July 01, 2019

SELECTION OF SUITABLE MECHANISMS AND ASSOCIATED PARAMETERS FOR A NETWORK SLICE IN A 5G NETWORK USING MACHINE LEARNING APPROACHES

Mukesh Taneja

Sudhir Jain

SELECTION OF SUITABLE MECHANISMS AND ASSOCIATED PARAMETERS FOR A NETWORK SLICE IN A 5G NETWORK USING MACHINE LEARNING APPROACHES

AUTHORS:
Mukesh Taneja
Sudhir Jain

ABSTRACT

Described herein are techniques for activating the appropriate set of mechanisms in a 5G network and for determining the appropriate set of parameters for those mechanisms. This may help create a network slice in the network with specific constraints (e.g., reliability, latency, throughput etc.). These techniques may also enable adjusting mechanisms and/or associated parameters as the situation changes in the network. Currently, it is difficult to select (and dynamically update if needed) the appropriate set of mechanisms and associated parameters. The methods described herein may be located at a digital network center for private 5G networks or at a Network Data Analytics Function (NWDAF) (i.e., in a 5G core network) for private enterprises or Service Provider (SP) scenarios. Alternatively, these techniques may be distributed across multiple entities (e.g., Multi-access Edge Computing (MEC), NWDAF, digital network center, etc.).

DETAILED DESCRIPTION

Different applications can have different types of requirements in terms of availability, reliability (including Mean Time Between Failures (MTBF)), survival time (which also impacts number of consecutive erroneous packets received at a destination node), update time, latency / jitter, and service bit rate. Some of these are deterministic periodic applications that send data periodically while others are deterministic aperiodic applications that send event-triggered data, such as safety critical information. These also have stringent constraints. Useful traffic is carried for non-deterministic applications as well, and some of these applications also have demanding constraints in terms of availability, latency, MTBF and service bit rate.

There are variety of techniques introduced in 5G systems to improve system performance in terms of reliability, availability, MTBF, latency, and service data rate. There are applications for which a network slice needs to be created in a 5G network with

Published by Technical Disclosure Commons, 2019

specific constraints in terms of reliability, latency, throughput, and other parameters. To do this, the appropriate set of mechanisms must be activated in these networks and the correct set of parameters for these mechanisms must be found. Furthermore, these mechanisms and/or associated parameters may need to be changed as the situation changes in the network. It is difficult to select (and dynamically update if needed) right set of mechanisms and associated parameters.

Figure 1 below illustrates different requirements for a first slice and a second slice.

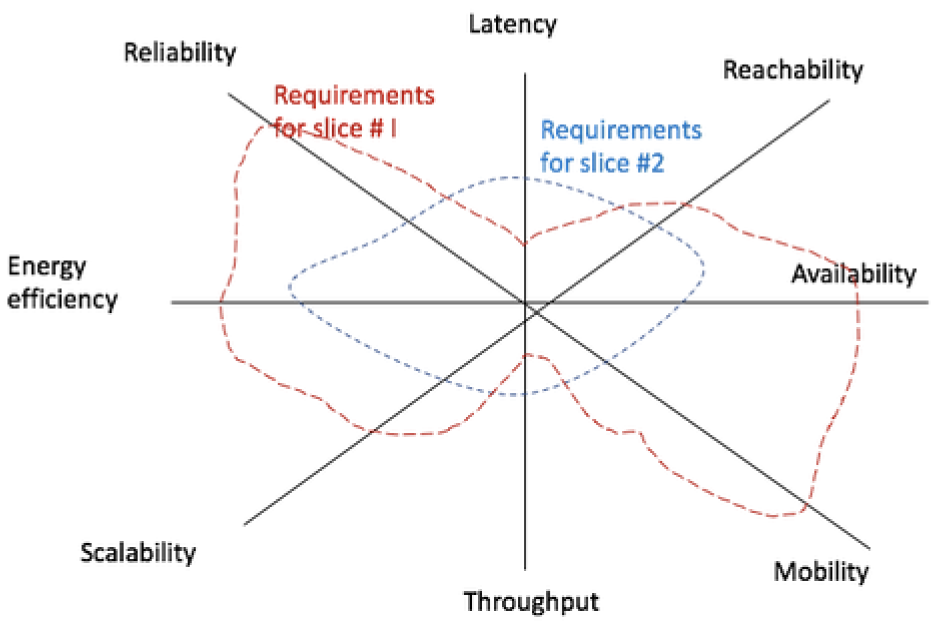

Figure 1

The techniques described herein enable determination of which mechanisms to activate in a 5G network to meet the requirements of the first and second slices. These techniques may also help determine which parameters to use for these mechanisms, and whether to change mechanisms or associated parameters dynamically.

5G systems support various capabilities for improving performance in terms of reliability, latency, throughput, and other parameters. Some of these capabilities are mandatory but the appropriate set of parameters must be determined. Additionally, some capabilities are optional and it must be determined whether to use these and, if so, a suitable set of parameters must be found. Some of these 5G capabilities include the following: inter - Access Point (AP) coordination for Downlink (DL) and Uplink (UL), use of different numerologies (including use of mini-slots), fast UL scheduling, preemption of enhanced

Mobile Broadband (eMBB) for Ultra-Reliable Low-Latency Communication (URLLC), exploitation of diversity techniques, robust coding and modulation selection, replication of packets across different paths or using different resources and multi-connectivity on different Radio Access Technologies (RATs).

There are multiple decisions given a set of performance parameters (and the current network situation) for each slice. One example is whether users at a cell edge have a desired performance for a given slice, and what action to take if they are not getting the desired performance. Another example is which inter-AP coordination technique to use (e.g., Inter-Cell Interference Coordination (ICIC), Enhanced Inter-Cell Interference Coordination (eICIC), DL Coordinated Multipoint (CoMP) / Coordinated Scheduling (CS), DL CoMP / Coordinated Beamforming (CB), DL CoMP/CS and CB, DL CoMP / Joint Processing (JP) / Joint Transmission (JT), DL CoMP / JP / Dynamic Point selection (DPS), UL CoMP/CS, UL CoMP / Joint Reception (JR), etc.) for a slice at a given time and for a given set of performance targets.

Another decision is which gNodeBs and User Equipments (UEs) should use these techniques for a given slice at a given time. Still another decision is at what timescale to share interference information between different gNodeBs for interference coordination to meet performance targets for users belonging to a slice. Yet another decision is which DL and UL Orthogonal Frequency-Division Multiplexing (OFDM) numerologies to use for each 5G UE for a slice to help achieve performance objectives, and whether and how these should be changed dynamically.

Further examples include whether to send packets using different resources over an air-interface (e.g., over different Physical Resource Blocks (PRBs) in 3GPP for ultra-reliable applications) for a given slice, how many times a packet should be replicated over a wireless part of the network, when they should these be replicated (e.g., how to decide activation and de-activation of such time windows), which modulation and coding schemes to use for these packets, and how many times to replicate each packet in each time window.

Another decision is whether to use multiple 5G DL and UL Bandwidth Parts (BWPs) to save energy and achieve other performance objectives for different devices associated with a slice, when to use the BWPs, and how dynamically change these for a slice. Still another decision is which diversity techniques to enable (e.g., for higher reliability), and when.

A specific application running on a UE that belongs to a slice is referred to herein as a 'slice-ue-app'.

For each slice (and each UE in that slice or for an application of a UE that belongs to a slice, i.e. slice-ue-app), several parameters may be collected and analyzed for DL as well as UL, as applicable. One example parameter includes target performance parameters for that slice or slice-ue-app as applicable (e.g., latency, reliability, throughput, scalability, energy efficiency, mobility and other parameters), observed performance violations for that slice or specifically for the slice-ue-app as applicable (e.g., percentage of slice-ue-apps not getting desired latency or reliability, per-slice throughput targets not being achieved, or device mobility or energy requirements not met as expected). Another example parameter is network topology (e.g., location of gNodeBs, UEs, areas where UEs experience interference from neighboring gNodeBs, etc.), traffic load, building topology, and other related parameters.

Still another example parameter includes analysis of 5G wireless DL/UL methods being used (and associated set of parameters) such as over a recent short interval (t,t-e) and over a long time interval (t, t-long) (and/or weighted average in some cases, as applicable). These may include the fraction of time for which each type of 5G OFDM numerology was used (e.g., frequency band, subcarrier spacing, slot length, use of mini-slots, etc.); the fraction of time for which a specific inter-AP coordination mechanism was used (this data may be obtained for each type of inter-AP coordination mechanism that could have been used in the system, such as ICIC, eICIC, DL CoMP/CS, DL CoMP/CB, DL CoMP/CS and CB, DL CoMP/JP/JT, DL CoMP/JP/DPS, UL CoMP/CS, UL CoMP/JR, etc.); the fraction of packets that were replicated over wireless and wireline portions of the network over a given interval to improve the reliability of the system, including the number of packets that were replicated n times for n = 2, 3, ... in these time intervals; the fraction of packets (or bytes) for which a more robust modulation and coding scheme was used instead of usual modulation and coding scheme chosen by the gNodeB scheduler, in a given interval, to improve the reliability of the system; the number of devices (or applications) that the network is able to support for each slice; the fraction of time for which a specific diversity technique was enabled (e.g., for higher reliability); and various other parameters related to some modules discussed above.

Figure 2 below illustrates an example flowchart for performing the techniques described herein. As shown, a variety of parameters are analyzed, the behavior of different modules of the 5G network are learned, and based thereon the modules that need to be activated and their associated parameters are determined. Supervised learning techniques (e.g., Support Vector Machine (SVM) or Support Vector Regression (SVR)) or Reinforcement Learning (RL) techniques may be used for this purpose.

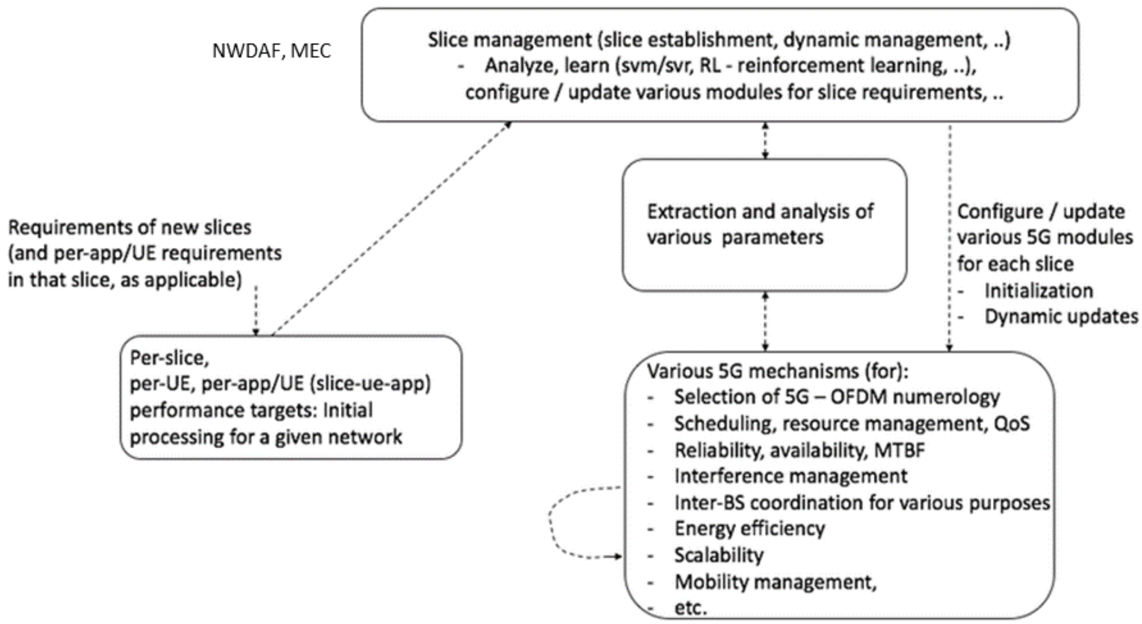

Figure 2

With supervised techniques, many factors may be considered, including (1) per-slice (and per slice-ue-app) performance targets, (2) per-slice (and per-slice-us-app) performance parameters, and (3) a variety of other parameters for various 5G modules as discussed above. This may be assigned a score or classified in certain categories (e.g., highly acceptable, acceptable, average, poor, or not at all acceptable). Techniques like SVM/SVR may be used to train the slice management machine learning module and to select 5G modules (and associated parameters) for new slices.

RL (e.g., Q-learning) may be used by a slice management RL module to meet the performance requirements of the newly introduced slice by identifying the right set of 5G modules (and their associated operating parameters) while continuing to meet the requirements of the existing slices. The system may move from one state to another state, such as from no inter-gNodeB coordination to DL CoMP-CS to DL CoMP-CS-CB to DL-CoMP-CS-CB and UL CoMP-CS, or from one set of 5G OFDM numerology to another or a set of valid numerologies for that scenario. The performance may be analyzed and actions in the Q-learning RL method may be suitably rewarded or penalized.

Once this model is trained (e.g., using RL or SVM/SVR as described above), it may be used as new slices are introduced. As requirements for new slices are provided to the slice management module, this module analyzes those requirements and optionally identifies a subset of techniques that should definitely be used in that scenario. It chooses a performance category (e.g., acceptable or highly acceptable category) depending on the slice requirements and provides those parameters to the machine learning module. The slice management module uses its machine learning models to provide a full set of 5G modules and associated parameters needed to obtain these performance targets. If it is not possible to meet those targets, a set of 5G mechanisms (and associated parameters) may be suggested for a lower performance category.

A network operator may want to initiate multiple slices at approximately the same time. In that case, the slice management module analyzes the requirements of multiple slices, arranges them in a certain order based on one or more policies, and determines 5G mechanisms and associated parameters therefor using the machine learning models described above. If the system is not able to meet the requirements of all the slices due to very demanding or conflicting requirements (e.g., one slice needing a certain kind of inter-AP coordination to be activated while another slice has high requirements for energy efficiency, or for reasons such as multiple mismatches for 5G-OFDM numerologies), it suggests what it can meet for each slice. The operator may accept the suggestions or modify the requirements. As users begin using these slices, the slice management module continues analyzing the behavior of different 5G mechanisms and associated parameters, and dynamically activates / deactivates various mechanisms and their operating parameters as needed. The machine learning model is used to determine whether the system can meet slice requirements more effectively than what it could originally allow in the network and suggests updated performance targets that it can support to the network operator.

In summary, described herein are techniques for activating the appropriate set of mechanisms in a 5G network and for determining the appropriate set of parameters for those mechanisms. This may help create a network slice in the network with specific constraints (e.g., reliability, latency, throughput etc.). These techniques may also enable adjusting mechanisms and/or associated parameters as the situation changes in the network. Currently, it is difficult to select (and dynamically update if needed) the appropriate set of mechanisms and associated parameters. The methods described herein may be located at a digital network center for private 5G networks or at a Network Data Analytics Function (NWDAF) (i.e., in a 5G core network) for private enterprises or Service Provider (SP) scenarios. Alternatively, these techniques may be distributed across multiple entities (e.g., Multi-access Edge Computing (MEC), NWDAF, digital network center, etc.).