

클라우드 데이터센터를 위한 네트워킹 기술에 관한 연구

최정열
성결대학교 컴퓨터공학부

A Study on Networking Technology for Cloud Data Centers

Jung-Yul Choi

Division of Computer Engineering, Sungkyul University

요 약 모바일 및 사물인터넷 기술의 발전, 대용량 빅데이터 처리, 그리고 클라우드 컴퓨팅 기술의 발전에 힘입어 기존 데이터센터는 클라우드 데이터센터로 변모하고 있다. 클라우드 데이터센터는 ICT 자원들을 가상화하여 운영함으로써 에너지 및 시설 자원을 효율적으로 관리하고 사용자들의 서비스 요구에 빠르게 대응하는 것을 목표로 하고 있다. 이에 따라 클라우드 데이터센터 네트워크는 가상화된 ICT 자원을 효율적으로 제공할 수 있도록 구성되어야 한다. 본 논문은 클라우드 데이터센터에 적합한 네트워크 구조 및 네트워킹 기술을 분석하고 이를 효과적으로 운용하기 위한 방안을 제시한다.

주제어 : 가상화, 네트워크 구조, 네트워킹 기술, 클라우드 데이터센터, ICT, VXLAN

Abstract Legacy data centers are transforming toward cloud data centers according to the advance of mobile and Internet of Things technology, processing of big data, and development of cloud computing technology. The goal of cloud data centers is to efficiently manage energy and facility, and to rapidly provide service demands to users by operating virtualized ICT(Information and Communication Technology) resources. Accordingly, it requires to configure and operate networks for efficiently providing virtualized ICT resources. This paper analyzes networking technologies suitable for cloud data centers and presents ways to efficiently operate the data center.

Key Words : Virtualization, network architecture, networking technology, cloud data centers, ICT, VXLAN

1. 서론

정보기술의 지속적인 발전은 사용자가 어떠한 상황에서도 다양한 서비스를 지체없이 사용할 수 있도록 도와주고 있다. 모바일 및 사물인터넷 기술의 발전, 대용량 빅데이터 처리 및 활용, 그리고 클라우드 컴퓨팅 기술 등

이 이들을 뒷받침해주고 있다. 데이터센터는 사용자가 요구하는 서비스와 데이터를 제공하는 허브로서 중요한 역할을 담당한다. 데이터센터는 컴퓨팅 서버와 스토리지, 그리고 데이터 전송을 위한 네트워크 등과 같은 수많은 ICT(Information and Communication Technology) 장비들을 운용하며, 이러한 ICT 장비들이 24시간 정상적으로

Received 29 December 2015, Revised 30 January 2016
Accepted 20 February 2016, Published 28 February 2016
Corresponding Author: JungYul Choi(Sungkyul University)
Email: passjay@gmail.com

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN: 1738-1916

작동할 수 있도록 공조 장치 및 전력공급장치 등과 같은 인프라 시설을 운용한다[1].

기존의 데이터센터들이 웹 호스팅 서비스 및 서버 임대와 같은 물리적인 자원을 제공하는 서비스를 제공하는 데 주력했다면, 클라우드 데이터센터는 ICT 자원을 클라우드화 하여 클라우드 서비스를 제공하는데 초점을 두고 있다[2]. 특히 서버 가상화, 스토리지 가상화, 네트워크 가상화 등과 같이 가상화 기술을 기반으로 물리적인 장치의 한계를 뛰어 넘어 클라우드화 함으로써 보다 효율적으로 자원을 관리할 수 있게 되었으며, 사용자의 요구에 보다 빠르게 대응할 수 있게 되었다.

대부분의 ICT 장비들이 가상화로 구성되어 동작하는 클라우드 데이터센터에서는 서버의 멀티테넌시(multi-tenancy)¹⁾ 지원으로 인해 수많은 가상머신들을 관리해야 하며, 가상머신 간의 데이터 이동 지원과 빈번히 발생하는 자원 할당 및 재분배, 그리고 ICT 자원의 효율적인 이용을 위한 자동화 및 통합이 요구된다[3]. 이러한 특징들로 인해 클라우드 데이터센터의 네트워크에서는 스위치 간 상하(north-south)의 트래픽 유통보다는 서버 간의 좌우(west-east)로의 트래픽 유통 비중이 훨씬 더 높다[4]. 이러한 클라우드화 된 데이터센터를 운용하기 위해서는 근본적으로 차별화된 네트워크 구조와 네트워킹 기술이 요구된다.

이에 따라 본 논문에서는 클라우드 데이터센터에서 ICT 자원을 보다 효율적으로 운용 관리하기 위해서 요구되는 네트워크 구조 및 네트워킹 기술을 논한다. 이를 위해서 먼저 2장에서는 클라우드 데이터센터 위한 네트워크의 특징 및 요구사항을 기술한다. 이를 바탕으로 3장에서는 클라우드 데이터센터에 적합한 네트워크 구조를 논한다. 4장에서는 클라우드 데이터센터에 적용이 가능한 네트워킹 기술을 가상머신 간 스위칭 기술과 네트워크 가상화 기술 측면에서 분석한다. 마지막으로 5장에서는 클라우드 데이터센터 위한 효율적인 네트워크 구조 및 운용 방안을 제시한다.

2. 클라우드 데이터센터 네트워크

2.1 클라우드 데이터센터의 특징

클라우드 데이터센터는 ICT 환경이 클라우드화 함에 따라 고밀도 ICT 장비를 보유하고, ICT 자원을 보다 효율적으로 활용하기 위해서 가상화 기술에 기반하여 자원을 통합 운영하며, 사용자의 서비스 요구에 따라 유연하게 ICT 자원을 제공할 수 있다[5]. 이러한 클라우드 데이터센터는 손쉽게 확장할 수 있도록 공간, 기반 시설 및 ICT 장비의 구성을 모듈화한다[2]. 데이터센터의 에너지 효율화를 위해서 ICT 장비를 고밀도로 운영하기 때문에 높은 전력밀도²⁾를 나타낸다. 따라서 고밀도 환경을 위한 전력공급장치 및 공조장치의 구성이 필수적이다. 또한 사용자에게 원활하게 클라우드 서비스를 제공하고, 데이터센터의 에너지 효율성을 평가하고, 시설을 효율적으로 관리하기 위해서 데이터센터 내의 시설 정보는 물론 ICT 자원의 운영 정보를 통합 관리한다.

2.2 클라우드 데이터센터 네트워크의 특징

전통적인 데이터센터의 목적은 데이터의 자료 보관 외에 외부의 사용자들에게 이메일, 웹, 데이터베이스, 멀티미디어 콘텐츠 등의 데이터와 서비스를 제공하는 것이었다. 따라서 대부분의 데이터 트래픽이 데이터센터 내부의 서버에서 데이터센터 외부의 사용자에게 전달되는, 즉 서버에서 집선 스위치를 거쳐 코어 스위치를 통해서 외부로 전달되는 상하(north-south) 경로를 이용했었다[6,7]. 하지만 클라우드 데이터센터에서는 빅데이터 및 분산 서비스 처리를 위한 서버 간 트래픽과 멀티테넌시 지원으로 인한 수많은 가상머신들의 관리 및 이들 간의 데이터 이동, 그리고 가상화를 통한 서버 통합 등으로 인해 서 좌우(west-east) 경로로의 트래픽 비중이 80%에 육박하는 상황이다[4].

2.3 클라우드 데이터센터 네트워크의 요구사항

본 절은 앞서 살펴본 클라우드 데이터센터와 이를 위한 네트워크의 특징을 효율적으로 수용하기 위해서 고려해야 할 사항들을 살펴본다.

[가상화]클라우드 데이터센터에서는 ICT 자원을 효율

1) 하나의 장치나 시스템에서 여러 사용자에게 서비스를 제공하는 기술. 개별 사용자는 '테넌트(tenant)'라 부르며, 장치나 시스템, 또는 소프트웨어를 고유화하여 사용할 수 있다.

2) 단위면적당 전력 소비량

적으로 운용하기 위해서 서버와 스토리지가 가상화되어 운용된다. 대다수의 서버들은 여러 개의 가상머신들을 운영한다. 사용자가 필요로 하는 작업이 같은 네트워크에 있는 가상머신들을 통하여 처리될 수도 있지만 다른 네트워크에 있는 가상머신들 간에 통신을 통해서 처리될 수 있다. 이를 위해서 가상머신 간에 효율적인 가상 네트워크 구성 기술이 필요하다[7,8,9]. 또한 사용자의 요구에 따라 ICT 자원 제공 및 반환이 즉시 이루어져야 하는 상황을 지원하기 위해서 각 사용자에 따라 가상 네트워크를 손쉽게 구성 및 운영할 수 있어야 한다.

[서버 통합 지원]에너지를 절감하기 위해서 이용률이 낮은 다수의 서버나 스토리지를 가상화 기술을 이용하여 소수의 서버로 통합하여 운영할 수 있다. 이때 단순히 서버의 이용률만을 기준으로 서버를 통합해서는 안 되며, 서버 내의 가상머신들 간의 네트워크 구성을 고려하여 진행되어야 한다. 네트워크를 고려하지 않고 서버를 통합할 경우에 가상머신 간의 통신량 증가로 인해 네트워크에 병목현상이 발생할 수 있기 때문에 오히려 부정적인 결과를 낳을 수 있게 된다. 이를 위해서 서버 및 가상머신의 운영 정보를 네트워크 상에서 관리되어야 한다.

[확장성]ICT 자원 수요의 급증에 따라서 클라우드 데이터센터에서는 ICT 장비가 고밀도, 고집적화되고 있다[2]. 데이터센터 내의 서버 및 스토리지의 수가 증가함에 따라 이들 간에 효과적인 통신이 이루어질 수 있는 네트워크 구성이 요구된다. 이는 초기 구축시는 물론 수요 증가에 따른 점진적인 확장성도 제공해야 함을 의미한다.

[좌우 트래픽 전달]기본적으로 클라우드 데이터센터의 네트워크에서는 상하간 트래픽의 비중보다 좌우간 트래픽의 비중이 높기 때문에, 상하간 트래픽과 신뢰성에 기반을 둔 다계층 이중화로 구성된 전통적인 트리 형태로는 대응하기에 부적합하다. 따라서 좌우간 트래픽 전달에 적합한 플랫폼 구조의 네트워크가 요구된다[10].

[신뢰성]네트워크 구성의 기본적인 원칙은 장애 발생 시에도 사용자에게 통신 서비스를 제공하는 신뢰성이다. 많은 경우에 장비 및 링크 이중화 및 이원화를 통해서 신뢰성을 제공하고자 하는데 이는 대역폭 낭비 및 비용 부담을 요구한다. 또는 라우팅 프로토콜 등을 활용하여 우회 경로를 설정하여 신뢰성을 지원하고 있다. 클라우드 데이터센터 내에는 무수히 많은 가상머신 간에 통신이 이루어지고 있으므로 네트워크 및 통신에 대한 신뢰성

지원이 요구된다.

[데이터센터간 통신]대규모의 클라우드 서비스를 제공하는 글로벌 데이터센터는 확장성과 서비스 품질, 그리고 신뢰성을 고려하여 지역적으로 분산된 다수의 데이터센터를 구축하여 운영한다. 사용자의 요구에 따라 인근 지역으로부터 서비스를 제공받을 수 있으며, 특히 데이터센터의 장애가 발생할 경우에 백업 센터를 이용할 수 있다. 따라서 데이터센터 간에 효율적인 데이터 전송을 위한 네트워크의 구성 및 운영 기술이 요구된다.

3. 클라우드 데이터센터 네트워크 구조

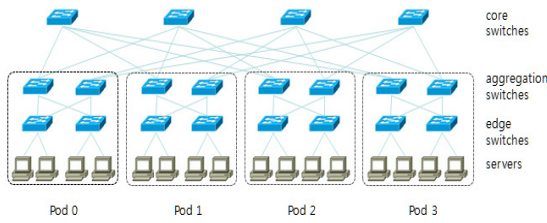
3.1 트리 기반의 네트워크

일반적인 데이터센터의 네트워크 구조는 서버들로 구성된 랙(rack) 상단에 스위치(top of rack switch, ToR switch)가 있고, 이들을 연결하는 집선 스위치(end of row switch, EoR switch), 그리고 이들을 연결하는 코어 스위치 등으로 구성된다[11]. 네트워크의 규모와 데이터센터의 구성에 따라 집선 스위치가 다단계로 구성되거나 집선 스위치가 없을 수도 있다. 서버와 ToR 스위치 간에는 주로 1Gbps 링크로 연결되며, 집선 스위치는 10Gbps 링크로 구성된다. 스위치 간에는 부하 분산과 신뢰성을 고려하여 이중 링크로 구성될 수 있다.

이러한 다계층 이중화 기반의 트리 구조의 네트워크에서는 서로 다른 ToR 스위치에 연결된 서버들이 서로 통신을 하기 위해서는 상단의 집선 스위치나 코어 스위치를 거쳐야 하기 때문에 이들 스위치나 스위치로의 링크 구간이 병목이 될 수 있다. 또한 이중화로 구성되므로 링크 대역폭을 절반 밖에는 사용하지 못하게 된다. 신뢰성을 중시하고 상-하 트래픽 비중이 높았던 기존의 데이터센터에서 일반적으로 사용하던 구조이다.

3.2 팻 트리(fat-tree) 기반의 네트워크

기본 트리 구조는 초과 가입(oversubscription)으로 인해 스위치의 상향 링크와 하향 링크의 불균형이 발생하게 되고, 결국 코어 스위치에서 병목현상이 발생할 수 있다. 스위치의 상향 링크와 하향 링크를 동일하게 구성하면 이를 해결할 수 있다[12]. n 개의 포트를 갖는 에지 스위치는 $n/2$ 개의 서버(하향)와 $n/2$ 개의 집선 스위치



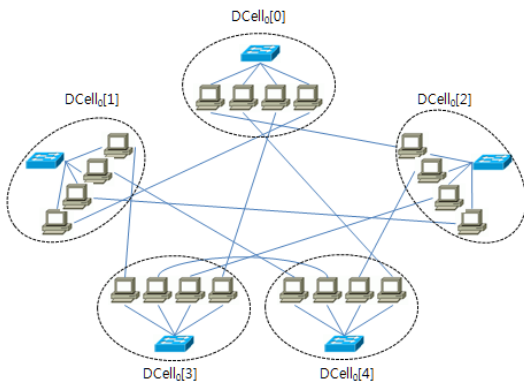
[Fig. 1] A fat-tree network with four ports switches

(상향)으로 연결된다. $n/2$ 개의 집선 스위치, $n/2$ 개의 에지 스위치, 그리고 에지 스위치에 연결된 서버가 하나의 팻(pot)를 형성한다. n 개의 포트를 가진 $(n/2)^2$ 개의 코어 스위치 각각은 모든 팻에 연결된다. 결국, n 개의 포트를 가진 스위치로 구성된 팻 트리 구조는 총 $n^3/4$ 개의 서버를 연결할 수 있다. [Fig. 1]은 4개($=n$)의 포트를 갖는 스위치로 구성된 팻 트리 구조의 데이터센터 네트워크를 보인다.

3.3 순환형 DCell 구조의 네트워크

데이터센터 네트워크가 갖추어야 할 근본적인 요소는 지속적으로 증가하는 서버들을 얼마나 효과적으로 상호 연결할 수 있는가이다. 하나의 서버와 다수의 스위치들이 작은 그룹을 형성하고, 이들 그룹이 보다 큰 그룹의 요소로서 완전 연결성을 제공한다면 이러한 문제의 해법이 될 수 있다[13]. [Fig. 2]는 4개($=n$)의 서버와 하나의 스위치로 구성된 DCell₀ 5개($=n+1$)로 이루어진 DCell₁ 네트워크를 보인다.

DCell₁ 내의 총 서버의 수는 DCell₀의 수와 DCell₀ 내



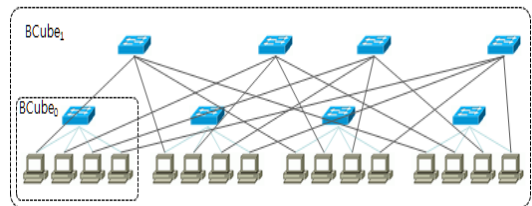
[Fig. 2] A DCell1 network with five DCell0s

의 서버 수의 곱으로 산출되어, 총 20개가 된다. 이를 일반화하면, t_{k-1} 개의 서버들로 구성된 DCell_{k-1}로 이루어진 k 단계의 DCell_k 내에 있는 서버의 수는 $t_{k-1} \times (t_{k-1} + 1)$ 개가 된다. 이때 각각의 서버는 총 $k+1$ 개의 링크(또는 포트)로 다른 서버($=k$)와 스위치에 연결된다. 예를 들어 5개의 다른 서버들로 구성된 4단계 DCell의 경우 약 86만대의 서버들로 구성될 수 있어 대규모의 데이터센터를 구성이 가능하다.

3.4 순환형 Bcube 구조의 네트워크

대용량의 데이터를 전세계적으로 신속하게 처리하기 위해서는 높은 수준의 이동성을 갖는 모듈형 데이터센터가 필요하다. 이를 위해서 선박 컨테이너를 이용한 모듈러 데이터센터(modular data center)가 제안되었다. 모듈러 데이터센터는 밀집된 공간에 많은 서버들을 배치해야 하므로 최소의 공간에 최대한 많은 서버-스위치를 연결해야 한다. BCube라 불리는 구조에서는 스위치가 서버와만 연결되고 다른 스위치와는 연결되지 않으며, 서버가 더미 크로스바처럼 동작할 수 있어 서버가 서로에게 트래픽을 전달하게 된다[14]. 이러한 구성은 아주 저렴하고 간단한 스위치를 사용하고 라우팅 기능을 서버에 전가할 수 있다. BCube는 DCell처럼 다단계로 구성될 수 있다. [Fig. 3]은 0단계 BCube 4개로 이루어진 1단계 BCube 네트워크를 보인다.

기본적으로 k 단계($k \geq 1$)의 BCube_k는 n 개의 BCube_{k-1}와 n^k 개의 n 포트 스위치로 구성된다. BCube_k 내의 각 서버는 $k+1$ 개의 포트를 갖는다. BCube₀은 4개의 서버와 1개의 스위치로 구성되며, 이들 각각의 서버는 1단계 BCube의 스위치와 연결된다. 이 BCube₁을 바탕으로 BCube₂를 구성할 경우, 총 서버는 64대

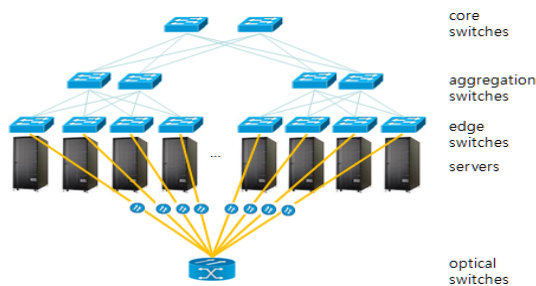


[Fig. 3] A 1-level BCube network with four 0-level BCubes

($= N = n^{k+1}$), 스위치는 48 대($= \frac{N}{n} \log_n N$)가 필요하다. 만약 8 포트 스위치를 이용해서 BCube₃을 구성할 경우 총 4096 대의 서버를 구성할 수 있으므로 컨테이너형 클라우드 데이터센터 네트워크를 구성하기에 충분하다.

3.5 광-전 혼용 구조의 네트워크

서버간의 높은 데이터 전송률을 충족시키기 위해서는 기존의 네트워크로 불충분할 수 있다. 이를 해결하기 위해서 ToR 스위치 간의 연결성을 제공한 트리 구조의 전기적인 네트워크에 재구성이 가능한 광 스위치 기반의 네트워크를 추가로 구성할 수 있다[15]. 버스티한 특성의 데이터 트래픽은 패킷 기반의 전기적인 네트워크를 통해서 전달되고, 대량의 트래픽 전송이 요구되는 응용은 회선 기반의 광 네트워크로 전송하면 보다 효과적으로 데이터를 처리할 수 있다. 광 스위치로 구성된 광 링크는 높은 비용과 넓은 대역폭을 제공하므로 모든 서버 랙 간에 완전 연결성을 제공할 필요가 없다. 사전에 서버 랙 간에 트래픽 정보를 알고 있다면 한 번에 한 랙 쌍 사이에 광 연결을 제공함으로써 보다 비용을 줄이면서 효과적으로 트래픽을 전송할 수 있다. [Fig. 4]는 광-전 혼용 구조의 네트워크를 보인다.



[Fig. 4] A hybrid network with optical and electric switches

4. 클라우드 데이터센터 네트워킹 기술

본 장에서는 클라우드 데이터센터를 위한 네트워킹 기술을 크게 가상머신 간 스위칭 기술과 네트워크 가상화 측면에서 살펴본다.

4.1 가상머신 간 스위칭 기술

한 서버가 다른 서버와 통신을 하기 위해서는 서버의 주소를 식별하기 위한 2계층 네트워크 인터페이스 카드(network interface card, NIC)가 필요하다. 마찬가지로 서버 내의 가상머신(virtual machine, VM)이 다른 가상머신과 통신을 하기 위해서는 별도의 가상 네트워크 인터페이스 카드(virtual NIC, vNIC)가 필요하다. 그런데 하나의 서버에 장착할 수 있는 NIC 수가 제한적이기 때문에 멀티테넌시 환경에서의 효율적인 vNIC의 구성 및 운영이 요구된다.

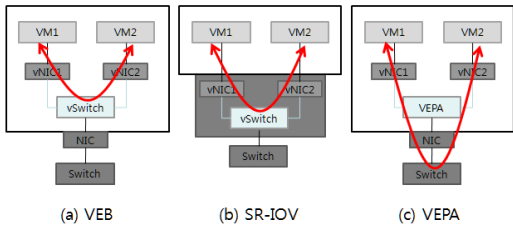
4.1.1 가상 에지 브리지

서버 가상화 환경을 제공하는 하이퍼바이저가 소프트웨어적으로 vNIC을 제공할 수 있다. 서버 내의 VM 당 vNIC이 존재하며 이들이 가상 스위치(virtual switch, vSwitch)에 연결되고, 이들은 NIC³⁾을 통해서 스위치에 연결된다. 하나의 서버 내의 VM 간에는 vSwitch에 연결된 두 vNIC을 통해서 데이터가 전달되며, 다른 서버 내의 VM신과의 통신은 NIC과 스위치를 거쳐서 이루어진다. 이러한 방식을 가상 에지 브리지(Virtual Edge Bridge, VEB)라고 하는데 투명하고 직관적이나 소프트웨어 오버헤드가 커서 확장성 및 성능에 제약이 있다. [Fig. 5(a)]는 가상 이더넷 브릿지에서 VM 간 통신이 vSwitch를 거쳐서 이루어지는 것을 보인다.

4.1.2 입출력 가상화

프로세서의 처리 부담을 덜어주고 별도의 NIC이나 스위치를 통하지 않고 다른 서버에 있는 VM과의 통신을 위해서 PCIe 상에서 단일 루트 입출력 가상화(single-root I/O virtualization, SR-IOV)를 사용하여 vNIC 포트를 구현할 수 있다[16]. [Fig. 5(b)]에서 보듯이 vNIC와 vSwitch가 NIC상에서 구현이 되며, VM과 vNIC 간에는 PCIe 버스를 통해서 연결된다. VM 간 통신을 하이퍼바이저가 제공하는 vSwitch가 아닌 SR-IOV가 적용된 NIC가 처리한다. 따라서 VM 트래픽을 감시하는데 어려움이 있으며 QoS도 제공하기 어렵다.

3) '가상'이라 칭하지 않으면 실제적인 물리 장치를 의미함



[Fig. 5] Switching technologies among VMs

4.1.3 에지 가상 브리징

프로세서의 처리 부담을 덜고 VM 트래픽 감시 및 QoS를 제공하기 위해서는 외부의 스위치를 거치는 것이 효율적일 수 있다. 즉, 두 서버 내의 VM 간의 통신이라고 하더라도 반드시 외부의 스위치를 거쳐서 내부로 되돌아오기 때문에 VM 간 트래픽 감시 및 포트별 정책 및 QoS를 제공할 수 있다. 이 방식은 서버 내의 VM 간 통신도 스위치를 통해서 이루어지므로 VEB에 비해서 성능이 떨어질 수는 있으나 VEB에 비해서 관리 측면의 장점을 갖는다. 한편, 스위치 입장에서는 한 포트에 들어온 트래픽이 다시 그 포트에 전달되어야 하는데 MAC 주소 플러딩이나 스페닝 트리 프로토콜에서 이를 제공하지 않으므로, 이를 구현하기 위한 기술로 가상 이더넷 포트 집계(Virtual Ethernet Port Aggregator, VEPA, IEEE 802.1Qbg)[17]과 VN-Tag(IEEE 802.1BR)[18]이 제안되었다. [Fig. 5(c)]는 에지 가상 브리징에서 VM 간 통신이 VEPA를 거쳐 외부의 스위치를 통해서 이루어지는 것을 보인다.

4.2 네트워크 가상화 기술

네트워크 가상화는 물리적인 네트워크 자원을 다수의 가상 네트워크로 나누어 구성할 수 있도록 지원한다[19]. 테넌트들이 다수의 VM에 할당되고 VM들이 여러 네트워크에 연결된 멀티테넌시 환경의 클라우드 데이터센터에서 네트워크 가상화는 필수적인 요소라고 할 수 있다.

4.2.1 네트워크 가상화의 요구사항

IT 자원이 가상화되어 운영 및 관리되고 있는 클라우드 데이터센터는 네트워크 운영에 있어서도 새로운 요구사항이 제시된다. 4.1절에서도 살펴본 바와 같이, 서버 내에는 다수의 VM이 존재하게 되는데 VM 간에 가상 네트워크를 구성하기 위해서는 VM별로 MAC 주소를 할당해

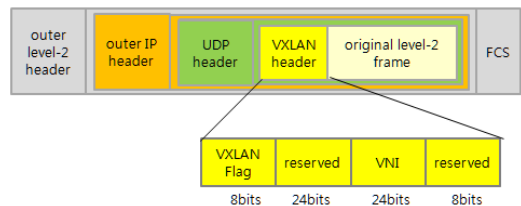
야 한다. 예를 들어 서버당 50개의 VM이 존재하고, 단일 L2 도메인 내에 서버가 4000대가 있다고 하면 하나의 스위치에서 관리해야 하는 MAC 주소가 200,000개가 되므로 이들을 구성 및 관리하는데 상당한 어려움을 겪게 된다. 또한 멀티테넌시 환경에서는 각 테넌시에게 별도의 VLAN을 할당해야 하는데, 기존의 Q-in-Q(IEEE 802.1ad, Provider Bridging)과 같은 기술의 경우에는 최대 4096개의 ID를 제공하기 때문에 클라우드 데이터센터에서는 적용이 불가능하다[20]. 또한 새로운 VM의 생성 및 소멸과 VM들이 네트워크를 통해서 빈번히 이동하는 상황을 적절히 지원할 수 있어야 한다.

4.2.2 VXLAN

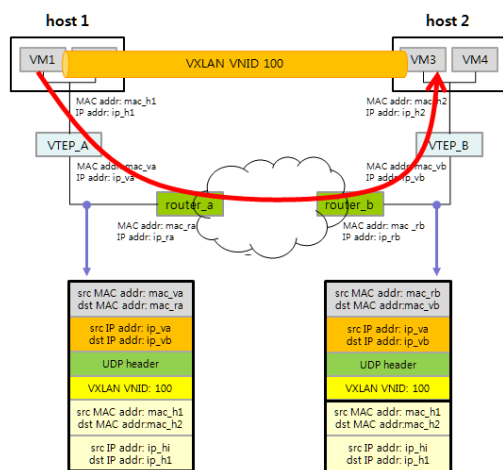
VXLAN(Virtual eXtensible LAN)은 대규모 클라우드 데이터센터 네트워크를 구성함에 있어서 발생할 수 있는 확장성 제약을 해결하기 위해서 개발되었다[21].

[Fig. 6]은 VXLAN의 프레임 포맷으로서, 원 2계층 프레임이 UDP로 캡슐화된 것을 볼 수 있다. VXLAN 헤더에 있는 VXLAN 네트워크 ID(VID)는 24비트로서, 이를 이용하여 최대 16백만 개의 유일한 가상 네트워크를 식별할 수 있다. VXLAN 프레임의 캡슐화는 VXLAN 터널 종점(VXLAN Tunnel End Point, VTEP)에서 이루어진다. 외부 IP 주소는 VTEP 사이의 데이터센터의 3계층 네트워크를 통해서 데이터를 전달하는데 사용된다. VTEP는 하이퍼바이저 내부에 있거나, NIC 또는 스위치일 수 있다.

[Fig. 7]은 VXLAN의 동작 과정을 보여준다. 호스트 A가 호스트 B로 프레임을 전송하기 위해서 VTEP1은 이 프레임을 UDP 패키지로 캡슐화를 시킨다. VTEP 간에 터널링을 통한 데이터 전송을 위해서 VXLAN 프레임 헤더에 VID이 할당된다. IP 네트워크 사이에서는 외부 IP 패키지의 목적지 주소에 따라서 패키지가 전달된다. 만약



[Fig. 6] VXLAN frame format



[Fig. 7] VXLAN operation

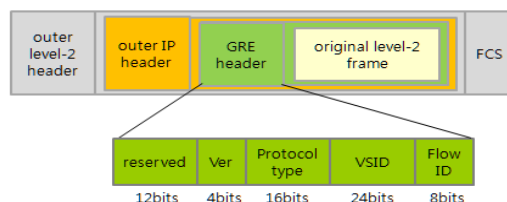
VXLAN 프레임에 있는 목적지 호스트의 MAC 주소가 알려지지 않았다면 IP 멀티캐스트를 이용해서 MAC 주소를 획득한다. 해당 VID에 연관된 모든 VTEP는 IP 멀티캐스트 그룹에 가입하게 되고, 그룹 내에서 목적지 호스트의 MAC 주소를 찾기 위해서 ARP 요구메시지를 플러딩한다. 목적지 호스트가 송신한 ARP 응답메시지를 수신한 VTEP은 자신의 ARP 테이블에 이 MAC 주소를 추가한다.

4.2.3 NVGRE

NVGRE(Network Virtualization using Generic Routing Encapsulation)도 VXLAN과 유사하게 대규모 클라우드 컴퓨팅 환경에서 확장성 문제를 해결하기 위해서 개발되었다[22]. NVGRE는 3계층 네트워크 상에서 2계층 프레임 캡슐화로 전달하기 위해서 GRE를 사용한다. GRE는 IP 망에서 가상 점대점 링크를 생성하여 다양한 네트워크 계층 프로토콜의 캡슐화를 제공한다.

[Fig. 8]과 같이 NVGRE의 헤더 포맷은 VXLAN과 유사하나, UDP가 아닌 GRE 헤더를 사용하여 캡슐화를 제공한다. GRE 헤더는 24비트 VSID(Virtual Subnet Identifier)를 이용하여 약 16백만 개의 테넌트를 식별한다. NVGRE는 내부 프레임에 VLAN 태그를 포함하지 않는다.

VM에서 생성된 프레임은 하이퍼바이저 내의 NVGRE 종점(endpoint)에서 NVGRE 프레임으로 캡슐화된다. 한



[Fig. 8] NVGRE frame format

서버에 있는 VM1이 다른 네트워크 상에 있는 VM3로 프레임을 전송하려고 할 때, VM1은 VM3가 지역 네트워크에 있는 것으로 생각하고 프레임을 전송하면 NVGRE 종점이 NVGRE 프레임으로 캡슐화하여 이들 간에 가상 터널을 생성하여 전송한다. NVGRE는 VLAN을 지원하지 않으므로 NVGRE 종점은 내부 프레임 안에 있는 VLAN 태그 정보와 VM이 속해있는 가상 네트워크 정보를 이용하여 VSID를 생성한다. NVGRE 종점은 VSID와 목적지 MAC 주소를 이용하여 VM3가 속해 있는 NVGRE 종점의 IP 주소로 전송한다. NVGRE 프레임을 수신한 NVE는 VSID와 목적지 MAC 주소를 이용하여 목적지 VM으로 이 프레임을 최종적으로 전송한다.

5. 효율적인 네트워크 구성 및 운영 방안

본 절은 앞서 살펴본 네트워킹 기술을 바탕으로 클라우드 데이터센터에 적용하기에 적합한 방안들을 제시한다. 먼저 클라우드 데이터센터 네트워크에 요구되는 사항에 따라 네트워크 구조를 분석한 결과는 <Table 1>과 같다. 일반적으로 널리 적용되는 구조는 안정성 및 신뢰

<Table 1> Analysis on cloud data center networks

Requirements	Tree	Fat-tree	DCCell	BCube	Hybrid
Virtualization support	moderate	moderate	good	good	bad
Scalability	bad	bad	good	good	moderate
West-east traffic transmission	bad	bad	moderate	moderate	moderate
Reliability	good	good	moderate	moderate	moderate
Application	general cases	with identical and economical switches	big-size data centers	modular data centers (container type)	constant but many traffics among VMs

성을 고려한 트리 구조이나, 좌우트래픽 전달에 불리하다는 단점이 있다. DCCell은 확장성이 높은 장점이 있지만 실제 데이터센터 구축에 적용된 예는 없는 것으로 파악된다. 컨테이너형과 같은 협소한 공간에 밀집된 구조에는 BCube과 같은 구조를 적용할 수 있다. 광-전 혼용 구조는 서버 간에 안정적인 트래픽이 유통되는 상황에는 유리하나, 서버 간에 트래픽이 빈번이 이동하는 클라우드 데이터센터에서는 적합하지 않다.

클라우드 데이터센터는 공공 및 사설 클라우드 컴퓨팅 서비스를 제공하기 위한 목적이 크므로 코로케이션이나 시설 임대형이 아닌 단독형으로 구축하는 경우가 많다. 물리적인 시설 한계를 벗어나 서비스를 제공하기 때문에 높은 신뢰성이 요구되므로 트리 기반의 네트워크가 적합하다. 사용자 요구에 따라 빈번히 발생하는 가상 머신의 생성 및 이주, 그리고 에너지 절감 및 이용률 향상 등을 고려한 서버 통합을 보다 효과적으로 지원하기 위해서는 좌우 트래픽 전달에 효율적인 플랫폼 2계층 트리 구조를 갖는 것이 바람직하다. 플랫폼 2계층 구조의 경우에는 스페닝 트리 프로토콜 적용에 따른 대역폭 낭비가 발생할 수 있으므로 TRILL 등과 같은 별도의 프로토콜을 적용해야 한다[23].

한편, 클라우드 데이터센터에 적용할 수 있는 네트워크 가상화 기술은 4.2절에 기술한 바와 같으며, 이들을 비교한 결과는 <Table 2>와 같다. VXLAN과 NVGRE 모두 수많은 VM들 간의 통신 및 네트워크 가상화 지원이라는 동일한 목표를 달성하기 위해서 개발되었다. 터널링 기술을 제공하고 16백만 개의 VLAN을 지원하는 것은 동일하나 패킷 캡슐화와 부하 분산 기술 등에서 차이점을 보인다. 통신 프로토콜은 기술 자체도 중요하지만 표준화 및 지지하는 벤더에 따라 네트워크에 적용여부가 결정되기도 한다. 표준화 완성도와 지지하는 측면을 본

<Table 2> Analysis on network virtualization technology

Features	VXLAN	NVGRE
Transmission	Tunneling	Tunneling
VLAN support	around 16M	around 16M
Packet encapsulation	IP/UDP	GRE
Load balancing	IP hashing-based	difficult
Supporters	Cisco, VMWare, Citrix, Redhat, etc.	MS, Intel, Dell, HP, Broadcom, etc.

다면 VXLAN이 약간 우세를 보이고 있다. 두 기술 모두 클라우드 데이터센터의 네트워크 가상화를 위해서 사용될 수 있으며, 기존 운용 장비와의 호환성 및 운용 기술의 보유에 따라서 적용하는 것이 바람직하다.

6. 결론

클라우드 데이터센터 네트워크는 ICT 자원들이 가상화되어 운영되며, 가상머신 간의 빈번한 통신으로 인해 좌-우간 트래픽의 비중이 높은 특징을 갖는다. 사용자에게 신뢰성있는 서비스를 제공하고 ICT 자원을 효율적으로 운용하기 위해서 플랫폼 2계층 트리 구조의 네트워크가 바람직하다. 스페닝 트리 프로토콜 적용에 따른 대역폭 낭비를 위한 기술이 요구되며, 효율적인 VLAN 구성을 위해서 VXLAN이나 NVGRE 등과 같은 네트워크 가상화 기술을 적용해야 한다. 본 논문에서는 다루고 있지 않지만, 신뢰성과 확장성을 위해서 여러 데이터센터가 지역적으로 분산되어 운영되고 있는데, 이들 간에 데이터를 효율적으로 전송하기 위한 네트워킹 기술에 대한 연구가 요구된다[7].

REFERENCES

- [1] D. Alger, "Grow a Greener Data Center", Cisco Press, 2010.
- [2] TTA.KO-09.0093, "Design guideline for cloud data centers", TTA standards, 2013.
- [3] H. Kim, "Network virtualization-based cloud data centers", presented in Cloud Data Center Application Networking Optimization 2011, 2011.
- [4] Cisco Global Cloud Index: Forecast and Methodology, 2011-2016.
- [5] TTA.KO-10.0762, "Evaluation framework for energy efficiency of cloud data centers", TTA standards, 2014.
- [6] Al. Greenberg, et al., "The cost of a Cloud: Research Problems in Data Center Networks", ACM SIGCOMM Computer Communication Review, Vol. 39, No. 1, pp. 68-73, Jan. 2009.

- [7] M. F. Bari, et al. "Data Center Network Virtualization: A Survey", IEEE Communications Surveys & Tutorials, Vol. 15, No. 2, 2nd Quarter 2013, pp. 909-928, 2013.
- [8] R. Jain and S. Paul, "Network Virtualization and Software Defined Networking for Cloud Computing: A Survey", IEEE Communications Magazine, Nov. 2013, pp. 24-31, 2013.
- [9] G. Lee, "Cloud Networking", Morgan Kaufmann, 2014.
- [10] J. Duff and etc, "Zetta network guideline for cloud data centers", IDG Tech Focus.
- [11] Cisco: Cisco data center infrastructure 2.5 design guide,
http://www.cisco.com/application/pdf/en/us/guest/netsol/ns107/c649/ccmigration_09186a008073377d.pdf
- [12] M. Al-Fares, A. Loukissas and A. Vahdat, "A scalable commodity data center network architecture", in Proc. of the ACM SIGCOMM 2008, pp. 63-74, 2008
- [13] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang and S. Lu, "DCCell: a scalable and fault-tolerant network structure for data centers", ACM SIGCOMM Computer communication Review, Vol. 38, No. 4, pp. 75-86, 2008
- [14] C. Guo, et al, "BCube: a high performance, server-centric network architecture for modular data centers", Proceedings of the ACM SIGCOMM 2009 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 63-74, 2009.
- [15] G. Wang, et al., "c-Through: part-time optics in data centers", Proceedings of the ACM SIGCOMM 2010 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 327-338, 2010.
- [16] "PCI-SIG SR-IOV Primer", Intel white paper, <http://www.intel.com/content/dam/doc/application-note/pci-sig-sr-iov-primer-sr-iov-technology-paper.pdf>
- [17] IEEE Std. 802.1Qbg-2012, "IEEE Standard for Local and Metropolitan Area Networks – Media Access Control(MAC) Bridges and Virtual Bridged Local Area Networks – Amendment 21: Edge Virtual Bridging," July 5, 2012.
- [18] IEEE Std. 802.1BR-2012, "IEEE Standard for Local and Metropolitan Area Networks - Virtual Bridged Local Networks - Bridge Port Extension", 2012.
- [19] M. Lasserre, et al., "Framework for Data Center (DC) Network Virtualization", IETF 7365, Oct. 2014.
- [20] IEEE Std. 802.1ad-2005, " IEEE Standard for Local and Metropolitan Area Networks - Virtual Bridged Local Area Network - Amendment 4: Provider Bridges".
- [21] M. Mahalingam, et al., "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", IETF RFC 7348, Aug. 2014.
- [22] M. Sridharan, et al., "NVGRE: Network Virtualization using Generic Routing Encapsulation", IETF RFC 7637, Sep. 2015.
- [23] R. Perlman,, "Introduction to TRILL", Internet Protocol Journal, Sep. 2011, pp.2-20.
- [24] Seong-Hoon Lee, "Actual Cases and Analysis of IT Convergence for Green IT", Journal of the Korea Convergence Society, Vol. 6, No. 6, pp. 147-152, 2015.
- [25] Seong-Hoon Lee, Dong-Woo Lee, "A Case Study in Japanese and Prospect of Cloud Computing Service in Convergence Age", Journal of the Korea Convergence Society, Vol. 6, No. 1, pp. 17-22, 2015.

최 정 열(Choi, Jung Yul)



- 2000년 2월 : 인하대학교 전자공학
과(공학사)
- 2006년 8월 : 한국정보통신대학교
공학부(공학박사)
- 2006년 9월 ~ 2011년 2월 : (주)
KT 선임연구원
- 2011년 3월 ~ 현재 : 성결대학교
컴퓨터공학부 조교수

- 관심분야 : 클라우드 데이터센터, 그린 통신망
- E-Mail : passjay@gmail.com