Virtual Network Allocation for Fault Tolerance with Bandwidth Efficiency in a Multi-Tenant Data Center

Yukio Ogawa
Hitachi, Ltd., Japan

Go Hasegawa and Masayuki Murata
Osaka University, Japan

Introduction

Research background and objectives

Modeling a multi-tenant data center network

A hypothesis on the failure recover time

Network model for a multi-tenant data center

Evaluation

Contents

● Introduction
  ▷ Research background and objectives

● Modeling a multi-tenant data center network
  ▷ A hypothesis on the failure recover time
  ▷ Network model for a multi-tenant data center
  ▷ Objective
  ▷ Recovery time model of a single virtual network

● Evaluation
  ▷ Data center network for evaluation
  ▷ Overview of a single virtual network mapping
  ▷ Trade-off between fault tolerance and physical bandwidth consumption
  ▷ Virtual network allocation policy derived from the results

● Conclusion

Research background

• A data center (DC) for the IaaS cloud computing
  – serves virtual DC for multiple client organizations, i.e. tenants
  – needs to host business-critical and mission-critical applications

• The virtual network (VN) for a tenant’s virtual DC
  – is an overlay network built by connecting VMs, based on VXLAN, etc
  – has a topology independent of the physical substrate network (SN)
  – should be appropriately assigned to the SN to share the SN's resources effectively and tolerate SN failures

• Goal: ensuring high availability for the VN so that mission critical applications can be hosted on it

Research objectives

• Mapping VNs to the shared physical SN is a kind of the Virtual Network Embedding problem

• Problems:
  in a multi-tenant data center,
  – nodes and links of VNs share a single component of the SN
  – a failure of a single SN component can cause multiple simultaneous failures in a VN
  – significantly disrupts the services offered on the VN, as compared to a traditional network

• Research objectives: clarifying how the fault tolerance of a VN is affected by a SN failure, from the perspective of VN allocation

A hypothesis on the failure recovery time in a single VN

• A hypothesis: multiple simultaneous failures can lead to a longer recovery time in physical and virtual networks

• Proposal: switching from hot- to cold-standby recovery with reference to the failure complexity

Network model for a multi-tenant data center

Map a VN onto the SN
Network model for a multi-tenant data center

Mapping a VN onto the SN

- Traffic flow is assigned to logical link.
- Logical node is mapped onto physical server.
- Logical Link is mapped onto physical path.

FW: Fire Wall
LB: Load Balancer
AP: Application
DB: Database

External network
Internal network
Physical substrate network (SN) (rack)
Mapping a VN onto the SN

- Traffic flow $f$ is assigned to logical link $l$
- Logical node $n$ is mapped onto physical server $v$
- Logical Link $l$ is mapped onto physical path $p$

Mapping a VN onto the SN

- Traffic flow $f$ is assigned to logical link $l$
- Logical node $n$ is mapped onto physical server $v$
- Logical Link $l$ is mapped onto physical path $p$

Goal of VN allocation: minimizing the bandwidth loss when a failure happens in the SN

Objective

- Objective: minimize
- Bandwidth loss of the VN
- Failure rate of physical servers
- VN's recovery time after a failure of physical server $v$
- Bandwidth of traffic flow $f$

Recovery time of a VM

- Hot-standby: 4 s, Cold-standby: 60 s

Data center network for evaluation

A single VN
- Three-tier web serving architecture
- 5.8 web and AP/DB servers, a total of 15.7 VMs on average.
- CPU cores per VM: 1
- Average bandwidth demand from an external network: $1.7 \times 10^8$ bit/s
- Recovery time of a VM
  - Hot-standby: 4 s, Cold-standby: 60 s

The SN
- Two-level fat-tree topology
- Max configuration: 8 core switches, 16 ToR switches, and 120 physical servers
- CPU cores/physical server: 32, bandwidth of each link: $1 \times 10^{18}$ bit/s
- Available CPU cores: 3,360
- Failure rates: physical server: 4/year, physical link-switch: 0.05/year (neglected)
Overview of a single VN mapping

- VN embedding problem is NP-hard:
  - Initially – Greedy Algorithm, refined – Tabu search
- VN recovery time depends on \( \theta \) (threshold for switching hot-to-cold-standby), which can not be defined in advance
  - \( \theta_s \) (a setting value of \( \theta \)) is initially chosen
- VN is allocated by using \( \theta_s \), then evaluated for various values of \( \theta \)

\[ \theta_s = 1(\text{min}) \]
- The VMs and logical links are scattered across many physical servers and links
- The VMs and links are consolidated in a few physical servers

VN Allocation Policy Derived from the Results

- Minimizing the bandwidth loss of the VN while avoiding holding too many redundant core switches
- Pareto optimality: \( \theta_s = 4 \)
  - Almost of the logical links were mapped onto the physical links between the physical servers and ToR switches.
  - The VN had almost no inter-rack traffic flows other than the one coming through the gateway

\[ \theta_s = 4 \]
(Pareto optimal)

Trade-off between fault tolerance and physical bandwidth consumption

- \( \theta_s = 1(\text{min}) \): one VM to one physical server mapping
  - The consumed bandwidth between servers/racks reaches the maximum
- \( \theta_s = 32(\text{max}) \): many VMs to one physical server mapping
  - The bandwidth loss is nearly the maximum for cold-standby recovery
  - The consumed bandwidth between servers/racks becomes the minimum

Conclusion

- The fault tolerance of each VN in an IaaS data center
  - Focusing on the situation of multiple simultaneous failures in each VN caused by a single physical failure
  - The trade-off between the bandwidth loss and the required bandwidth between physical servers
  - Balancing by assigning every four VMs to a physical server,
    - the required bandwidth of the outside racks was minimized

Future work
- Investigation of resource allocation over WANs, i.e., in a hybrid cloud environment