June 20, 2018

SERVICE LEVEL AGREEMENT ENFORCEMENT AND MEASUREMENT OF NETWORK SLICES

Robert Barton

John Graybeal

Bart Brinckman

Jerome Henry

Catherine Rosenberg

Recommended Citation

Barton, Robert; Graybeal, John; Brinckman, Bart; Henry, Jerome; and Rosenberg, Catherine, "SERVICE LEVEL AGREEMENT ENFORCEMENT AND MEASUREMENT OF NETWORK SLICES", Technical Disclosure Commons, (June 20, 2018)
https://www.tdcommons.org/dpubs_series/1266

SERVICE LEVEL AGREEMENT ENFORCEMENT AND MEASUREMENT OF NETWORK SLICES

AUTHORS:
Robert Barton
John Graybeal
Bart Brinckman
Jerome Henry
Catherine Rosenberg

ABSTRACT

The embodiments presented herein disclose a method to predict, score, and ultimately select networks such as radio access networks, as well as heterogeneous network layers such as macro, piccolo, and the like. Technologies that are optimally able to meet service level agreements for different types of network slices, such as enterprise Wi-Fi and SP 3GPP, may also be selected.

DETAILED DESCRIPTION

In a 5G network, each device may potentially require a different type of service, such as services for Internet of Things devices, mobile phones, or surveillance cameras. Applications that depend on low latency, such as augmented reality (AR), virtual reality (VR), or autonomous guided vehicles (AGVs), may require reliable service level agreements (SLAs). Meeting SLA commitments may be difficult, especially in instances where millimeter wave technology is employed, when heavy channel reuse occurs, or in heterogeneous wireless environments that use both licensed and unlicensed spectrum bands. In addition, since carriers may attempt to reuse spectrum as much as possible for different slices, the chances of co-channel interference can increase, thus impacting SLAs; conventional approaches to provisioning and managing SLAs do not take this into account. In addition, performance of wireless networks can vary (especially in the ISM band), meaning adherence to SLAs is not always possible.

Embodiments presented herein consider network slices in a 5G network and compare the radio access technologies (RATs) allocated for each slice, both in the unlicensed and licensed spectrums. Even in a licensed spectrum in which a mobility provider owns the spectrum, not all frequencies are necessarily clean or optimized for a given type of slice due to heavy frequency reuse by the provider. In the ISM band, offering a slice with an SLA is even more difficult. When a 5G client requests to join a network, the carrier may offer a connection proposal based on a band or aggregated carriers. However, it may be generally unknown whether the performance a given band or aggregation can support a particular client or fully meet an SLA for a given customer.

Different types of SLAs may include one or more of a low latency SLA (e.g. able to meet 1 ms latency), a high bandwidth SLA (e.g. able to meet 1 Gbps), a high volume-of-clients SLA, and application-specific SLAs, such as video conferencing SLAs (e.g., able to meet ITU requirements for video conferences), AR/VR SLAs, and AGV SLAs. It may be necessary to ensure that an SLA can be met for different types of slices, and once offered, the SLA must be enforced. Embodiments presented herein profile the performance of different elements of a network for their ability to meet specific SLA types, and then probe the connection that best ensures that an SLA can be enforced at the time of connection proposal.

Before an access point (AP) or e-UTRAN Node B (eNB) is deployed in a given geography, prior statistical data may be helpful for making a zeroth-order selection (such as for starting a call or session, or for deciding which bands/layers/techs will be probed, and in which order). Any number of elements may be used for this pre-deployment evaluation, including prior historical data regarding the location, time of day, user, application type, and the like, and/or crowd-sourced data. Then, when an AP or eNB is deployed, a testing server may determine the performance and reliability of each radio access network and associated wired network. A variety of metrics can be monitored, including application-specific metrics (such as latency, packet loss, and jitter), as well as RF metrics (such as a link quality indicator). These metrics are combined to provide an initial score for each type of SLA, thus ranking the connection's ability to meet the available SLAs. Later, as clients use the radio access network, band metrics continue to be recorded, with breaking thresholds (points in time when deviation from previous measurements switch the probability SLA as below).

The initial score is then compared with minimum requirements of each SLA type and a probability is generated to determine if the radio access network is able to meet each type of SLA, as shown below in Table 1:

SLA Type	**Probability of Meeting the SLA on Band j**
Low Latency (less than 3 ms)	99% - suitable for general low latency
High Volume of Clients	33% - not suitable for high volume of clients
Application: Video Conference	99% - suitable for video applications
Application: AR/VR	88% - suitable for AR/VR applications
Application: AGV	99% - suitable for AGV applications

Table 1

A similar table may be generated for each frequency band. The SLA probability is associated with a time index based on previous measurements to determine the points in time when the SLA is most likely to deviate from the provided value range. When the user equipment is moving, movement is accounted for to anticipate the requirements in other cells. The goal is to offer a connection slice with the highest probability of meeting the SLA over the duration for which the SLA is likely to be maintained. As radio conditions change over time, more and more information may be gathered about a radio access network's ability to meet a certain type of SLA. For example, cyclical occurrences such as the peak hours can also be recorded and anticipated in the future. As more metrics are collected for a given radio access network, it is possible to obtain a clearer picture of the network's ability to meet an SLA for a given time interval.

Statistical modeling such as Bayesian inference may be used to combine previous probabilities with new information in order to make accurate predictions and to proactively account for future time periods. Additionally, user equipment movement can be monitored to evaluate the SLA related-issues relating to handover and user speed. It should be noted that the SLA can be offered on a single radio access network, or on a combination of radio access networks and technologies. For example, while one band may marginally offer a required SLA, combining that band with another radio access network, RAT, or heterogeneous network may increase the probability of meeting the target SLA. Thus, if a single RAN does not reach a satisfactory probability level, the offering may be a combination.

For example, Table 2 depicts several networks that have been examined for a client that is requesting an AR/VR SLA.

Network with RAN/RAT	**Probability of Meeting the AR/VR SLA**	**Deviation Probability**
f1	33%	± 5% over next day
f2	87%	± 12% over next 12 hours
f3	92%	+1%, - 20% over next 5 hours
f4	98% *(this is the band with the highest probability of meeting an AR/VR SLA)*	- 5% over next 4 hours
f5	88%	+ 5 %, - 8% over next 6 hours
. . . . *fn*	53%	±14% over next 8 hours

Table 2

In this example, the network associated with technology *f4* has the best statistical ability to meet the SLA (note that in this simplified example, a single band is sufficient to provide the SLA). Before a proposal is made, a test can also be conducted to confirm that the band's or heterogeneous network's current reliability has not deviated and it is still capable of meeting the SLA.

If a particular connection begins to degrade before the deviation probability point and can no longer meet the SLA, a new connection may be selected. An alternate band or heterogeneous network with the next-highest probability of meeting the SLA may be selected. Furthermore, a negative score for the previous connection may be integrated into the statistical model to provide a revised summary of the first connection's ability to meet this type (e.g., AR/VR) of SLA.

Meeting an SLA for a slice involves more than just one RAT (or combination of RATs). For example, if two customers are participating in a long-distance VR connection, the link between them may experience delay, jitter, and packet loss beyond just the radio access network. Thus, it may be necessary to address the end-to-end SLA in two components: the 5G radio access portion, and the rest of the network. The system may then make a performance measurement of the backhaul network from the eNB to the far-end remote station. While a certain band may appear to meet the SLA from a wireless-only perspective, when combined with the performance metrics of the backhaul network, it can be determined whether the overall SLA will actually be met. For example, an overall SLA may require an end-to-end latency of 10 ms or less between stations. Measurements are taken of the backhaul network and the latency is observed to be 8 ms. Thus, to meet the SLA requirements, a band that supports 2ms or less must be used. Thus, the system may dynamically select a wireless band that is able to meet these requirements in order to satisfy the end-to-end SLA. This approach provides significant advantages for mixed wireless environments that use both licensed spectrum with heavy channel reuse and ISM band access technologies.