

VIRTUAL ROUTING FUNCTION DEPLOYMENT IN NFV-BASED NETWORKS UNDER NETWORK DELAY CONSTRAINTS

Shin-ichi Kuribayashi

Department of Computer and Information Science, Seikei University, Japan

ABSTRACT

NFV-based network implements a variety of network functions with software on general-purpose servers and this allows the network operator to select any capacity and location of network functions without any physical constraints. The authors proposed an algorithm of virtual routing function allocation in the NFV-based network for minimizing the total power consumption or the total network cost, and developed effective allocation guidelines for virtual routing functions.

This paper evaluates the effect of the maximum tolerable network delay on the guidelines for the allocation of virtual routing functions, which minimizes the total network cost. The following points are clear from quantitative evaluations: (1) The shorter the maximum tolerable network delay, the greater the number of areas where the routing function must be allocated, resulting in an increase in the total network cost. (2) The greater the routing function cost relative to the circuit bandwidth cost, the greater the increase in the total network cost caused by the maximum tolerable network delay. This paper also provides the possible guideline how to decide the value of maximum tolerable network delay when the condition of allowable increase in network cost is given.

KEYWORDS

NFV, resource allocation, virtual routing function, minimum total network cost

1. INTRODUCTION

In recent years, network carriers have begun to implement network functions based on a new technology called network functions virtualization (NFV) [1]-[3]. In an NFV-based network, network functions are implemented in a virtual manner by software on general-purpose servers. NFV makes it possible to select both the capacity and location of each function without physical constraints. Therefore, to build an economical NFV-based network, it is important to determine where each network function should be allocated in the network and what its capacity should be. The authors proposed an algorithm of virtual routing function allocation in the NFV network for minimizing the total power consumption or the total network cost, and developed effective allocation guidelines for virtual routing functions [4],[5]. Specifically, it was found that the greater the power consumption of the routing function relative to that of the circuit bandwidth, and the greater the peak coefficient of traffic between areas is, the smaller the total power consumption or the total network cost can be made by integrating the routing functions of some areas into those of a small number of areas. It was also indicated that there can be cases where distributing low-capacity routing functions in the network, which has been difficult to implement in conventional networks but is made possible by NFV, can dramatically reduce the network

power consumption. The above studies were not meant just to address the VNE (virtual network resource allocation) problem [6],[7]. Rather, the aim was at developing function allocation guidelines, which show how much capacity each virtual routing function is required to have and what the effective way of allocating each function is depending on the ratio of the cost of a routing function to that of a circuit bandwidth and the peak coefficient of traffic between areas.

References [4] and [5] did not take network quality into consideration. This paper evaluates the effect of the maximum tolerable network delay on the guidelines for allocating the virtual routing function. The objective function in this paper is to minimize the total network cost.

The rest of this paper is organized as follows. Section 2 explains related works. Section 3 proposes to enhance the existing routing function allocation algorithm, which was proposed in References [4] and [5]. Section 4 evaluates the effect of the maximum tolerable network delay on the function allocation guidelines, with a ladder-shaped network model that simulates the network structure of Japan. Finally, Section 5 presents the conclusions. This paper is an extension of the study in Reference [17].

2. RELATED WORK

One of the related studies is the VNE (virtual network resource allocation) problem, a problem of allocating virtual nodes and links to a physical network efficiently [6],[7]. Reference [8] evaluates the allocation of multiple VNFs. Reference [9] proposes a functional allocation assuming a hybrid arrangement in which dedicated hardware and virtual functions are used. References [10] and [11] propose dynamic resource allocation and scheduling. Reference [12] evaluates the virtualization of S-GW (serving GW) and P-GW (packet data network GW) functions in a mobile network in terms of network load and network delay. Based on an evaluation made on a network that simulates the network structure of the U.S.A., it shows that four data centers that completely virtualize these GWs are required. Reference [13] proposes a method for solving the virtual DPI allocation problem, which is aimed at minimizing the cost.

We proposed a resource allocation method in a cloud environment in which computing power and access bandwidths are allocated simultaneously [14]-[16]. The key point of this method is how to pack as many requirements (simultaneously requiring two different types of resource of computing power and access bandwidths with service quality being not uniform) as possible into multiple sets of computing power and bandwidths.

Most of the above evaluations deal with how to pack (allocate) virtual network functions to finite physical network resources efficiently. In contrast, references [4] and [5] and this paper try to develop allocation guidelines that indicate how much capacity is required for each virtual network function and what is the effective way of allocating each function, both of which depend on the relative costs of the routing function and the circuit bandwidth and the inter-area traffic distribution. We believe that these guidelines provide network carriers with essential information when they design and build an NFV-based network in a medium to long term.

3. ENHANCED ROUTING FUNCTION ALLOCATION ALGORITHM UNDER NETWORK DELAY CONSTRAINTS

It is proposed to enhance the existing routing function allocation algorithm proposed in References [4] and [5], in order to ensure that the network delay does not exceed the maximum tolerable network delay. In this paper, the constraint of the maximum tolerable network delay is applied to only intra-area communication and communication between adjacent areas. This is because, in communication between remote areas, the locations of routing functions do not greatly affect communication distances. Moreover, we evaluate the maximum tolerable network delay with the maximum number of transit areas, N .

In the existing allocation algorithm [4]-[5], the routing function is first allocated to each area. Next, an area with the smallest number of packets received and transmitted over a certain period is selected. If integrating the routing function of this area (integration-origin area) into the routing function of another area (integration-destination area) reduces the total network cost, this integration is done. If not, the integration processing will not be done. Then, the integration judgment is similarly performed on the next integration-origin area candidate. This process is terminated when there is no integration-origin area left.

The following extension is proposed in the existing routing function integration process, to ensure that the network delay does not exceed the maximum tolerable network delay. That is, the integration processing will not be carried out if there is even one communication flow that does not satisfy N in the integration judgment for each integration-origin area candidate.

4. EVALUATION OF THE PROPOSED ROUTING FUNCTION ALLOCATION ALGORITHM

4.1 EVALUATION CONDITIONS

(1) We have developed a simulation program in C language that executes the enhanced allocation algorithm proposed in Section 3.

(2) Network structure: The ladder-shaped model illustrated in Figure 1, which simulates the shape of Japan, is used as in References [4] and [5].

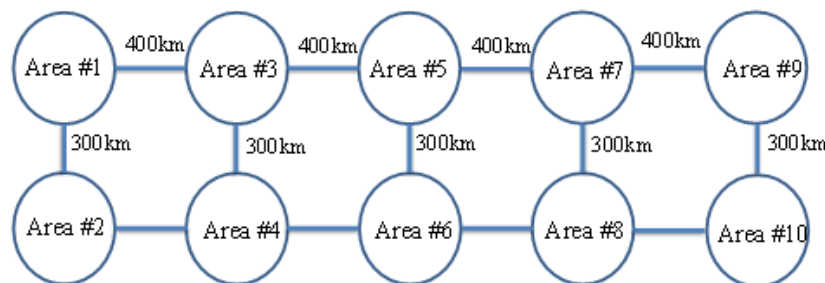


Figure 1. Ladder-shaped network topology which simulates the shape of Japan

(3) Traffic flow: The traffic flow is the same as that used in References [4] and [5]. The amount of traffic between Area #5 and each of the remaining areas is M times (peak coefficient) that of traffic between other areas. Thus, M=1 means that traffic flows between areas are uniform.

(4) Since the ratio of the cost of the routing function to the cost of the circuit bandwidth greatly affects the total network cost, the following ratio is defined as in Reference [5]:

$$Z_c = \alpha c / \beta c \tag{1}$$

Where αc is the cost of the routing function per pps (packets per second), and βc the cost of the circuit bandwidth per Mbps per 10 km. The fixed costs of the routing function and circuit bandwidth are not taken into consideration here.

(5) Routing policy: We revise the routing policy in References [4] and [5] so that the cost will be lower.

The route selection procedure is explained below.

- ① Select the route from the source area to the destination area in the horizontal direction.
- ② If one of the source area number and destination area number is an odd number and the other is an even number, select the route until a destination area appears upward or downward. Afterwards if the destination area number is odd, the upward direction is selected. If it is an even number, the downward direction is selected.
- ③ If there is no area where the route selection function is deployed on the route based on ① and ②, select the area in which virtual routing function is deployed as the transit area, that will be the shortest distance when relaying the packet.
- ④ The route selection both from source area to transit area and from transit area to destination area follows ① and ②.

An example of route selection based on the above route selection procedure is shown in Figure 2. Route selection function is allocated in Area # 4 and Area # 7, and shows the route when packet is transferred from Area # 1 to Area # 6, Area # 9, and Area # 10.

Moreover, the allowable maximum number of transit areas in the ladder-shaped network topology in Figure 2 could be 7 as shown in Figure 3.

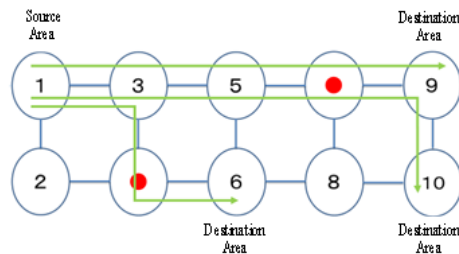


Figure 2. Supposed route selection policies in the ladder-shaped network topology in Figure 1

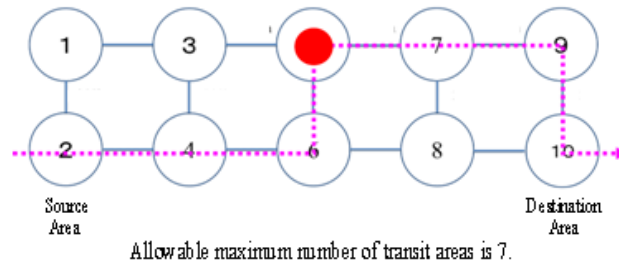


Figure 3. Allowable maximum number of transit areas in the ladder-shaped network topology in Figure 1

4.2 EVALUATION RESULTS AND DISCUSSIONS

The evaluation results are shown in Figures 4 to 10. Figure 4 shows the result when Z_c is 2.0 and M is 2. Figures 5, 6, and 7 illustrate graphs in which Z_c is changed to 1.0, 0.5, 0.1, respectively. Figure 8 shows the result when M is 20 and Z_c is 1. Figures 6, 7, 5 illustrate graphs in which M is changed to 10, 5, 2, respectively. The horizontal axis in Figures shows the allowable maximum number of transit areas (N). The vertical axis shows the total network cost (normalized value). Moreover, Figures 5, 7 and 9 illustrate the cost of optimal deployment (dashed line), in addition to the cost based on the proposed algorithm (solid line). Here, we calculate all the deployment patterns and decide the deployment with the lowest cost as the optimal deployment. Figure 11 shows the final allocations when N is varied under the conditions of Figure 5. The following points are clear from these Figs:

1) The smaller N is (i.e., the more severe the network delay requirement is), the greater the total network cost. However, in cases where routing functions end up being allocated to multiple areas (as in Figure 7), a decrease in N does not greatly increase the total network cost unless N is extremely small (for example, N is 2).

<Reason> As shown in Figure 11, in cases where there is no restriction on N (N is 7) under the conditions of Figure 5, concentrating the routing function on Area #5 alone is the most economical. However, as N becomes smaller, it becomes necessary to allocate the routing function to many areas in order to satisfy the smaller N .

The value of maximum tolerable network delay can be decided if the condition of allowable increase of network cost is given. For example, if the allowable maximum number of transit areas is limited to 3 in the example of Figure 4, the total network cost increases by about approximately 40%, resulting in 40% increase in service fee as well. If the carrier would like to limit the increase in service fee to 20% or less, it should limit the allowable maximum number of transit areas to 4 in this example. One of the main objectives of this paper is to be able to provide such network design guidelines.

2) The greater Z_c is and the greater M is, the greater an increase in the total network cost caused by the restriction on N .

<Reason> If Z_c is small, the cost of the routing function is small relative to the cost of the circuit bandwidth. As a result, the cost increase caused by an increase in the number of areas (to which the routing function is allocated) in order to satisfy N becomes relatively small. As M increases, the cost reduction achieved by the integration of the routing function increases and it becomes more desirable to integrate the routing function.

3) The proposed algorithm could get results that are close to the optimal deployment by 85% or more in this evaluation, as in Figures 5, 7, and 9.

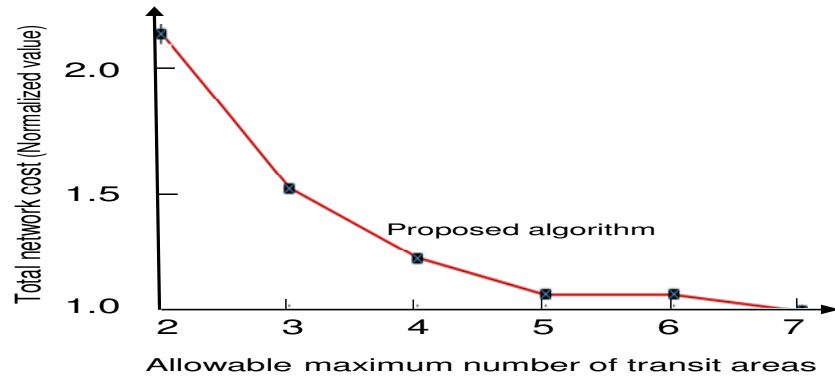


Figure 4. Simulation result 1 ($Z_C=2.0, M=2$)

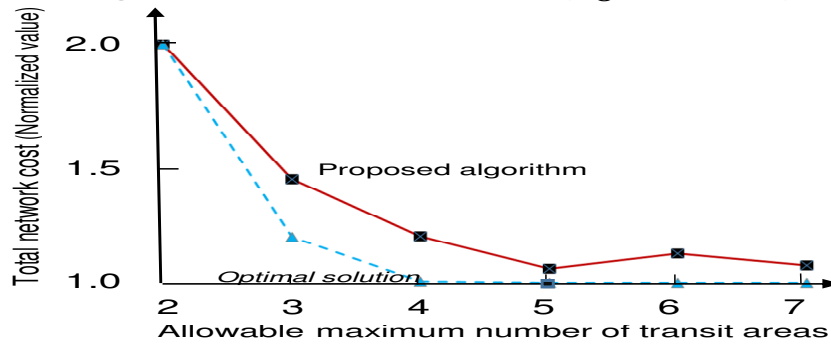


Figure 5. Simulation result 2 ($Z_C=1.0, M=2$)

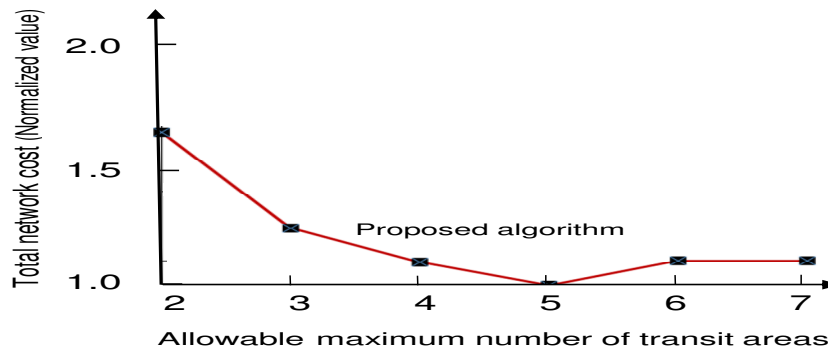


Figure 6. Simulation result 3 ($Z_C=0.5, M=2$)

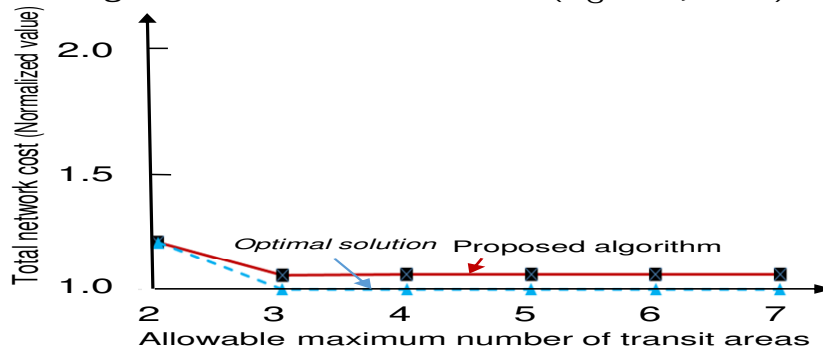


Figure 7. Simulation result 4 ($Z_C=0.1, M=2$)

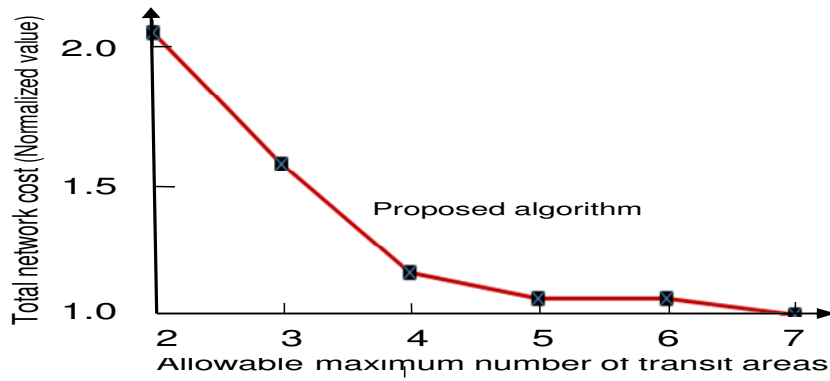


Figure 8. Simulation result 5 ($Z_C=1.0$, $M=20$)

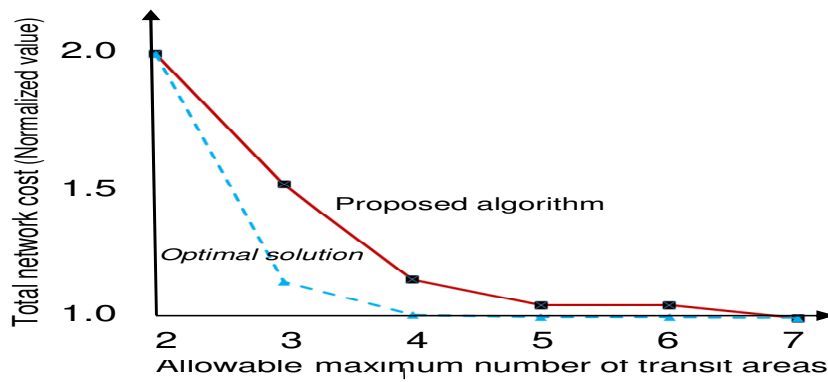


Figure 9. Simulation result 7 ($Z_C=1.0$, $M=10$)

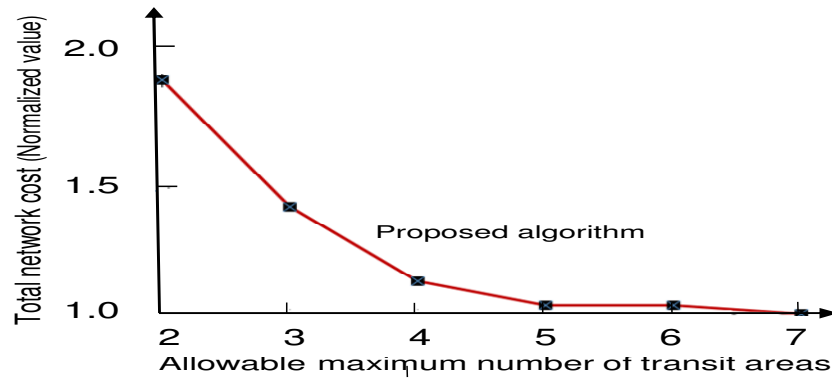


Figure 10. Simulation result 7 ($Z_C=1.0$, $M=5$)

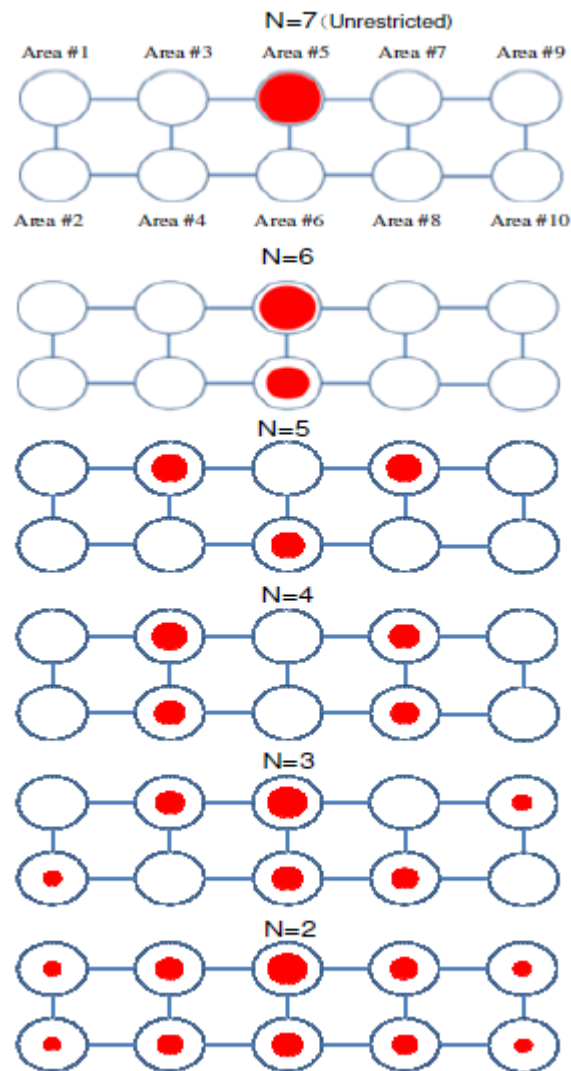


Figure 11. Final topology of virtual routing function allocation ($Z_C=1.0$, $M=2$)

Although the specific network topology is assumed in the evaluation of this paper, the major trend and the network design guidelines clarified this time could be applicable in the same way, even if the number of areas is increased to 100, or a star or loop topology is assumed. As the proposed algorithm, which repeats the judgment as to whether the total network cost is lower if the virtual routing functions in the specific area is integrated in the other area, is also applicable, the similar evaluation results such as ‘the shorter the maximum tolerable network delay, the greater the number of areas where the routing function must be allocated, resulting in an increase in the total network cost’ could be expected. It is noted that the areas in which virtual routing functions are actually deployed depend on the network topology.

5. CONCLUSIONS

This paper has evaluated the effect of the maximum tolerable network delay on the guidelines for the allocation of virtual routing functions. Our results with the ladder-shaped network have revealed the following: (1) The shorter the maximum tolerable network delay, the greater the number of areas where the routing function must be allocated, resulting in an increase in the total network cost. (2) The greater the routing function cost relative to the circuit bandwidth cost, the greater the increase in the total network cost caused by the maximum tolerable network delay. This paper has also provided the possible guideline how to decide the value of maximum tolerable network delay when the condition of allowable increase of network cost is given. Moreover, this paper has also indicated that the proposed algorithm could get results that are close to the optimal solution by 85% or more.

It is required to study cases where the fixed costs of the routing function and circuit bandwidth are taken into consideration, and where multiple virtual network functions (ex. firewall function, DPI function) are to be allocated.

ACKNOWLEDGMENT

We would like to thank Mr. Kenichiro HIDA for his help with the evaluation.

REFERENCES

- [1] M. Chiosi et al., "Network Functions Virtualization - An Introduction, Benefits, Enablers, Challenges and Call for Action," ETSI NFV, Oct. 2012.
- [2] "Network Functions Virtualisation (NFV); Architectural Framework," ETSI GS NFV 002 v1.2.1, Dec. 2014.
file:///C:/Users/Kuribayashi/Downloads/gs_NFV002v010201p.pdf
- [3] R.Mijumbi, J.Serrat, J.Gorricho, N.Bouten, F.D.Trurck and R.Boutaba, "Network Function Virtualization: State-of-the-art and Research Challenges," IEEE Communications Surveys & Tutorials, Vol. 18, Issue 1, pp.236-262, 2016.
- [4] K.Hida and S. Kuribayashi, "Virtual Routing Function Allocation Method for Minimizing Total Network Power Consumption," 18th International Conference on Information and Communication Systems (ICICS Venice 2016), Aug. 2016.
- [5] S.Kuribayashi, "Flexible Virtual Routing Function Deployment in NFV-based network with Minimum Total Network Cost," International Journal of Computer Networks & Communications (IJCNC) Vol.8, No.5, pp.15-26, Nov.2016.
- [6] A.Fischer, J.F.Botero, M.T.Beck, H.Meer, and X.Hesselbach, "Virtual Network Embedding: A Survey", IEEE Communications Surveys & Tutorials, Vol. 15, No.4, 2013.
- [7] X.Wei, S.Hu, H.Li, F.Yang and Y.Jin, "A Survey on Virtual Network Embedding in Cloud Computing Centers", The Open Automation and Control Systems Journal, Vol. 6, pp.414-425, 2014.
- [8] S. Mehraghdam, M. Keller, and H. Karl, "Specifying and placing chains of virtual network functions," in Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on, Oct 2014, pp. 7-13.
- [9] H. Moens and F. D. Turck, "VNF-P: A model for efficient placement of virtualized network functions," in Network and Service Management (CNSM), 2014 10th International Conference on, Nov 2014, pp. 418-423.
- [10] R.Mijumbi, J.Serrat, J.Gorricho, N.Boutenz, F.Turckz and S.Davy, "Design and Evaluation of Algorithms for Mapping and Scheduling of Virtual Network Functions," 1st IEEE Conference on Network Softwarization (NetSoft), pp.1-9, 2015.

- [11] S.Clayman, E.Mainiy, A.Galis, A.Manzaliniz and N. Mazzocca, "The Dynamic Placement of Virtual Network Functions,"IEEE Network Operations and Management Symposium (NOMS),pp.1-9, 2014.
- [12] A.Basta, W.Kellerer, M.Hoffmann, H.J.Morper, and K.Hoffmann, "Applying NFV and SDN to LTE Mobile Core Gateways; The Functions Placement Problem," 4th Workshop on All Things Cellular: Operations, Applications and Challenges 2014, pp.33-38, Aug. 2014.
- [13] M.Bouet, J.Leguay and V.Conan, "Cost-based placement of vDPI functions in NFV infrastructures," International Journal of Network Management, pp.490-506, 2015.
- [14] S.Tsumura and S.Kuribayashi, "Simultaneous allocation of multiple resources for computer communications networks", APCC2006, 2F-4 (2006.8)
- [15] S.Kuribayashi, "Optimal Joint Multiple Resource Allocation Method for Cloud Computing Environments", International Journal of Research and Reviews in Computer Science (IJRRCS), Vol.2, No.1, pp.1-8, Feb. 2011.
- [16] S.Kuribayashi,"Resource Allocation Method for Cloud Computing Environments with Different Service Quality to Users at Multiple Access", International Journal of Computer Networks & Communications (IJCNC) Vol.7, No.6, pp.33-51, Nov.2015.
- [17] K. Hida and S. Kuribayashi, "NFV-based virtual routing function deployment under network delay," 19th International Conference on Computer Communications and Applications (ICCCA Kyoto 2017), pp.862-866, Nov. 2017.

AUTHOR

Shin-Ichi kuribayashi Received The B.E., M.E., And D.E. Degrees From Tohoku University, Japan, In 1978, 1980, And 1988 Respectively. He Joined Ntt Electrical Communications Labs In 1980. He Has Been Engaged In The Design And Development Of Ddx And Isdn Packet Switching, Atm, Phs, And Imt 2000 And Ip-Vpn Systems. He Researched Distributed Communication Systems At Stanford University From December 1988 Through December 1989. He Participated In International Standardization On Atm Signaling And Imt2000 Signaling Protocols At Itu-T Sg11 From 1990 Through 2000. Since April 2004, He Has Been A Professor In The Department Of Computer And Information Science, Faculty Of Science And Technology, Seikei University. His Research Interests Include Resource Management For Nfv And Sdn-Based Networks, Qos Control, Traffic Control For Cloud Computing Environments And Green Network. He Is A Member Of Ieee, Ieice And Ipsj.

