Technical Disclosure Commons

Defensive Publications Series

June 26, 2018

EFFICIENTLY ALLOCATING RADIO RESOURCES IN THE DOWNLINK DIRECTION TO MEET QUALITY OF SERVICE REQUIREMENTS OF 5G WIRELESS NETWORKS

Akram Sheriff

Gonzalo Salgueiro

Santosh Patil

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Sheriff, Akram; Salgueiro, Gonzalo; and Patil, Santosh, "EFFICIENTLY ALLOCATING RADIO RESOURCES IN THE DOWNLINK DIRECTION TO MEET QUALITY OF SERVICE REQUIREMENTS OF 5G WIRELESS NETWORKS", Technical Disclosure Commons, (June 26, 2018)
https://www.tdcommons.org/dpubs_series/1283

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

EFFICIENTLY ALLOCATING RADIO RESOURCES IN THE DOWNLINK DIRECTION TO MEET QUALITY OF SERVICE REQUIREMENTS OF 5G WIRELESS NETWORKS

AUTHORS:
Akram Sheriff
Gonzalo Salgueiro
Santosh Patil

ABSTRACT

Slice Differentiator (SD) machine learning techniques are provided herein to efficiently allocate resources to meet Quality of Service (QoS) requirements in 5G wireless networks. This enables adaptive training based optimization of QoS weights for different services.

DETAILED DESCRIPTION

Network slicing based 5G systems will face mobility management challenges caused by the potentially ultra-high density of 5G networks combined with high-mobility Internet of Things (IoT) endpoints and high-density end devices. Consequently, new mobility management schemes need to be developed for network slicing based 5G systems to be applied specifically to different network slices to support seamless user experience with better mobility quality, performance, and scalability.

Ultra-Reliable Low Latency Cellular (URLLC) slices, IoT slices, and enhanced Mobile Broadband (eMBB) slices are the three fundamental categories of network slicing in 5G systems. In the vertical network slice domain, the physical computation/storage/radio processing resources in the network infrastructure (by the servers and base stations) and the physical radio resources (in terms of time, frequency, and space) are sliced to form end-to-end vertical slices via properly designed slice pairing functions. The criterion can be different when slicing the radio, Radio Access Network (RAN) and the Core Network (CN). Slice pairing functions are defined to pair the radio, RAN, and CN slices to form end-to-end slices for different services and applications. The mapping between RAN slices and CN slices are not necessarily specific. Depending on this, the QoS of the IoT devices present in the IoT network slice or Premium Mobile Broadband (PMB), eMBB, or URLLC network slice may vary. The techniques described herein provide better QoS for such a device.

In 5G, the mapping between horizontal and vertical network slices is not necessarily one-to-one. In the horizontal domain, the physical resources (in terms of processing, computation, and storage) in the adjacent layers of the network hierarchy are sliced to form horizontal slices of the core cloud, edge cloud, and device layer. A device may operate on multiple network slices. For instance, a 5G smartphone can operate in a vertical slice on an MBB service and a vertical slice on a health care IoT application service (in an IoT network slice) since an IoT application is running on the same smartphone. In this scenario, the QoS requirements for the same 5G device is different with respect to application, latency, and bandwidth requirements on each of the network slices. Adding to this QoS requirement complexity there is tiered interference, for example in URLLC slicing scenarios, which causes communication devices to be more sensitive to time delay and require lower transmission rates than those in other slices. There may be mutual interference between small cells and macrocells, which provide services (e.g., video streaming) for eMBB slices and for IoT slices respectively. Logic is described herein which may be implemented in the Network Slice Selection Function (NSSF).

The slice-specific RAN architecture inherently supports slice on/off, slice-based admission control, and slice-specific load balancing.

When turning on a slice, an access point may allocate radio resources for the slice and enable all radio and network functions associated with the slice, such as the corresponding physical channels. The triggers for turning on a slice at an access point may include the traffic load of that slice exceeding a certain threshold, a number of active devices operating on that slice exceeding a certain threshold, service continuity requirements, and QoS requirements such as low latency, ultra-reliability, etc. Slice-on at an access point (Wi-Fi® or Long-Term Evolution (LTE), radio agnostic) may be triggered by a device or by the network. When triggered by a device, the network may decline the device slice-on request if, for example, the network assesses that the cost/overhead of serving the slice outweighs the service benefit.

Figure 1 below illustrates the User Equipment (UE) requirements index record for every UE identifier (ID).

UE Requirements Index				
Request Time	Capabilities of Network Slice	Failure of Network Service access Previously	Network Service as per requirement	QOS Weights

Figure 1

Figure 2 below illustrates a network slice register for every slice (URLLC, M-IOT, eMMB).

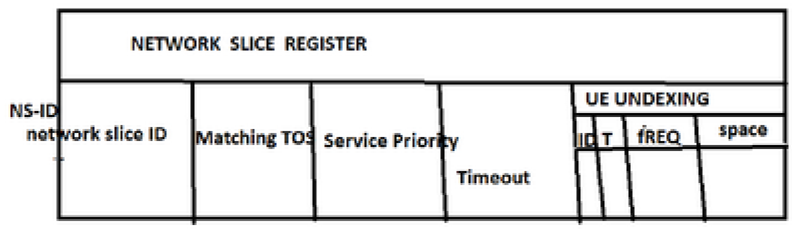

Figure 2

Figure 3 below illustrates example logistic regression based training solution logic.

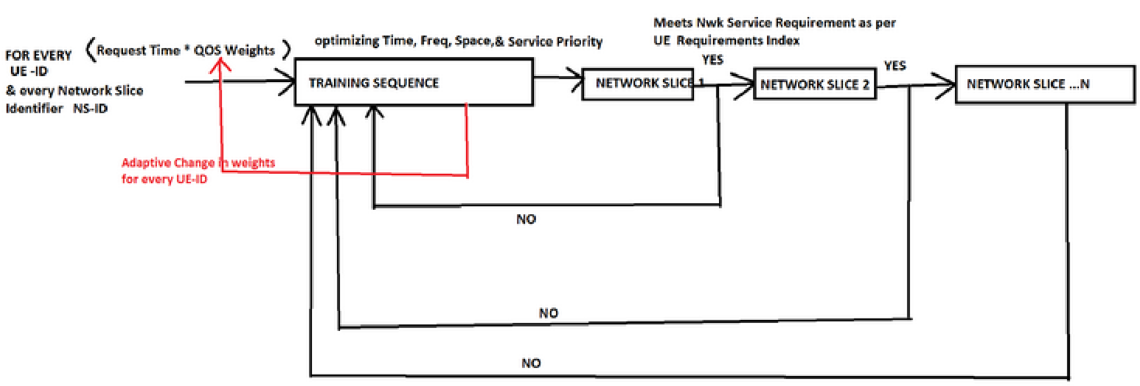

Figure 3

The techniques described herein may use a past historical data based UE ID application profiling dataset. Various UE Registration (UR) types represent different UE capabilities in terms of QoS and data usage. This information is exchanged as part of 3rd Generation Partnership Project (3GPP) UE registration procedure. UE application profiling data may be used as historic information. Based on the past application detected from the UE, the UE may be profiled into delay sensitive application use, data heavy application use, application usage duration, low data application use, low attach frequency device, etc.

The UE application profile may be stored as part of a policy management entity to refer back as part of network slice selection/NSSF. The policy management entity may be central so that this solution can scale beyond control and data plane boundaries.

An example packet flow is provided as follows. First, when the UE originally connects to the network operator or there is an initial registration to the Public Land Mobile Network (PLMN), and there is not enough information for the RAN network connection request, it is routed to an appropriate common control plane function. The UE then sends a network connection request to the wireless access network. Next, the wireless access network forwards the connection request to the slice selection function (Slice/Service Type (SST)).

Second, the slice selection function/NSSF determines which slice the UE can select by accessing the subscriber repository and authenticates the UE. The slice selection function/NSSF checks whether the UE is allowed to access the network. The Slice Differentiator (SD) may be used by the slice selection function/NSSF for a particular UE.

Third, the common control plane function selection determines the common control plane functionality to be connected by considering the information in the network connection request from the UE. In addition, other information in the subscriber repository may be considered (e.g., the subscription information for the UE).

Fourth, based on the response of the common control plane function sent to the RAN node, the UE may attach to the function.

Fifth, the UE completes the registration on the common control plane and managed by mobility management in the common control plane. The RAN establishes a connection between the UE and the common control plane by creating a signaling radio bearer on the Resource Reservation Control (RRC) and establishing a logical channel between the RRC and the common control plane.

Sixth, the slice selection function routes the network connection request of the UE to a dedicated core network control plane function. This request may contain additional information to select a Container Networking Interface (CNI) that is specific to a service type. If the CNI's dedicated control plane function rejects the UE's network connection request, the CNI's dedicated control plane function sends a Network Access Server (NAS) reject message to the UE.

Seventh, a training based solution is applied. Single direct binding of the UE may be accomplished by associating the endpoint UE with various types of network slices. When an endpoint UE is registered to a 5G network, it obtains a unique UE ID which is used for routing the data to the CN. The UE may be associated with a single Network Slice Identifier (NSID) or multiple NSIDs. The former allows the UE to use the NSID or direct binding to specific NSID over specific wireless links, directly associated with the corresponding Network Slice Instance (NSI). QoS is not an issue in a direct binding scenario.

Scenarios involving multiple binding to different network slices may create a QoS problem for the UE entity. In this example, the UE is bound to different slices as per 5G network business requirements. When the UE is connected to multiple NSIDs or different network slices and if different SST values (e.g., 1, 2, or 3) are seen, then the physical UE is composed of multiple logical UEs associated to multiple NSIDs. As illustrated in Figure 1 above, the UE / User Connection Request Field may be used to specify a request time, capabilities of requested network slices, and the SD requested by the UE.

The following formula provides an example algorithm for calculating the machine learning training step size.

ML Training Step Size = Request Time (From Figure 1) * SD requested (From Figure 1) * SST

This may be repeated for every UE ID / NSID pair by reducing the QoS weights (as per "Step Size") until the desired service capability is achieved. After multiple iterations, if the outcome of the training sequence meets the QoS requirements of the UE ID and NS-ID, then training based optimization may be performed (as shown in Figure 3 above). As per one or more requirements, the network service may be stopped for improvements. Correspondingly, that particular NSID may be used to handle the specific traffic from that UE ID meeting the Type of Service (ToS)/Layer 3 requirements.

After multiple iterations, if the outcome of the training does not meet the QOS requirements of the UE ID and NSID, the machine learning training step size may be

Published by Technical Disclosure Commons, 2018

changed again adaptively and the UE traffic may be allowed in the current network slice to which it is assigned until the best possible fit network slice is identified.

Eighth, the authentication and authorization function performs an authentication / slicing authorization process by examining the UE identity with the subscriber repository, which determines whether the UE is authorized to access the slice, and authenticates and allows the UE to attach / connect to the operator's network.

Ninth, the control plane connection may be set for the default or UE specified type of slice.

Tenth, the network slice instance may send a network connection response to the slice selection function. The response information may include the NSS-ID, the temporary UE ID, and other available information that facilitates connection to the slice.

Eleventh, the slice selection function sends a response to the UE that the network connection is accepted. In this response, it contains the temporary UE ID, NSS-ID, and the information to be configured by the UE.

Twelfth, the UE sends a new service session request, the UE requests a session to establish a communication service, which contains the UE ID, the NSS ID, and the information to be configured by the UE.

Thirteenth, the RAN then uses the information received from the common control panel to forward the service session request to the appropriate control plane function that is specific to the core network slice instance.

The techniques described herein provide adaptive training based optimization of QoS weights for different services (types of traffic, IoT slice, URLLC slice, eMBB slice, etc.). The NSSF may dynamically assign network slices for every UE ID based on different category types of UE IDs as defined in the 3GPP LTE UE category (http://www.3gpp.org/keywords-acronyms/1612-ue-category). An SD may also be used in the training model as shown in Figure 3 above for optimizing the QoS weights adaptively. The time, frequency, and spatial domain may be optimized for every UE-ID and network slice functionality combination. There is no tradeoff in performance required even when a UE endpoint has to present in multiple vertical network slice domains. The UE application profile may be based on historical QoS experienced by the UE. The policy management entity may be centralized so that these techniques can scale beyond control and data plane boundaries. The network slices are dynamically selected for a specific time span for every UE ID and NSID so that it meets the QoS requirements of the UE in a 5G Network.

In summary, SD machine learning techniques are provided to efficiently allocate resources to meet QoS requirements in 5G wireless networks. This enables adaptive training based optimization of QoS weights for different services.

Published by Technical Disclosure Commons, 2018