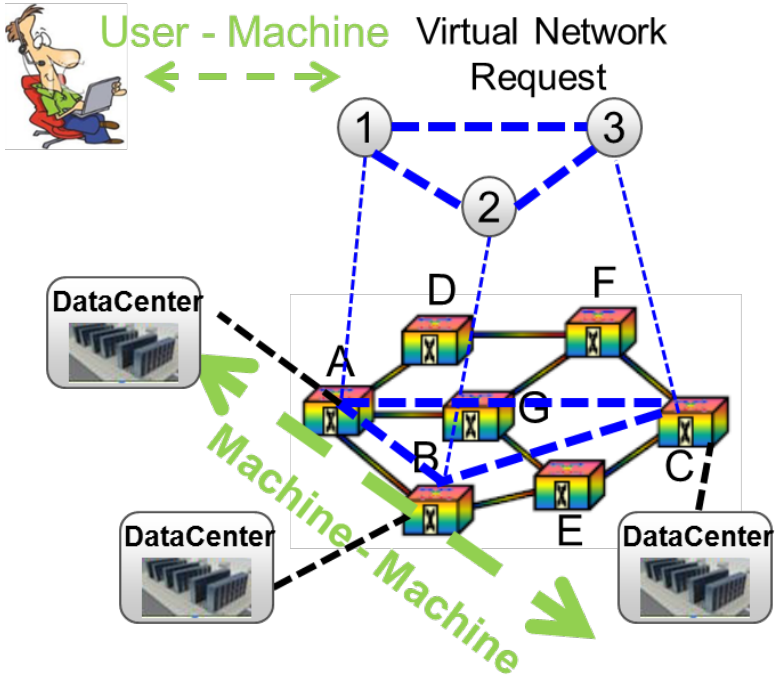
Hit-less spectrum defragmentation to enable higher network utilization

Kyosuke Sone, et al., ECOC 2012, post-deadline paper Th.3.D.1 (2012).

X. Wang et.al. “A Hitless Defragmentation Method for Self-optimizing Flexible Grid Optical Networks,” ECOC 2012.

Changes Brought by Applications





Current:

- A single request -> a single data center
- Low volume of M2M traffic
- Fixed sources and/or destinations
- Separate IT and network resources

Future:

- A single request -> distributed data centers
- Significant increase in M2M traffic
- Flexible sources and/or destinations
- Joint IT and network resources

Motivation

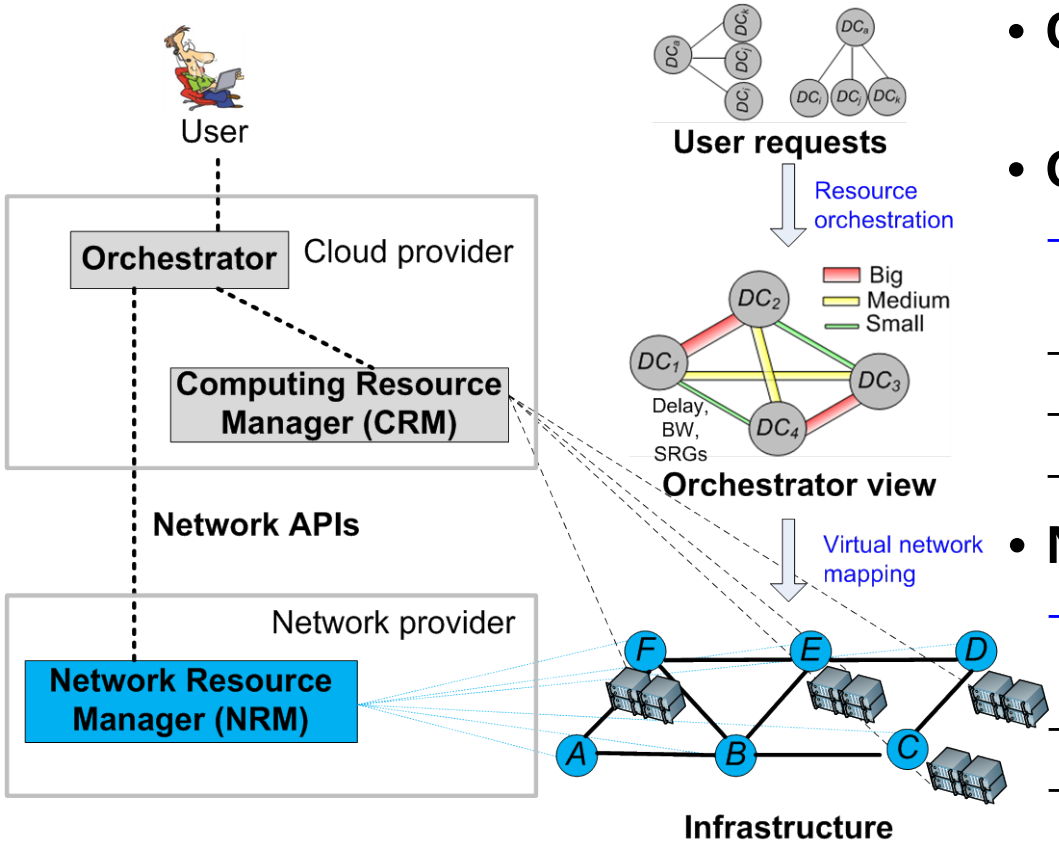
FUJITSU

- Cloud applications rely on distributed data centers (DCs) for improved user experience by service locality
- However, cloud providers may not own network infrastructure and count on network providers to optically interconnect distributed DCs
- Network providers are unwilling to expose their full network topology information to cloud providers

- Hence, it is critical to investigate an overlay framework that enables cloud providers to control cloud network connections and optimize resource orchestration without having detailed network information

- Previous research provide reliable anycast or manycast by disjoint routing on physical network topologies

Overlay Framework



- **CRM**
  - Manage DCs
- **Orchestrator**
  - Resource orchestration for user requests on overlay networks (our focus)
  - Use SDN APIs for network info & control
  - Point-to-point connections
  - Adjustable bandwidth
- **NRM**
  - Virtual network mapping on the physical infrastructure (our previous work)
  - Network control similar to a SDN controller
  - Infrastructure supports flexible optical data planes

Advantages

- No physical network topology exposed to cloud providers
- Easy to setup cloud services without considering intermediate network devices

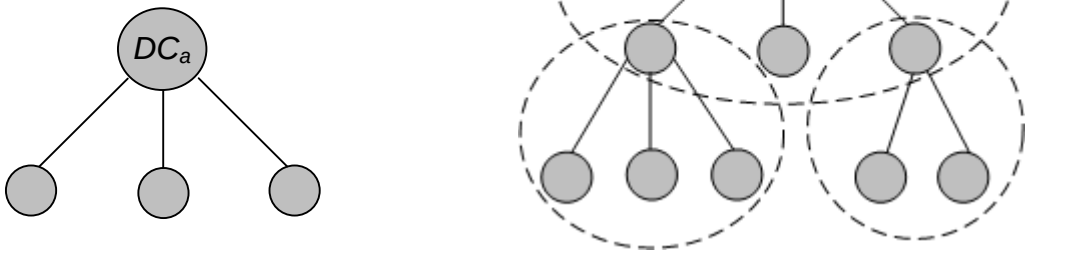
Q. Zhang et.al. "RWA for network virtualization in optical WDM networks," OFC 2013.

X. Wang et.al. "Flexible virtual network provisioning over distance-adaptive flex-grid optical networks," OFC 2014.

Aggregation Datacenter

- **Many cloud applications are arranged in an aggregation communication pattern, whereby an aggregation data center (DC_a) collects data processed at distributed DCs and outputs final results to users**

- **Complicated communication patterns can be constituted by scheduling a sequence of data aggregations**



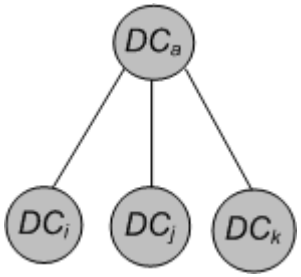
(a) Aggregation (b) A sequence of aggregations

Communication Patterns

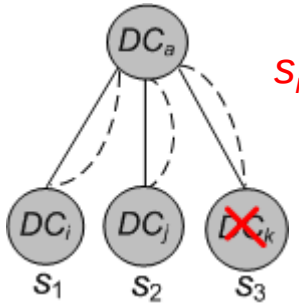
K-Connect Survivability

FUJITSU

- **K-connect survivability**
  - Defined as at least *K* DCs remain connected with DC_a for any shared risk group (SRG) failure

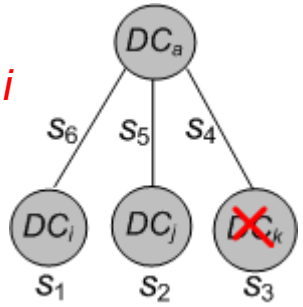


(a) Basic request



(b) Separate protection

s_i indicates SRG *i*



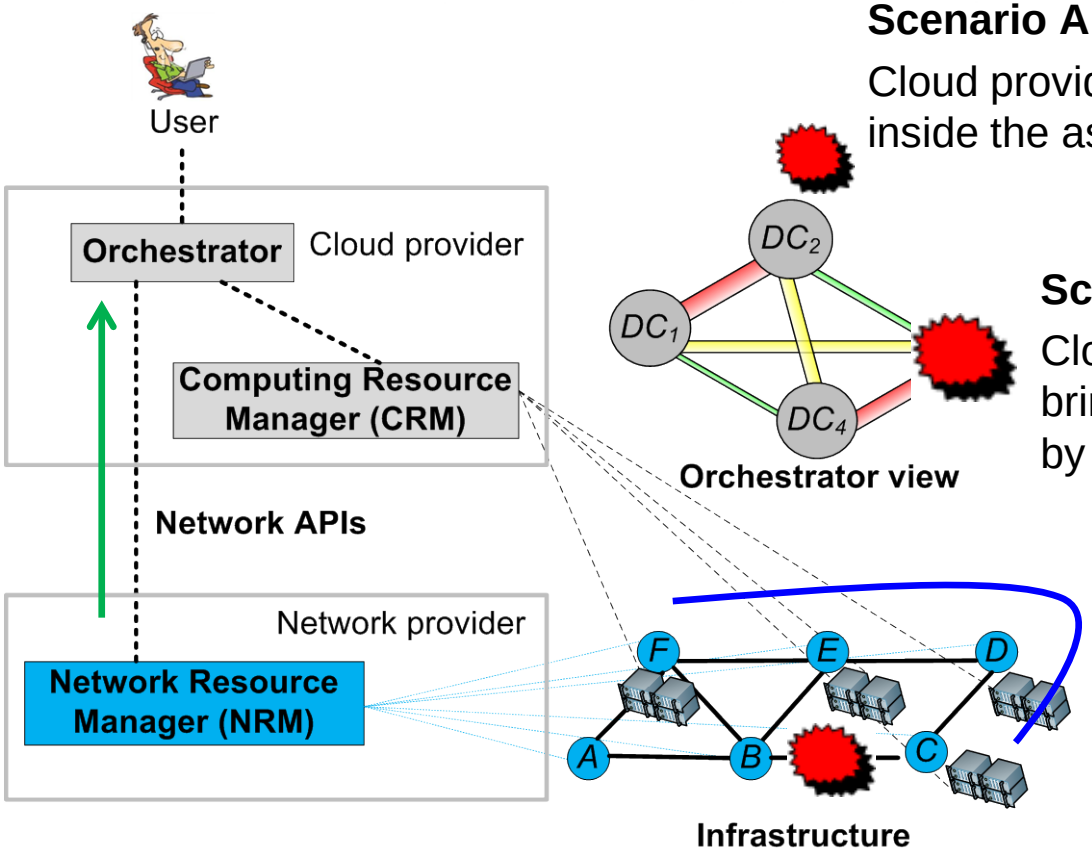
(c) Joint protection

Guarantee 2-connect survivability

  - Jointly protecting connections and DCs significantly saves network resources (a saving of three protection connections compared to separate protection)

- **Surviving *K* DCs can be allocated additional virtual machines in order to maintain computing capacity for a request when a failure occurs**

Failure Scenarios



Scenario A: VM failure

Cloud provider recovers VM failures inside the associated DCs via CRM

Scenario B: DC failure

Cloud provider recovers DC failures by bringing up protecting DCs provisioned by *K*-connect

Scenario C: Network failure

- Network provider recovers network failure (for protected connections)
- Cloud provider is notified of network failure and recovers failure by *K*-connect

- **Guarantee QoE for users by connecting VMs**
  - NRM provides network connections for minimum required DCs with Min Delay
    - Optical layer protection avoids impact to multiple connections caused by a network failure
  - CRM keeps the same total number of VMs for a user request
    - Connections are not necessary to be protected with *K*-connect survivability

How to Design for Survivability?

FUJITSU

Problem A – Finding the least DCs

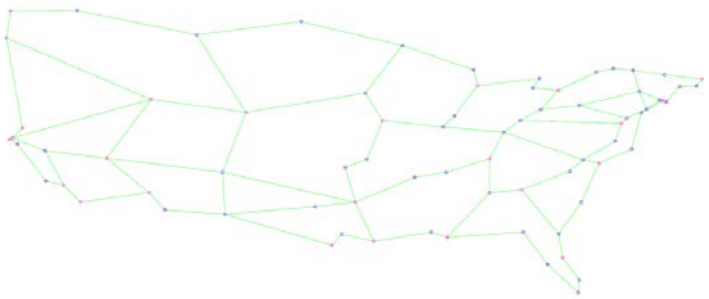
- **Find at least *M* DCs for guaranteeing *K*-connect survivability with minimum delay**
- Given: SRGs: $S = \{s_1, s_2, \ldots, s_l, \ldots, s_L\}$; Overlay network includes *N* DCs; Connections between DCs with associated risks
  - A request has a pre-defined DC_a and the requirement of *K*-connect survivability
  - Assume that DC_a can not fail; otherwise a new request should be initiated
- Proposed DelayBased and RiskBased schemes

Problem B – VM Allocation

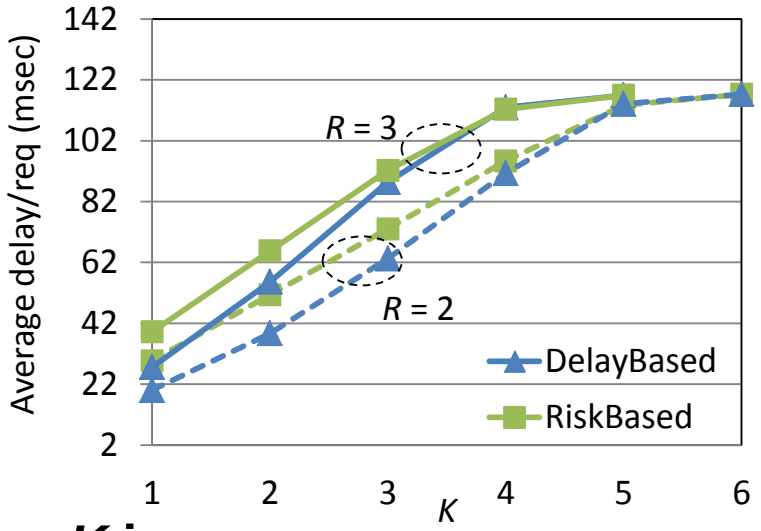
- **Allocate VMs at DCs for a set of requests in order to reduce the total number of VMs required on an overlay network**
- Given: A set of cloud requests, each represented by (DC_a, *K*, *V*)
  - *K* is the *K*-connect survivability to guarantee
  - *V* is the number of VMs that needs to maintain for any failure
- For each request, *M* DCs are selected based on the solutions for Problem A
- Proposed VM sharing within requests and between requests for reducing VMs

Simulation

- Overlay networks are generated with randomly chosen 10 DCs among the 75-node network topology
- The shortest paths are used for connections between DCs
- A request is generated by assigning DC_a to each DC at overlay networks until 10^5 requests are successfully allocated



75-node CORONET



- **The least number of DCs increases as *K* increases**
- **RiskBased requires up to 12% fewer DCs than DelayBased**
- **RiskBased results in longer delay than DelayBased, but the difference in delay reduces as *K* increases**

Q. Zhang et al., "Survivable resource orchestration for optically interconnected data center networks," Optics Express, vol. 22, no. 1, pp 23-29 (2014).

Conclusion

FUJITSU

- **An overlay framework enabled by SDN and network virtualization technologies is necessary for resource orchestration that jointly considers IT resources in DCs and network resources**

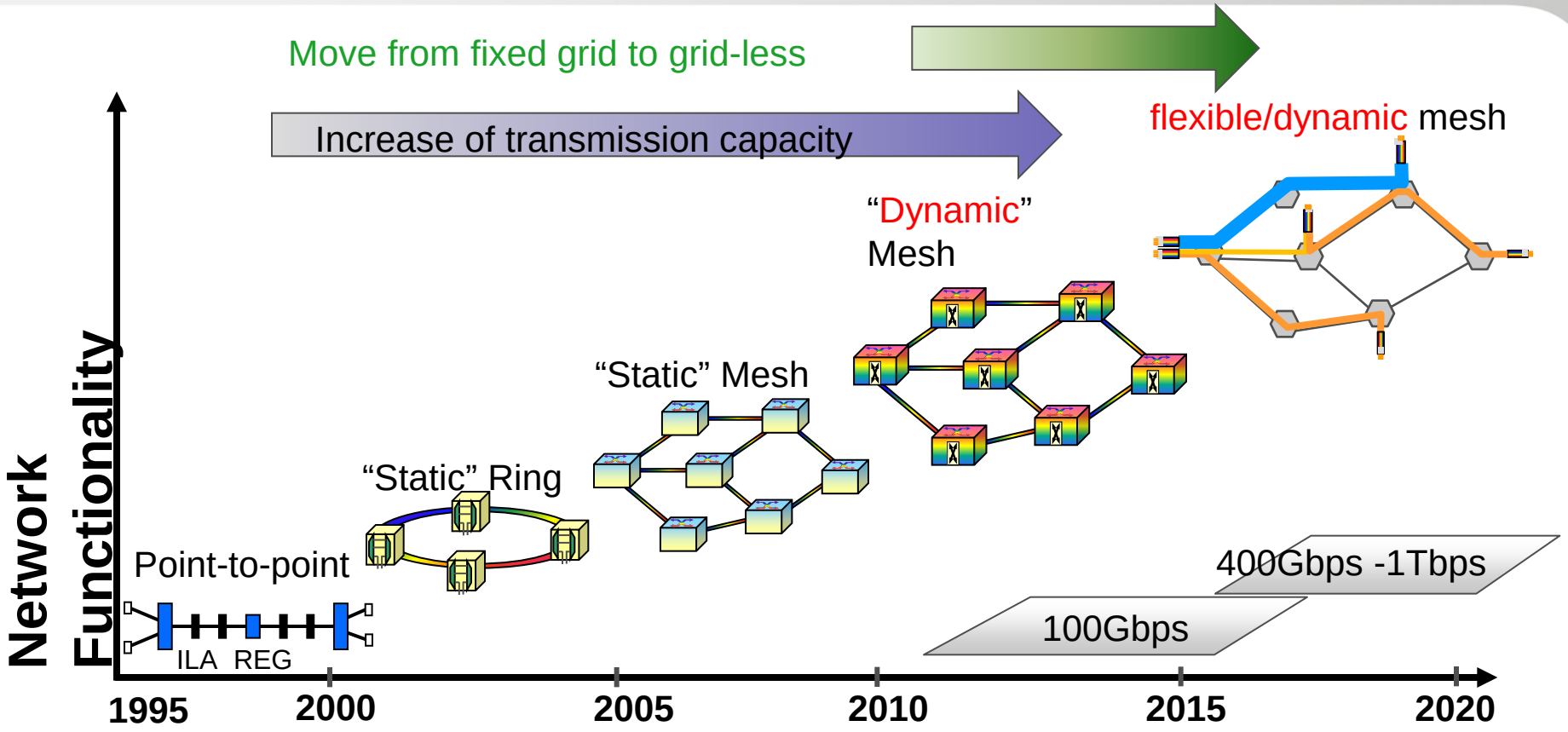
- **Presented resource orchestration schemes on SDN-managed overlay networks to provision the fewest data centers and the fewest VMs, while guaranteeing *K*-connect survivability**

Thank you!

FUJITSU

Backup slides

Evolution of Photonic Networks

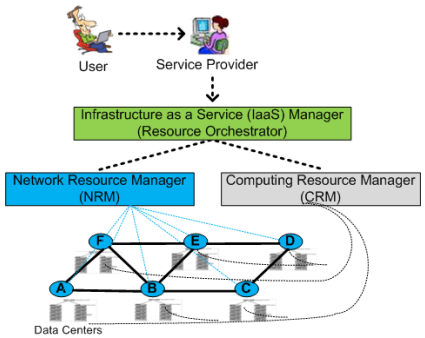


- Transmission capacity continue to increase
- Optical networks have evolved from point-to-point to ring to mesh to dynamic mesh networks
  - From channel-based to grid-less

Intelligent Network Management Utilize SDN

- Manage resources efficiently with centralized view
  - Allocate path and slots considering balance with reach, demand and resource
  - Manage resource considering fragmentation measured by utilization entropy
  - Iterative bridge and roll for connection rerouting (Qiong Zhang et.al TuQ3-2 OECC2013)
- Virtualized photonic network enabling new service
  - Virtualized optical networks, RWA algorithm and interconnected distributed DCs with reliable and VM service quality (Qiong Zhang et. Al Jth2A.65 OFC2013 , ECOC2013)

SDN enables new services in flexible grid networks



VON manage

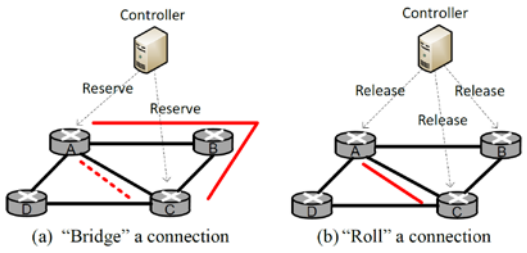


Fig. 1. Bridge and Roll a connection using SDN controller.

Bridge and roll