The devices in the Internet are nodes

Hiroki Kashiwazaki[1,a)]

Abstract: The researchers engaging to wide-area geographically distributed platform "Distcloud" have kept maintaining and operating the platform since the start of Distcloud, 2012. They have already lost their enthsiasm for expasion of Distcloud and development of Distcloud and DESTCloud that evaluate and validate fault tolerance of Distcloud have been slugging for a long time. Although an author have also sit around for years, he started to exert himself to the universe and his childfood dream again.

1. Introduction

Confidentiality, integrity and availability, also known as the CIA triad, is a model designed to guide policies for information security within an organization. Availability is highly challanged under circumastances of large scale natural disasters such as typhoons, eruptions and earthquakes. Especially in Japan, 2011 Tohoku earthquake and tsunami remain deep in our memories. Since the disaster, reserchers have started taking deeper note of the availability of information communication systems.

The birth of a term "Cloud computing" is beliebed to be in August, 2006. The term was given by Eric Schmidt of Google and he described their approach to SaaS as "cloud computing" at Search Engine Strategies Conference[*1]. By operation optimization of large scale computing environment, cloud computing showed a good availability. In fact, cloud computing environment such as Amazon Web Service, Google Apps and Sakura Cloud provided by Sakura Internet was used as an effective way to help evacuees in the 2011 earthquake.

In the correspondence of the 2011 earthquake, the term "beyond the cope of their assumption" (unexpected) was frequently used. The black swan theory or theory of black swan events is a metaphor that describes an event that comes as a surprise, has a major effect, and is often inappropriately rationalized after the fact with the benefit of hindsight[*2]. If an epicenter earthquake on the center of Tokyo occurs, any cloud computing environment in Tokyo can not work. To make good availability, various inter-operations among various cloud computing environments are necesarry.

A lot of approaches were tried to deal with the disasters to recover or to continue their business. Some reserchers tried to combine several cloud services and showed superiority from the aspect of agilities [1]. Also a lot of method to maintain or gain network availability was developed. Under circumstances of disasters, routing algorithms that make routing table according to network information such as delay time, available bandwidth and traffic demands can work well [2], [3]. Some routing methods collaborated with cloud computing environment also show good results [4]. But these approaches show superiority from the aspect of network connectivities but not the one of performances of application on the networks.

2. Related works

Thus, a lot of works to evaluate and validate the performance of a network itself or an application on the network quantitatively were published. Several research works have achieved to solve the problems. Consequently, new issues were also revealed.

2.1 Distcloud (v1)

In November, 2011, researchers including an author established Regional Inter-Cloud sub-Committee (RICC[*3]) under the Japan Society for the Promotion of Science (JSPS[*4]) 163rd Committee on Internet Technology (ITRC[*5]). Then the reserchers started a "Distcloud" research project[*6], which each participant provides computing and network resources and connect them each other to make a wide-area distributed virtualization infrastructure. Especially, the wide area distributed storage by clustering many computer resources located in geographically distributed areas where the number of sites is more than 2 ($N > 2$) attracted a high level of interest in the aspect of disaster recoveries and business continuity plans.

The file system works in an intercloud environment, that is, on the top of multiple independent cloud services. The important capability of the intercloud extension is that there is no need for users to take care of infrastructure layer, while the file system stores data into multiple cloud services, transparently and automatically. As the result, cloud users may have diversity about vendors and locations for data platform, as discussed in various

[1] Cybermedia Center, Osaka University, Ibaraki, Osaka, 567–0047, Japan
[a)] reo@cmc.osaka-u.ac.jp
[*1] https://www.google.com/press/podium/ses2006.html
[*2] https://en.wikipedia.org/wiki/Black_swan_theory
[*3] https://ricc.itrc.net/
[*4] https://www.jsps.go.jp/
[*5] https://www.itrc.net/
[*6] https://ricc.itrc.net/distcloud

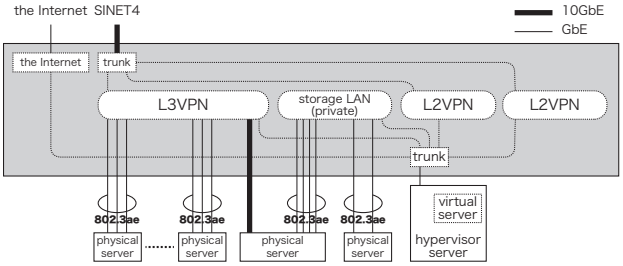

Fig. 1 A diagram of Distcloud network design

intercloud research projects. In 2013, the reserchers achieve to build a wide-area distributed storage system, which is now called "Distcloud File System (Distcloud-fs)" [5]. The storage supports a shared single POSIX file system so that LDLM (Long Distance Live Migration) of VMs (Virtual Machines) works well between multiple sites. On the SC13[*7], the reserchers demonstrated a nation wide live migration (about 500 km) in Japan and transpacific live migration (over 24,000 Km) [6].

They used Science Information NETwork (SINET)[*8] and Japan Gigabit Network (JGN)[*9] as network resources of Distcloud. In an early stage of Distcloud, the researcher use L2VPN[*10] and L3VPN[*11] services of SINET. VPN provides a secure communication environment for forming virtual organization on the network. SINET L2VPN provides a high degree of freedom and can be designed and used according to the application. SINET L3 VPN provided easy management for the increase and decrease of group members. A network designa and its diagram in the stage of Distcloud is shown in Fig.1. In the design and implementation of Distcloud network, availability of Distcloud is strongly relied on a robustness of SINET. Although an availability of Distcloud-fs for network faults was demonstrated and validated on SC13, a condition of faults is well-designed only to demonstrate, but was not emulate various situation of the disorders. They had a question how they can valiadate and evaluate an availability of distributed systems?

In Distcloud, every virtual and phisical computing resource are located under the several, fixed VLAN (IEEE 802.1Q), or virtual L3 network controlled by SINET. It is necesarry to increase or decrease a number of VLAN to construct more complex networks. L2 on demand of SINET[*12] is a good candidate to add VLAN for temoral research experiments in wide area network. But inside every organization, it is not always easy to add or remove VLAN. At that time after SC13 demonstration, some constraints were also found by them. In this paper, the author formally called the design and implementation of Distcloud as "Distcloud v1"

2.2 DESTCloud

The question turned to a motivation for them to start the research to validate and evaluate an availability of the system. They constructed a platform of virtual disaster drill to evaluate faults

[*7] http://sc13.supercomputing.org/
[*8] https://www.sinet.ad.jp/
[*9] http://www.jgn.nict.go.jp/
[*10] https://www.sinet.ad.jp/connect_service/service/l2vpn
[*11] https://www.sinet.ad.jp/connect_service/service/l3vpn
[*12] https://www.sinet.ad.jp/connect_service/service/l2ondemand

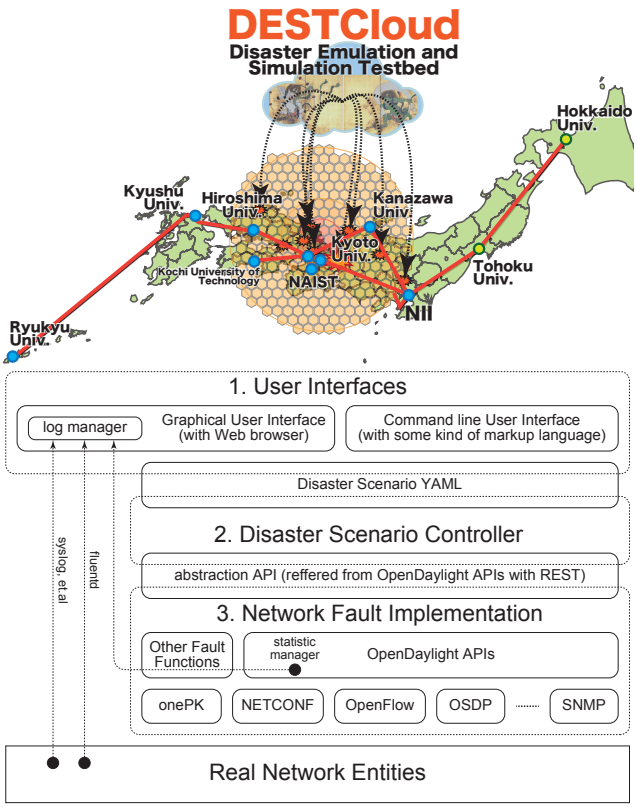

Fig. 2 A diagram of DESTcloud implementation and its platform overview

tolerance of information systems by emulating faults of networks, called "DESTCloud" in 2015. The platform enables the widely distributed system, especially the internet based distributed system to be designed and operated on an objective basis for widely distributed multiple faults. Natural disasters causes various wide area faults concurrently and continually. Especially, for wide area distributed system, it is prohibitively difficult and extraordinary costs too much for operators to implement network faults to geo-distributed network entities. There is a challenge to execute disaster drill to validate and evaluate the wide-area distributed system quantitatively and objectively. This is a motivation for authors to construct the validation and evaluation platform for a wide area distributed system automatically by using software defined network (SDN). By using SDN, network operators can be released from hard works of the disaster drill (Fig.2).

DESTCloud can control every network connection to shutdown, to change packet loss rate and delay time between every couple of two (phisical or virtual) network interface cards. If an administrator who try to use DESTCloud to evaluate network or application availability wants to change specific network connection condition, the connection between a specified couple of two nodes must be isolated VLAN from other VLAN. There are several ways to isolate the specified connection, but design and implementation of present Distcloud network can not solve to isolate. This difficulties became a motivation to change and improve Distcloud network design.

2.3 Distcloud (v2)

After the development of DESTCloud, the network design of Distcloud was altered. The change was also affected by the tran-

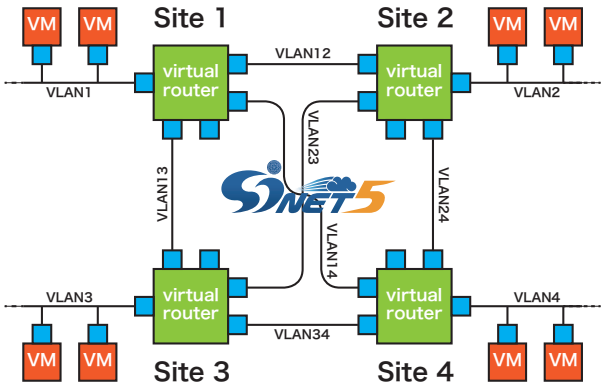

Fig. 3 A new design and implementation of Distcloud (v2)

sition of generation from physical machines to virtual machines. Computing resources provided by the participants of Distcloud became more powerful. Physical machines also became to be used as virtualization infrastructures. This change provides easy way to deploy the implementation of network functions virtualization (NFV). NFV is defined as a network architecture concept that uses the technologies of IT virtualization to virtualize entire classes of network node functions into building blocks that may connect, or chain together, to create communication services[*13].

By using NFV implementations such as VyOS[*14], whish is an open source network operating system that can be installed on physical hardware or a virtual machine on any servers, or any cloud platforms, the researchers could make a L3 network more easily under the single L2VPN of SINET. The new network design and implementation is shown in Fig.3.

Although virtual routers are located on the same VLAN, each couple of two virtual routers is connected with isolated unique VLAN. Each virtual router connects to other virtual routers aside from itself, so the network topology of virtual routers become fully connected mesh topology. The ethernet frame of the isolated unique VLAN is encapsuled by IEEE802.1ad[*15], also known as QinQ, on each virtual router. When the incoming frame is arrived at a specific virtual network interface on a virtual router, the frame is decapsuled and forwarded according to MAC address table. The new design of Distcloud let every virtual network connection line controllable.

L2 network issues on the network design of Distcloud v1 was solved by the way. But new L3 network issue revealed. Network segment on each site can be routed by some Interior Gateway Protocol such as RIP [7], OSPF [8]. But for some public cloud computing services, it is difficult to connect L3 networks of their virtual infrastructure environment to Distcloud because they recommend to use BGP-4 [9]. One of other issue is obscureness of inter-operability of various NFV implementation such as proprietary implementations that include Cisco ISO XRv 9000 Router[*16] and Cloud Services Router 1000V[*17]. According to the data sheet, Cisco 1000v is a virtual-form-factor router that delivers comprehensive WAN gateway and network services functions into virtual and cloud environments. Using familiar, industry-leading Cisco IOS ® XE Software networking capabilities, the CSR 1000v enables enterprises to transparently extend their WANs into provider-hosted clouds. Similarly, cloud providers themselves can use it to offer enterprise-class networking services to their tenants or customers[*18]. Juniper vMX[*19] and vSRX[*20][*21] also can be candidates. According to vMX's data sheet, vMX Virtual Router can be a virtualized MX Series 3D Universal Edge Router that helps network operators of all types improve customer experience and profitability by increasing network and service agility and accelerating time to market for new services, while streamlining their operations environment. Although Brocade vRouter 5400 and 5600 can be candidates, on 2nd June 2017, AT&T bought the Vyatta assets of Brocade Communications, including the Vyatta vRouter, so these implements of NFV have unknown consequences[*22]. Arista vEOS Router[*23] and so on[*24] is one of strong candidates. Arista have offered vEOS to Distcloud project until the end of March 2018.

The number of Participation organization of Distcloud was 13 including Hokkaido University[*25], Tohoku University[*26], National Institute of Informatics (NII)[*27]), Tokyo Institute of Technology[*28], Kanazawa University[*29], Nara Institute of Science and Technologyhttp://isw3.naist.jp/, Kyoto University[*30], Osaka University[*31], Hiroshima University[*32], Kochi University of Technology[*33], Kyushu Sangyo University[*34], University of Ryukyu[*35] and University of California San Diego[*36] in September 2017. With the collaborative research with NII, some public cloud computing services including Amazon Web

[*13] https://en.wikipedia.org/wiki/Network_function_virtualization
[*14] https://vyos.io/
[*15] http://www.ieee802.org/1/pages/802.1ad.html
[*16] https://www.cisco.com/c/en/us/products/collateral/routers/asr-9000-series-aggregation-services-routers/datasheet-c78-734034.html
[*17] https://www.cisco.com/c/en/us/products/routers/cloud-services-router-1000v-series/index.html
[*18] https://www.cisco.com/c/en/us/products/collateral/routers/cloud-services-router-1000v-series/datasheet-c78-733443.html
[*19] https://www.juniper.net/us/en/products-services/routing/mx-series/vmx/
[*20] https://www.juniper.net/us/en/products-services/security/srx-series/vsrx/
[*21] It can also work as a router.
[*22] http://about.att.com/story/att_to_acquire_vyatta_software_technology_from_brocade.html
[*23] https://www.arista.com/en/products/eos
[*24] Although the author have not validated, SEIL/x86 developed by IIJ can be one of the candidates.
[*25] http://www.iic.hokudai.ac.jp/
[*26] http://www.riec.tohoku.ac.jp/
[*27] http://www.nii.ac.jp/
[*28] http://www.gsic.titech.ac.jp/
[*29] https://www.imc.kanazawa-u.ac.jp/
[*30] http://www.media.kyoto-u.ac.jp/
[*31] http://www.cmc.osaka-u.ac.jp/
[*32] https://www.media.hiroshima-u.ac.jp/
[*33] http://www.kochi-tech.ac.jp/kut/research_activities/kikou.html
[*34] http://welcome.is.kyusan-u.ac.jp/
[*35] https://ie.u-ryukyu.ac.jp/
[*36] https://ucsd.edu/

Fig. 4 Distcloud network diagram 2018

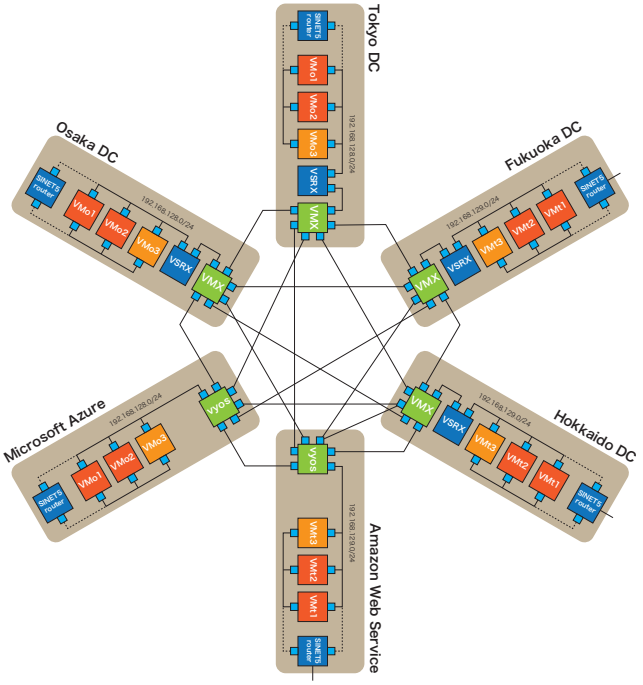

Fig. 5 Next generation infrastructure diagram of Distcloud.

Service[*37] and Microsoft Azure[*38] were avilable to connect to Distcloud and NFV environments in 4 SINET datacenter (Sapporo, Tokyo, Osaka and Fukuoka) were also avilable in October, 2017. To utilize those additional computing and network resource, the desing and implementation of Distcloud have to be improved (Fig.4).

3. New Distcloud design

To collaborate and inter-connect with several public cloud computing services, L3 routing protocol is determined to BGP-4. Each site should have private AS number and advertise each other in a private L2VPN network provided by SINET5. Full connected mesh topology can easily add the new participating organization. But issue of scaleability is apparent because of IEEE 802.1Q 12 bits constraint of VID (VLAN Identifier). Exclude VLAN ID 0 and 4095, 4094 VLAN ID is available with using IEEE 802.1Q. New design of Distcloud network require unique VLAN ID in each couple of two virtual router. If the number of participating organizations is n, the total number of VLAN required is $\frac{n(n-1)}{2}$ and should be less than 4094. The inequation means that upper limit of n is 90. To resolve the issue of upper limit, other protocol of L2 extension such as VXLAN [10] and NVGRE [11] should be considered. New Distcloud diagram of infrastructure including NFV environment and public cloud computing service is shown in Fig.5.

[*37] https://aws.amazon.com/
[*38] https://azure.microsoft.com/

4. Conclusion

This paper review the complementary evlolution betwenn Distcloud and DESTCloud as an anti-these of Distcloud. New design is described and other new issues are also revealed. Quantitative evaluations of new design of Distcloud will be shown in the IOT40 meeting.

References

[1] Kashiwazaki, H.: Practical Uses of Cloud Computing Services in a Japanese University of the Arts Against Aftermath of the 2011 Tohoku Earthquake, *Proceedings of the 40th Annual ACM SIGUCCS Conference on User Services*, SIGUCCS '12, New York, NY, USA, ACM, pp. 49–52 (online), DOI: 10.1145/2382456.2382467 (2012).
[2] Kashiwazaki, H., Kobayashi, S., Kawai, S., Ohishi, N. and Takai, Y.: An Adaptive Approach for Network Traffic Load Balancing by Using One-Way Delay, *2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet*, pp. 345–350 (online), DOI: 10.1109/SAINT.2012.62 (2012).
[3] Kashiwazaki, H. and Takai, Y.: The Simulation Evaluation for an Adaptive Network Routing by Using Latency Information, *2011 IEEE/IPSJ International Symposium on Applications and the Internet*, pp. 480–485 (online), DOI: 10.1109/SAINT.2011.89 (2011).
[4] Kashiwazaki, H. and Takai, Y.: An Evaluation of Adaptive Traffic Engineering for Delivering High Bandwidth Movie on Overlay Routing Network, *2014 IEEE 38th International Computer Software and Applications Conference Workshops*, pp. 614–619 (online), DOI: 10.1109/COMPSACW.2014.103 (2014).
[5] Nakagawa, I., Kashiwazaki, H., Shimojo, S., Ichikawa, K., Kondo, T., Kitaguchi, Y., Kikuchi, Y., Yokoyama, S. and Abe, S.: A Design and Implementation of Global Distributed POSIX File System on the Top of Multiple Independent Cloud Services, *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 867–872 (online), DOI: 10.1109/IIAI-AAI.2016.75 (2016).
[6] Nakagawa, I., Ichikawa, K., Kondo, T., Kitaguchi, Y., Kashiwazaki, H. and Shimojo, S.: Transpacific Live Migration with Wide Area Distributed Storage, *2014 IEEE 38th Annual Computer Software and Applications Conference*, pp. 486–492 (online), DOI: 10.1109/COMPSAC.2014.71 (2014).
[7] Malkin, G. S.: RIP Version 2, STD 56, RFC Editor (1998).
[8] Moy, J.: OSPF Version 2, STD 54, RFC Editor (1998).
[9] Rekhter, Y., Li, T. and Hares, S.: A Border Gateway Protocol 4 (BGP-4), RFC 4271, RFC Editor (2006).
[10] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M. and Wright, C.: Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks, RFC 7348, RFC Editor (2014).
[11] Garg, P. and Wang, Y.: NVGRE: Network Virtualization Using Generic Routing Encapsulation, RFC 7637, RFC Editor (2015).