Computing in Network Research Group
Internet-Draft
Intended status: Informational
Expires: May 5, 2021

P. Liu
H. Yao
L. Geng
China Mobile

November 1, 2020

Differential Computing Resource Reservation
draft-liu-coin-differential-reservation-01

Abstract

Computing in the network may require the embedded computing capability in the network device, such as gateway, switch, etc, and there might be so much distributed computing task in the network. Some new applications like AR/VR, motion control put forward higher demand of network than before, and AI is also considered to be used in the app and network. In order to satisfy the demands, it needs to guarantee both the bandwidth resource and the computing resource which is linked by the network.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 5, 2021.

Table of Contents

1. Overview

From cloud computing to edge computing, computing power is distributed and extends to customers. In the future network and computing integration system, computing power will be distributed in all nodes as ubiquitous endogenous resources. The user's request can be satisfied by calling the nearest node resource, which is no longer limited to a specific node.

The basic topology abstraction of traditional Internet architecture is the end-to-end model: the network is in the middle, the computing is in the periphery, and the host realizes the logical virtual full connection through the network. In the trend of network and computing convergence, computing resource may be embedded in the network. From the perspective of completing users' computing tasks, embedded resources are no longer peer-to-peer relationship, but need to consider the different distances and network conditions.

There are two kinds of ideas of the convergence, One is from the perspective of the network, to realize the perception of computing resources based on the network, so as to perform routing, scheduling, etc. The other is from the perspective of the data center, to realize the perception of network status based on the data center, and apply the scheduling of microservices and other architectures to a wide range network.

Some researching on computing and network convergence has been carried out in standardization organizations, including many network architectures proposed by operators. However, no matter who is the subject of perception, it is to provide better services, so the network and computing will develop in a more refined direction. Based on the perspective of network aware computing resources, this draft analyzes the problems of resource reservation in the trend of network and computing convergence, and put forwards the corresponding reference schemes.

The reservation of traditional network resources is same in an end-to-end path, which means the reserved bandwidth resources will not change from the client to the server, but computing is different. Distributed computing will bring different computing power, and different resources need to be reserved for different nodes. For example, AI algorithm now has a model of step-by-step iteration at multiple nodes. The previous iteration will affect the next calculation results, and the computing resources required for each iteration are not the same. From the perspective of network standard, we hope to regard computing resources as the dimensions to measure network performance, such as the same bandwidth, path, etc., while the traditional technologies of resource reservation have not considered the reservation of computing resources, and have not considered the differentiated resource reservation model.

2. Serial Distributed Computing Model

In the model of computing in the network, the computing resource may be distributed in multiple nodes. A task may be divided into several parts to be executed by multiple nodes, including serial distribution and parallel distribution. Parallel distribution can reserve resources separately. However, in the serial computing model, the calculation process of serial distribution algorithm is sequential, and the results of the previous calculation need to be used in the later calculation, so it will bring the following two characteristics:

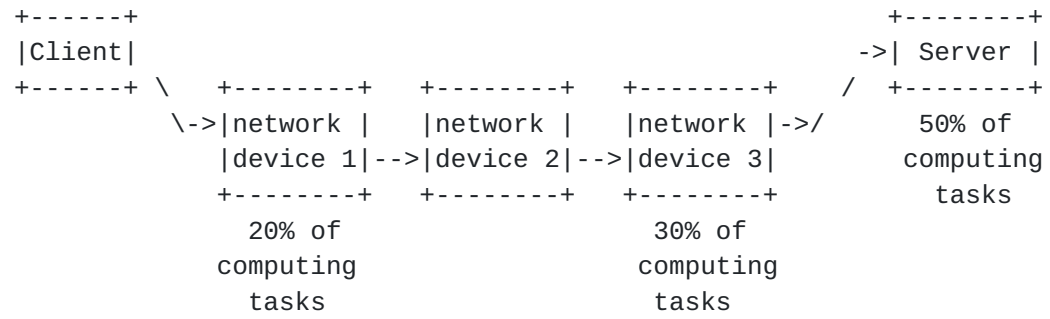
Different computing nodes on the same path need different reserved computing resources.

The bandwidth resources to be reserved maybe different after the previous calculations in the same path.

A typical example is the artificial intelligence algorithm, which involves the multi-layer convolution iterative process and can be completed by multiple computing device in serial. As shown in the figure, 20%, 30% and 50% tasks are calculated on network device 1, 3 and server respectively, and the calculation results of device 1 will affect the subsequent calculation of device 3 and server. Then,

Network device 1, 3 and server need to reserve corresponding computing resources respectively.

Since devices 1 and 3 calculated, the traffic will change after passing through devices 1 and 3, so the bandwidth resources to be reserved are different.



Serial distributed computing model

3. Problems of Existing Protocol

Existing resource reservation protocols work on different layers of network, such as Resource ReSerVation Protocol(RSVP) and Path Computation Element Protocol (PCEP) . RSVP is a traditional protocol, which only focuses on how to initiate the reservation of resources, not the establishment of path. Later, RSVP-TE protocol was developed for MPLS. PCEP was designed to separate the path calculation and path establishment functions of RSVP-TE firstly, which means that the path calculation part before resource reservation can be realized. Therefore, RSVP and PCEP can be used together or separately.

However, thoes protocols have some problem when meets the computing tasks:

First, they do not consider the computing attribute, also can't carry the value of reserved computing resource.

Second, The reserved value of bandwidth resource along the path is unchanged.

It should be noted that we only analyzes the resource reservation protocol in the network field. For the resource reservation of microservice architecture, there may be problems of applying the microservice architecture in the operator network, so it will not be analyzed for the time being.

4. Reference Method

This section provides distributed and centralized resource reservation reference scheme based on the existing protocol of network. It should be noted that for serial distributed computing, we assume that the application side implements the following functions:

The number of steps are involved in the calculation.

The computing proportion of calculation required at each node.

For bandwidth changes after each step of calculation, if this item cannot be implemented, the same bandwidth resources will be reserved by default.

4.1. Distributed Resource Reservation

Distributed resource reservation can be implemented by extending RSVP or RSVP-TE protocol. The server receives the client's service request, calculating the resource reservation strategy and return it. The process is as follows:

1. The client sends the service request, carrying the service requirements and the collected resource status of each node on the path. They will be collected and added to the information that carried by the service request.

2. The server receives the client's service request, then generates the resource reservation strategy for target nodes on the path based on the the service requirements and the resource status of each node, and return the resource reservation strategy to each target node along the path to reserve the resource.

The resource status at least includes the computing resource status such as the catergery of chip, algorithm, etc. It can also includes the network resource status such as bandwidth, delay, etc.

The resource reservation strategy at least includes the computing resource reservation information of target nodes, which is as follows:

1. Determine the serial distributed computing subtasks and computing resources required by each computing subtask based on the service request.

2. Select the target nodes for each computing subtask and generate the computing resources reservation information to inform each target node to reserve resource based on the computing resource status of each node and the computing resources required by each computing subtask.

Moreover, if the bandwidth change after each subtask can be calculated, the resource reservation strategy can also carrying the bandwidth resources reservation information.

It can be realized by defining new object of RSVP or RSVP-TE to reserve different resources in each target nodes. The object can be customized and extended with variable length. For example, redefining a new class num as 30, carries the following message body:

[L = 0, IPv4, 64, IP address1, bandwidth 1, computing resource 1]

[L = 0, IPv4, 64, IP address2, bandwidth 2, computing resource 2]

[L = 0, IPv4, 64, IP address3, bandwidth 3, computing resource 3]

[L = 0, IPv4, 64, IP address4, bandwidth 4, computing resource 4]

.....

It should be noted that the extended object can not only carry the collected resources status of each node in the PATH message, but also return the resource reservation strategy in the RESV message.

4.2. Centralized Resource Reservation

Centralized resource reservation can be realized by the network manager. The manager receives the service request, calculates the network and computing resources needed, and initiates resource reservation configuration for the target nodes along the path.The process is as follows:

The client sends a service request to the network manager.

Network manager selects the path according to the service request and get the resource status of each node on the path.

Network manager generates the resource reservation strategy based on the client's service request and resource status of each node.

Network manager sends resource reservation strategy to target nodes to reserve the resource.

The resource status at least includes the computing resource status. The resource reservation strategy at least includes the computing resource reservation information of each target node. Which are the same with chapter 4.1.

If at least one node in the selected path does not meet the resource reservation requirements, it is necessary to re-select at least one node in the path and get the resource status of the re-selected node until the path meets the requirements of the resource reservation strategy.

4.2.1. PCEP

By adding calculation force resource reservation field to resource reservation object in PECP message, each calculation force flow has a dynamic resource range based on the minimum reserved resource.

Object Type	Label ID	Reserverd Bandwidth	Interface IP Address	In/ Out

PCEP extension

4.2.2. Netconf/Yang

It can also send resource reservation configuration to the target nodes by netconf and defining the Yang structure. The reference Yang module is as follows.

```
module: rs-computing-network
  +--rw rs-computing-network
     +--rw added-device[id]
     |  +--rw service id         string
     |  +--rw user id            string
     |  +--rw bandwitdh            mbps
     |  +--rw computing resource    tbd
     +--rw deleted-device[id]
```

Yang Module

5. Conclusion

The draft proposes a method of differential reservation of computing power and bandwidth resources based on the network protocol. Because the traditional network does not include computing power, the reservation of network resources is the same on the path. This scheme can accurately reserve computing power and network resources for the serial distributed computing services. It also present the reference methods to realize different resource reservation.Of course, there may be more and more appropriate methods to achieve the computing and network resource reservation, which may require more analysis and discussion.

6. Security Considerations

TBD.

7. IANA Considerations

TBD.

8. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/info/rfc5440>.

Authors' Addresses

Peng Liu
China Mobile
Beijing 100053
China

Email: liupengyjy@chinamobile.com

Huijuan Yao
China Mobile
Beijing 100053
China

Email: yaohuijuan@chinamobile.com

Liang Geng
China Mobile
Beijing 100053
China

Email: gengliang@chinamobile.com