July 24, 2019

AUTOMATED SYSTEM TO ENFORCE ENTERPRISE DEFINED SERVICE LEVEL AGREEMENTS IN A 5G SYSTEM

Indermeet Gandhi

Robert Barton

Jerome Henry

AUTOMATED SYSTEM TO ENFORCE ENTERPRISE DEFINED SERVICE LEVEL AGREEMENTS IN A 5G SYSTEM

AUTHORS:
Indermeet Gandhi
Robert Barton
Jerome Henry

ABSTRACT

Techniques are described herein for an automated system to enforce enterprise defined Service Level Agreements (SLAs) in a 5G System (5GS). This enforcement allows for management platform exchanges and permits a Network Data Analytics Function (NWDAF) to define target SLAs for specific applications, regions, and time windows. The NWDAF may additionally dynamically adjust differentiated parameters upward to achieve the target SLA where/when applicable, and downward to optimize the overall SLA cost where/when possible.

DETAILED DESCRIPTION

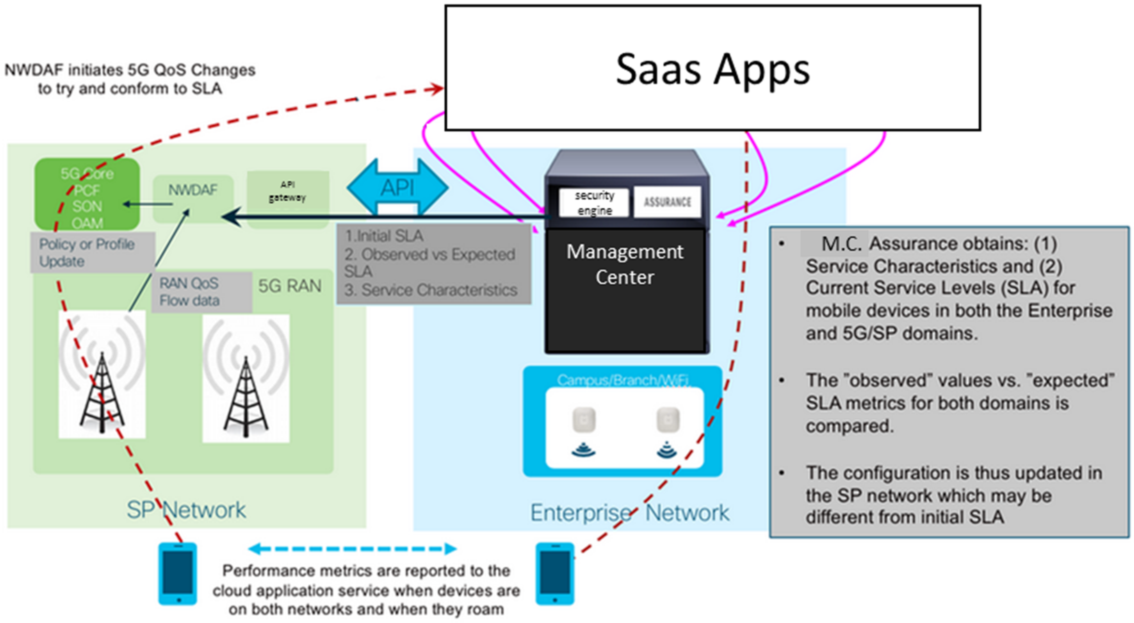
In enterprise networks, centrally-established Quality of Service (QoS) policies and network operation reports provide differentiated services as well as visibility for enterprise network managers as to users, devices, and application Quality of Experience (QoE) on the network. Unfortunately, enterprise network managers have no or very little visibility or control over application performance in cellular networks. Their view is limited to purchasing SLA levels, and letting the Service Provider (SP) translate that SLA into the QoE.

Control of QoE, and consistency of that QoE, is critical in network scenarios where traffic may be load balanced across multiple access technologies (e.g., Wi-Fi® and Long Term Evolution (LTE) or Citizens Broadband Radio Service (CBRS)). Additionally, autonomous and mobile systems with higher bandwidth requirements (e.g., vehicles with camera or Augmented Reality (AR) / Virtual Reality (VR) capabilities, drones, etc.) that move in and out of the Wi-Fi domain into the cellular domain need consistent QoE. The techniques described herein may provide the same level of QoE visibility and enforcement to cellular networks that is currently available to enterprise networks.

Enterprises negotiate SLA targets (against associated costs) with SPs. A management platform may be a single part of the enterprise network. Assurance applications collect application metrics from enterprise elements (e.g., routers, switches, Wireless Local Area Network (WLAN) controllers, etc.) and are able to identify applications using Application Visibility and Control (AVC), which includes Network Based Application Recognition (e.g., NBAR2). Assurance applications also collect service characteristics and SLAs from cloud-based application providers.

By correlating the metrics collected from the cloud-based application providers and from its own network elements, enterprise network managers may suggest improved policy or network configuration changes to minimize the difference between the expected and observed SLAs.

As illustrated in Figure 1 below, the enterprise network exchanges policies with the SP network regarding service characteristics and expected QoE metrics for target/critical applications. This occurs when the devices roam to the cellular network.


NWDAF initiates 5G QoS Changes to try and conform to SLA
Saas Apps
5G Core
PCF
SON
OAM
NWDAF
API gateway
API
Policy or Profile Update
RAN QoS Flow data
5G RAN
1.Initial SLA
2. Observed vs Expected SLA
3. Service Characteristics
security engine
ASSURANCE
Management Center
Campus/Branch/WiFi
SP Network
Enterprise Network
Performance metrics are reported to the cloud application service when devices are on both networks and when they roam

- M.C. Assurance obtains: (1) Service Characteristics and (2) Current Service Levels (SLA) for mobile devices in both the Enterprise and 5G/SP domains.
- The "observed" values vs. "expected" SLA metrics for both domains is compared.
- The configuration is thus updated in the SP network which may be different from initial SLA

Figure 1

For example, a Mean Opinion Score (MOS) may be used as an SLA benchmark for Voice Over Internet Protocol (VOIP) calls. A value above 4.0 (where the maximum theoretical value is 5.0) may be defined as the minimally acceptable score when running the VOIP application. In one such policy, an enterprise network expects the SP to allocate a level of service allowing a MOS score above 4.0 for 95% of enterprise subscriber calls on all days except weekends between 7 AM and 10 PM. In other words, the target SLA is for a specific application, region, and time window(s)).

An initial SLA(s) for the given application service is provided by a management center. The 5G System (5GS) maps the initial SLAs to the 5G QoS parameters to ensure differentiated service performance. However, the initial SLAs may be over-demanding for certain applications and result in high network cost. On the other hand, if the initial SLA provided is low, the 5GS may use the default QoS flow (which has low network cost), but will likely result in poor service performance when the device transitions to 5G.

As described herein, once the initial policy and SLAs are exchanged with the SP network, the management center continues to receive application SLA information for its workforce/devices from the cloud-based application providers. This data is collected under three distinct situations: (1) when the devices are within the enterprise network; (2) when the devices are transitioning from the enterprise network to a cellular network (e.g., the user moved from the enterprise network to the cellular network during the call); and (3) when the devices are within the cellular domain (e.g., call was originated and terminated in the cellular network).

In one example, the management center exchanges the application signatures [Application, Application Identifier, Application Type, Destination IP Address, Destination Port, Protocol] with the 5GS on a real-time basis for the detected application. This enables the 5GS to track and provide the QoE to encrypted applications, applications for which signatures are missing in the 5GS, and for enterprise specific application. This also helps if application connection establishment was originated in the enterprise domain and is being continued as the user moves into the 5G domain. The capability to detect encrypted applications or assign an existing encrypted flow to an application as clients move to the 5GS may be very limited in the 5GS on a standalone basis.

As a result, the management center may obtain metrics indicating adherence to the SLA at all times irrespective of the access medium and the location of its subscribers. The management center monitors the difference between the observed performance metrics and the expected SLA in the enterprise or SP networks as the device moves or roams. Accordingly, the management center triggers these observed differences to the Network

Data Analytics Function (NWDAF) using an Application Programming Interface (API) gateway.

In existing 5G SP systems, the NWDAF collects the QoS flow level network data from 5G Core (5GC) Network Function(s) (NF(s)) including 5G Radio Access Network (RAN), Performance Management (PM) measurements, and QoE data from the User Equipment (UE). The NWDAF may be enhanced to collect metrics relating to adherence to the SLA from the management center. The NWDAF may receive the difference in service experience (expected vs. actual) from the management center and evaluate the current QoS profile. Deriving an optimum QoS profile may be a typical multi-objective optimization and may involve Multiple Criteria Decision Making (MCDM), which means that the single best QoS parameter combination may not exist with respect to all the objectives. Instead, there may exist a set of QoS parameters (or a combination of parameters) which is superior when considering all the objectives together. Deep learning (e.g., deep Q learning) methods at the NWDAF may be used to arrive at the optimal MCDM.

By observing the previously-correlated and analyzed information available in the management center (e.g., the observed metrics while the mobile device was in the cellular network compared to its observed adherence to the SLA in the enterprise network), and the data from the NF(s), the NWDAF may provide statistical information that enables operators to change network deployment and configuration parameters to improve the end-to-end QoS. In effect, the management center may observe geolocation and time slices for which the observed SLA is lower than the expected SLA. Similarly, the management center may observe locations and times where the observed SLA exceeds the expected SLA. This information may be retrieved by the NWDAF. The NWDAF may apply a QoS correction (e.g., bandwidth, QoS Class Identifier (QCI), etc.) to bring the assigned SLA closer to the intended target, while remaining in the overall enterprise budget. This correction may include increasing differentiated service parameters (e.g., bandwidth or QCI / QoS Flow Identifier (QFI) in some locations, for some target time windows), but also lowering QoS parameters, for example outside of SLA time windows, or in locations where lower differentiated service would be acceptable.

The NWDAF thus enables the 5GS to derive a new set of QoS parameters that is different from those derived from the initial SLA, along with the addition of resources in the RAN and in the core network allocated to the slice. In parallel, a management center agent running on the enterprise endpoints may provide SLA information (e.g., "your call in this area will likely not exceed 3.1 for the next 2 miles, do you want to wait to place the call?"; "your navigation system is guiding you toward a low VOIP SLA zone, would you want to be rerouted around the low SLA area?"; etc.). This feedback mechanism allows the enterprise user to optimize SLAs while navigating through inconsistent zones.

In one example, the enterprise controller may share the initial SLA requirement to the 5GS for its users using the 5GS, collect application specific metrics from the application servers for its users (irrespective of location or access type of user), and monitor the difference between observed performance metrics and the expected SLA in the enterprise or SP networks as the device roams. The enterprise controller may further provide application specific signatures to the 5GS for the encrypted application, applications for which signature are missing in the 5GS, or for enterprise specific applications. The NWDAF may dynamically receive the difference in service experience (expected vs. actual) from the management center, derive an optimum QoS profile using deep Q learning to perform MCDM, and orchestrate changes in network deployment and configuration parameters to update the end-to-end QoS based on the trigger from the enterprise controller. An agent on the user equipment may provide the observed SLA directly to the enterprise controller to perform the aforementioned operations or to navigate through the QoS optimized route (not necessarily the shortest route) for its movement in the 5G network.

In summary, techniques are described herein for an automated system to enforce enterprise defined SLAs in a 5GS. This enforcement allows for management platform exchanges and permits a NWDAF to define target SLAs for specific applications, regions, and time windows. The NWDAF may additionally dynamically adjust differentiated parameters upward to achieve the target SLA where/when applicable, and downward to optimize the overall SLA cost where/when possible.