Dynamic Network Provisioning with AI-enabled Path Planning

Hong-Nam Quach
Department of Electronics and
Computer Engineering
Chonnam National University
Gwangju, South Korea
Email:quachhongnam1995@gmail.com

Chulwoong Choi
Department of Electronics and
Computer Engineering
Chonnam National University
Gwangju, South Korea
Email: sentilemon02@gmail.com

Kyungbaek Kim
Department of Electronics and
Computer Engineering
Chonnam National University
Gwangju, South Korea
Email: kyungbaekkim@jnu.ac.kr

Abstract— **As the number of mobile devices increases and the concept of 5G networks becomes popular, various network infrastructures and services emerge. Also, more users request user-specific network services within limited network resources. Under this complex situation, in order to provide a guaranteed QoS level to users, it requires to consider various factors which affect the performance of network service, and dynamic network provisioning is required. In this paper, we propose a dynamic network provisioning system with AI-enabled path planning, which uses the side channel information such as disaster events, maintenance events, and distribution of users. In this system, we design a user request handler that understands user-specific QoS and spatio-temporal requirements in order to maximize the utilization of a given network resource. Also, this system utilizes side-channel information to optimizing the network provisioning in a real-time manner.**

Keywords—Software-defined Networking, Q-routing, network provisioning, Artificial Intelligence, location-aware network, bandwidth demand constraints.

I. INTRODUCTION

In recent days, when the COVID-19 virus broke out, people could not go out, resulting in the need to use Internet services increased significantly. For example, when teachers start online classes, many students access and use limited local network resources. The sudden increase in network services creates various problems, such as node congestion. As a result, many users cannot use network services. This fact raises the question of how to leverage the limited network resources to provide personalized network services and respond to a variety of user requests with various constraints, for example, QoS guarantee, selecting locations, and duration.

In this paper, we design a new framework using AI-based network planning to make the most of limited network resources and provide services to as many users as possible as efficiently. Data such as region, time, and bandwidth are collected from the users who requested the services. We combine the use of a big-data system with external data such as weather, disaster, network maintenance, pre-processed user activities, and analyzed. The AI-enabled big-data system supports the analysis and offers the results as information about network planning. Besides, based on customer information and massive data system results, we use artificial intelligence, machine learning technicals to planning network resources, and processing customer requests. Also, it is expected that extensive data accumulated from big data systems and artificial intelligence learned from them will be able to efficiently provide network services even in urgent situations such as disasters.

II. RELATED WORK

In our previous studies [1], [2], we have proposed an SDN-based framework that handles user requests, including locations, QoS levels, and implements a proper network service dynamically. To support the location-aware dynamic network provisioning, researchers proposed a method based on the user who has requested locations for generating a flexible network slice. Through a map-based web interface, the required areas which represent the required bandwidth and delay asked by users. After obtaining the user request, our framework will map out the corresponding network switches in the resource, then generate network paths between these network switches and guarantee the given QoS level. After that, an SDN controller implements the flow rules to the switches above to launch the generated network paths. However, how to calculate and provide a reasonable route between selected switches becomes an important issue. Thus, researchers need to answer the question of how network providers can make decisions on whether the requests are acceptable to their limited resources. Thus, M.Huang et al. proposed online algorithms with an auxiliary graph for unicast and multicast requests, including a bandwidth constraint, and network throughput maximized [4] to solve this problem. Besides, W.Liang et al. also proposed algorithms with an auxiliary graph for NFV-enable requests, including bandwidth and delay constraints, to minimize the operational cost and maximize the network throughput [5]. Additionally, in another research [3], Huu-Duy et al. also adopt algorithms with an approach to accept users' requirements (Spatio-Temporal QoS requests), including location, usage duration, and QoS.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2017R1A2B4012559). This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2020-2016-0-00314) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation). This research was one of KOREN projects supported by National Information Society Agency (No.1711117098).

*Prof. Kyungbaek Kim is the corresponding author.

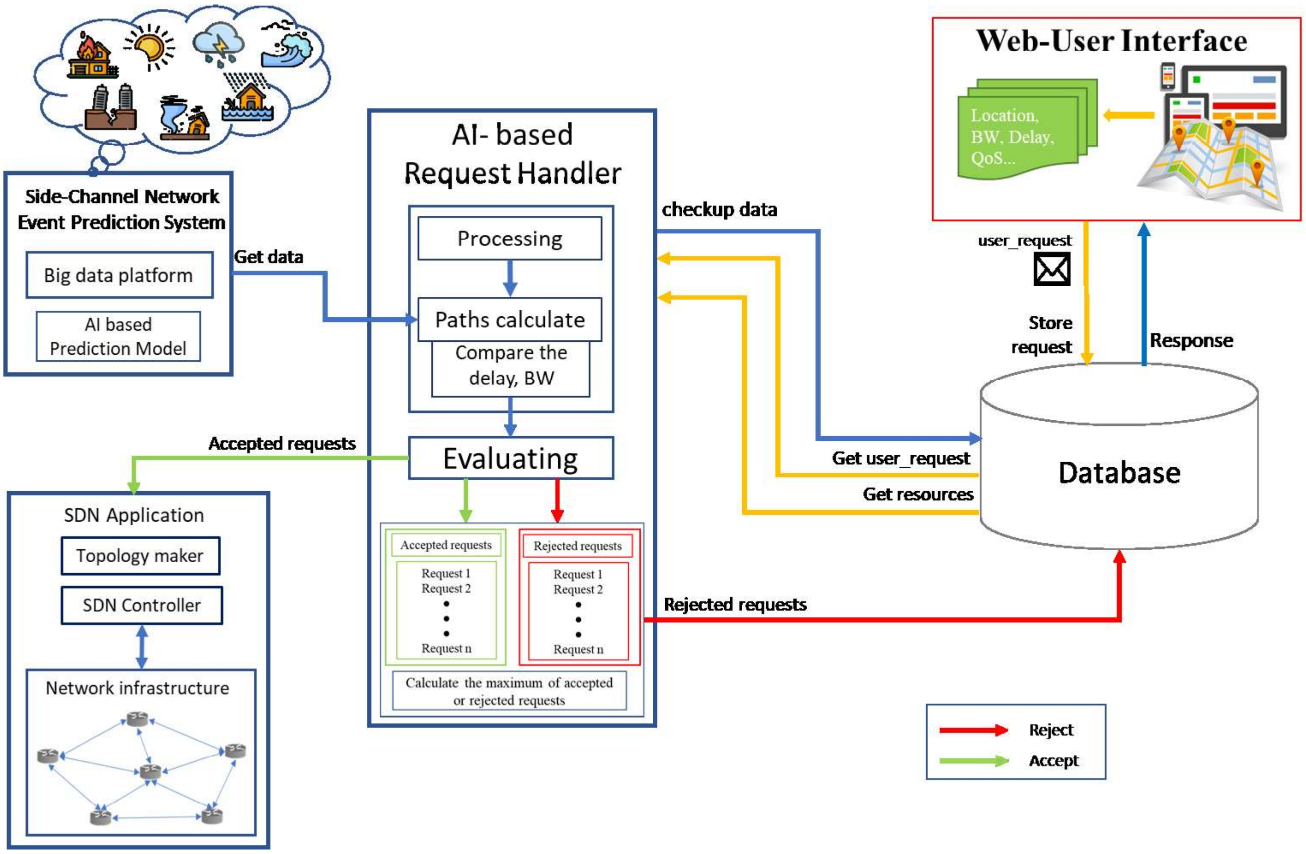

Figure 1. An overview of the architecture framework

Although the above researches have solved the problem of calculating the route, optimizing the network throughput. Still, they have not considered the ability to troubleshoot when the network suddenly appears, such as a bottleneck, broken connection devices, broken connection links between devices, and natural disaster issues. The actual Internet is a dynamic network and constantly changes state. It is difficult to break the connection, but in a real-world environment, when user traffic is high, it is likely to lead to congestion at a node in the network. When congestion occurs, we need to calculate and find alternative paths promptly with little effect on the user. Also, machine learning techniques are increasingly widely used in many fields. Some studies use machine learning to solve the deadlock problem in network operation, management [6] [7], and smart, customizable, and detailed routing management for SDN. Reinforcement learning is a technique that can continuously explore the surrounding environment through agents without prior knowledge, familiar with the entire climate after several training cycles, and make the right choice finally. Thus, it is suitable to address the management challenges in the network.

III. DESIGN OF PROPOSED OUR FRAMEWORK

In this section, we describe the architecture of a framework that developed on our previous framework [2], [3] for location-aware dynamic network provisioning with the user's constraints. In our framework, we propose five primary components of the architecture that have to define for implementation, as illustrated in figure 1.

Web-User Interface (Web-UI): The Web - UI is a website interface, map-based selector - regions to assist users in sending service requests to Internet services providers. In this web interface, we use an API map from google map and create the region selector as a grid to support the location-aware request. Besides, Web-UI implemented using Socket.io and HTML with node.js. We designed Web-UI into two parts, one for the client-side for the user to assist in collecting requests from the client. The rest linked as a web server by using socket.io, which tasked with listening, receiving, and processing requests from the user. The grid will be used by users to set up and select areas for service registration. After selecting the location to request the service, users can choose the requirements for service quality, such as bandwidth and delay. Moreover, web-UI not only assists customers in sending service requests to ISPs but also responds to them in real-time. After obtaining the user request through the web-interface, we store those requests to the database for processing. User's requests obtained from web interface could be requested, such as location, time for service, QoS, number of users using the service, bandwidth, delay, etc.

Database – To support location-aware network provisioning, obtain the user's requests, and handle the demands, the database is an essential component. We use MySQL to deploy and design the database. Our database not only contains information about device resources but also used to store user requests. When a new request sent from the user via the Web-UI, the web server will process the request's information into the database. Besides, our databases also contain request information in and after processing and make it possible for users to track requests. The database is a warehouse for storing all useful information retrieved dynamically and periodically.

AI-based Request Handler (Network planner) – Which gets the user's requests from the database to handle. After getting the request from the database, the request handling performs a reverse query with the database to get the necessary information and map the selected locations in the request to the resource devices. Then, the database sends the resource to request handling, that is the list selected devices. When the request handling receives the list selected, it goes to the next step for calculating and finding paths to make the connection between the selected devices. Here, the algorithms for calculating and finding the shortest path (Dijkstra, LARAC, Bellman-Ford) used to deploy. Besides, finding the shortest path, the calculation step also compare the QoS binding parameters obtained from the user's request with the existing resources to decide to accept or reject the user requests. After deciding to accept or reject the user request, we perform evaluating to review the number of accepted requests and the number of dismissed requests. In the evaluating, we deploy the table which lists the accepted requests and rejected requests - the accepted requests sent to the *SDN Application* to implement the logical topology network and then store in the database. The accepted request contains information that is the link between the selected device and the required QoS constraint parameter of the user. The rejected requests stored in the database to waiting for re-handle. Through the database and web – UI, we respond to the user's request with the request status.

SDN Application – The SDN-Application defined as a component that generated and implemented the logical topologies network based on the accepted requests. The SDN-Application includes the topology maker and network controller. After receiving the accepted request, it sent to topology maker to generate topology which contains switches along the candidate routing paths calculated by Request Handling. The network controller is responsible for receiving and processing the logical topology network derived from the topology maker. When the network topology forwarded to the network controller, it creates the flow rules on the controller and deploys them on the corresponding switches to generate a network for the user. After a network slice successfully implemented, it sends a success message to Request Handling to reply to the user; otherwise, it will send a failure message asking to perform it again.

Side-Chanel Network Event Prediction System – The Bigdata platform holds information such as weather, natural disasters (earthquakes, tsunamis, floods, volcanoes, etc.). The Bigdata platform will evaluate the reliability and stress of links and nodes in a network and apply the observations to the Network Planner based on that data. The Bigdata platform is in our environment connected directly to network planners, and its outputs are used by network planning to find new routes to implement the network slice in our network infrastructure and to generate new rules for SDN-application. The QoS constraint is the bandwidth and delays that obtained through the Web-UI. This QoS information was used to locate the network routes between the switches and is considered a problem with the Constrained Shortest Path. Currently, we deploy algorithms to find the least expensive ways, such as Lagrange Relaxation based Aggregated Cost (LARAC) with delay constraint and implement the algorithms in Path Calculate.

Additionally, due to the information from the Side-Chanel Network Event Prediction System, we can share resources as well as create the Network planner smartly and efficiently. For example, each geographic area will have its main path, which can be affected by impacts from natural disasters. Suppose an earthquake happens that the primary path brake. Our framework can rely on existing resources and information from big data to find alternative ways and split resources into other areas to help maintain uninterrupted connectivity and support repair, maintenance. Besides, our framework can also share resources with other regions to support the maintenance of the main path without interruption. From there, it is possible to find the best route based on the division of resources and associated with disaster information.

IV. PROMISING EXAMPLE

In this section, with our new framework, we will discuss our previous research as promising. In our studies[2]and[3], we performed performance testing of the frameworks. Research[2] is the first system, and[3] is an improved framework based on the concept of time-slices with a method of calculating and discovering pathways. We conducted several experiments for different network settings using mininet to assess the proposed approach, and then we measure the number of accepted requests and the cumulative bandwidth. In these experiments, we vary the scale of the switch network in 30, 50, 100, 150, and 200, and each network link has a bandwidth capacity of 100 ~ 1000 Mbps and delays of 2 ~ 5ms. We assume that up to 1000 flow inputs can behold by a switch. The proposed method evaluated by increasing the network size from 30 to 100, and the number of requests increases up to 2000, while other parameters set.

We evaluate the proposed method by adjusting network size from 30 to 100, and by setting specified parameters, the number of requests increases to 2000. In all instances, our time-slice concept-based method outperforms the original framework. Also, at the very first step of the simulation, when the number of requests is quite low (less than 50), the number of requests accepted between the two methods appears to be the same. Nevertheless, if the number of requests continues to increase, the number of requests our proposed method accepts at the end of the simulation is more than the other slightly up to 1.5 times. To put it another way, the more demands the system receives, the more requests it supports than the original framework. It is because, in the origin system, when a request is met by the network resource, which can not be used for other deals until the end of its duration. Nonetheless, with the time-slice concept-based approach, multiple requests can share the network resource within a similar common time-slice of them. Such tests also show that our time-slice concept-based method provides more cumulative bandwidth than the original framework.

In the future, we expect to design the new framework and experiment with the same environment as in this example, with expected models to be more efficient than the original generation framework. In addition to this, we will extend the experiment with larger real structures provided by the topology-zoo. Also, thanks to the Side-Channel Network Event Prediction System, along with the power of Artificial Intelligence and machine learning techniques, we believe that our proposed structure will be able to devise, provide intelligent network planning. The combination with the big data platform and the AI-based predictive model shows the enormous potential of the framework, and promising the ability to plan networks, maximize performance intelligently more. Moreover, our proposed framework can make full use of the resources of Internet service providers.

V. CONCLUSION

In this paper, we proposed a framework, which supports dynamic network provisioning with AI-enabled path planning using the SDN network. Besides, we use information of network requests of users through the Web-UI and various external data obtained through bigdata platform, which uses the side channel information such as disaster events, maintenance events, and distribution of users. We build an AI-based user request handler in this system that understands user-specific QoS and spatio-temporal requirements to maximize the use of a given network resource. This system also makes use of side-channel information to improve real-time network provisioning. A natural extension of this paper is evaluating the performance of the proposed system with real-world scenarios with multiple user requests including complex spatio-temporal constraints. Also, we are planning to study GNN (Graph Neural Network) and DRL (Deep Reinforcement Learning) based request handlers.

REFERENCES

[1] Van-Quyet Nguyen, Sinh-Ngoc Nguyen, Deokjai Choi, Kyungbaek Kim, "Location-aware Network Provisioning," in Proceeding of APNOMS 18th Conference, Korea, 2017, pp. 239-242.K. Elissa, "Title of paper if known," unpublished.

[2] Huu-Duy Nguyen et al. "Implementation of a Framework for Location-aware Dynamic Network Provisioning." in Proceeding of KIPS Conference, Korea 2018, pp. 133-135.

[3] Huu-Duy Nguyen et al. "Handling Spatio-Temporal QoS Requests for Dynamic Network Provisioning." in Proceedings of the 2019 KICS Korean-Vietnam International Joint Workshop on Communications and Information Sciences, November 20-22, 2019, Hanoi, Vietnam.

[4] M.Huang et al., "Dynamic routing for network throughput Maximization in Software-defined networks," in Proceeding of IEEE INFOCOM, USA, 2016, pp. 1-9.

[5] Mike Jia, W.Liang, M.Huang, Z.Xu and Yu Ma, "Routing Cost Minimization and Throughput Maximization of NFV-enabled Unicasting in Software-defined networks," IEEE Transaction and Network service Management, vol. 15, no. 2, pp. 732-745, June 2018.

[6] J. A. Boyan and M. L. Littman, "Packet routing in dynamically changing networks: A reinforcement learning approach," Advances in neural information processing systems, 1994, pp. 671-678

[7] R. S. Sutton and A. G. Barto, "Reinforcement learning: an introduction", MIT Press, Chp.1-3, 1994.