On overlay channels choice in Network Powered by Computing Environment

Smeliansky R.L., Stepanov E.P., Borisov A.V., Semenikhin K.V.

The network powered by computing environment is the next step in the development of computing infrastructure after client-server, based on data processing centers. The following problem is under consideration: given a network where the vertices are computing resources (data processing centers, HPC, Edges). Between each pair of such vertices there are several overlay channels (further just channels). The number of such channels available for one computational node is about dozen. It is required to develop a method that will ensure that the quality of service (QoS) of the channel/channels used for transmission meets the requirements of the Service Level Agreement (SLA) for the transmitted data flow. As QoS of a channel, as SLA of a flow, is an ordered set of interrelated variable values such as bandwidth, jitter, loss, delay. The report considers the relationship between these values and proposes a method for dynamic selection of a channel available for flow transmission based on their compliance degree with the flow SLA requirements in the NPC infrastructure.

It is assumed that in each computational node, the parameters characterizing the QoS of each available channel are measured at a certain frequency. The report proposes a channel choice method based on these measurements. When a flow arrives for transmission, it is assumed that its SLA is known. In accordance with this SLA, one channel is selected whose QoS satisfies the requirements of the SLA.

In [1.Network Powered by Computing] the functional architecture of a new generation computing infrastructure that meets the requirements of modern applications has been formulated. Let us briefly repeat its main provisions, following [2.On Fair Traffic allocation and Efficient Utilization of Network Resources based on MARL]. Recently, the landscape of computational infrastructure is in dramatic changes under the pressure of application requirements. The suit of the properties of modern applications can be summarized as follows: distributed self-sufficient, work in real time, elastic, cross-platform, actively interact and synchronize, and are easy to update. The definitions of these terms are in [1.Network Powered by Computing]. For further understanding, it is important to recognize that an application is made up of interrelated components, which we will refer to as application functions. The analysis of requirements of modern application to the computational infrastructure presented in [1.Network Powered by Computing] shows the trend of ubiquitous application deployment. We are moving to the era when data processing resources and data transmission resources form a single space for computing - computational infrastructure. In other words, the time has come for the implementation of the slogan "Network is a Computer". Further we will call such computational infrastructure Network Powered by Computing (NPC). It is should be noted that similar concept was proposed under the name Computing Power Network [3. Computing Power Network: A Survey].

Several versions of Functional Architecture for such new generation of computational infrastructure were proposed. Briefly NPC functional architecture (fig. 1) can be described as following. It consists of data processing (DP) plane, data transmission (DT) plane, data processing control (DPC) plane, data transmission control (DTC) plane, administration, orchestration and management plane (AOM plane). DP plane covers all computational resources of NPC. DT plane is an overlay network over underlying physical network. Actually data transmission plane is data transmission network (DTN). DPC plane is responsible for preparation of the application for execution, planning the placement of application components; calculation of the quality of service (QoS) requirements based on the service level agreement (SLA) specified by the user; generation of DTN control plane instructions for setting up overlay tunnels in accordance with the application function interaction topology of the application. DTC plane is responsible for control and monitoring of DTN. AOM plane orchestrates interactions between application components in accordance with application topology, collects NPC resource consumption statistics by every application component, secures management and administration of NPC.



Figure 1NPC Functional Architecture

In this architecture there is an important point, which is weakly explored so far – the facility that is responsible for the integration of every data processing resource (computational node - CN) with data transmission network (DTN) and external sources of computational service requests. Call this facility NPC router (NPCR). NPCR has several overlay channels with different QoS, some of that can be used to transmit data to the same destination. Suppose that number of such channels available for NPCR is about a dozen. NPCR replaces several devices at once - a task manager, traffic and task router, VPN-gateway, CPE, and supplies the following functionality:

1. distribution of application functions (ApF)/ virtual network functions (VNF) across computational nodes (CN) of DP plane;
2. decision making: is it worth to execute the certain ApF/VNF on the CN connected to this current NPCR or not;
3. forwarding ApF/VNF that was not accepted by the current facility under some reason to other facilities where their computational resources are much more promising from the point of Application execution efficiency as a whole;
4. optimal data traffic routing as between ApF as between corresponding VNF;
5. provision of the transport connection that meets the required Service Level Agreement (SLA).

Some of the NPCRs provide the input for external sources of data, applications, and computational services requests to the NPC resources.

The cited paper [2] has been dedicated only to one problem from the listed above – to the optimal traffic routing by NPCR in overlay DTN of NPC. It should be noted that the solution proposed there for optimal traffic engineering is also applicable in traditional data network. Here we will consider the problems 4 and 5 from listed above. More specific the way to choose the NPCR overlay channel based on the degree of QoS compliance with the requirements of the flow SLA.

The channel choice is based on the estimation of the transmission cost for the given flow with its SLA requirements. To meet the loss requirements, it is assumed that transport agent uses Forward Error Correction (FEC) like methods based on Reed-Solomon codes [4. RS with Fourier Transform]. FEC methods introduce the redundancy that depends on the channel QoS parameters. Redundant data should be taken into consideration while checking bandwidth requirements. It is worth to mention that QoS characteristics of the transport connection are greatly influenced by the used congestion control algorithm. This work presents the approach based on the machine learning methods to estimate the transport connection QoS characteristics depending on the channel quality and congestion control algorithms BBR [5. BBR] and Cubic [6.Cubic] to properly choose the optimal channel.