

AFFECT-AWARE QoE-TO-SLA MAPPING FOR 5G-ADVANCED AND 6G NETWORK SLICING

Viktor Stoynov

Received on July 1, 2026

Presented by S. Hadjitodorov, Corresponding Member of BAS, on September 29, 2026

Abstract

This paper proposes an Affect-to-Service Level Agreement (A2S) framework for affect-aware network slicing in 5G-Advanced and 6G networks. A2S estimates the user affective state in a normalised valence–arousal space and converts it into a Mean Opinion Score (MOS)-based Quality of Experience (QoE) objective. The resulting QoE requirement is mapped to Service Level Agreement (SLA) parameters: guaranteed bitrate (GBR), packet delay budget (PDB), packet error rate (PER), and scheduling weight. The framework combines affect-to-QoE translation and QoE-to-SLA mapping as a lightweight operator-domain extension around the Network Slice Selection Function (NSSF). A 5G-LENA/ns-3 urban macro evaluation with 150 user equipment units and mixed traffic is used to assess the framework. The results show that A2S improves average MOS from 3.92 to 4.11, increases 10th-percentile MOS from 3.18 to 3.57, and reduces 99th-percentile latency from 59.4 ms to 47.1 ms, while preserving Jain fairness above 0.85.

Key words: network slicing, 5G-Advanced, 6G, affective computing, QoE, SLA orchestration

Introduction. Human-centric service provisioning is becoming a major design direction for 5G-Advanced and 6G, where connectivity is expected to support immersive and context-aware services rather than peak throughput [1, 2]. In this context, network performance should increasingly be interpreted through its effect on perceived service quality, especially for interactive applications. Network slicing

already provides the architectural mechanism for differentiated service treatment in the 5G system, allowing service requirements to be associated with specific slice profiles and resource-management policies [3, 4]. However, current slicing decisions are still based mainly on traffic class, load, radio conditions and static service profiles. They do not explicitly account for the fact that the same technical degradation may be perceived differently depending on the user state.

Affective computing studies indicate that user state can influence perceived quality, tolerance to degradation and the subjective impact of impairments such as delay, stalling or quality oscillation [5, 6]. This does not imply that the network should infer clinical or psychological conditions. Rather, affective information can be interpreted as a compact service-sensitivity signal, indicating higher vulnerability to quality degradation. Russell's circumplex model is suitable for this purpose because it represents affect through valence and arousal rather than unstable discrete emotion labels [7]. The low-valence/high-arousal region can therefore be treated as a higher-sensitivity region, while positive or low-arousal states may tolerate smaller short-term quality fluctuations without changing the service class.

The 5G system already separates slice selection, management and resource enforcement. The NSSF selects suitable slices, while management objects such as slice and service profiles describe latency, throughput and reliability expectations [8, 9]. Recent network-slicing research has also advanced QoE-aware and resource-aware orchestration, but a reproducible bridge between a human-state descriptor and enforceable SLA parameters is still missing [10]. A2S makes that bridge explicit as a cross-layer control chain from valence-arousal through target MOS to 3GPP-compatible GBR, PDB, PER and scheduling weight. Compared with prior emotion-aware mobile cloud concepts [6], the novelty is the operator-side affect-to-QoE-to-SLA mapping and its enforceable output, not a new psychological model or radio scheduler.

Proposed A2S framework. The proposed architecture is shown in Fig. 1. The A2S control path starts from the user equipment/wearable affect sensing block, where a local component estimates valence, arousal and confidence. The edge privacy filter performs data minimisation, confidence checking and smooth-

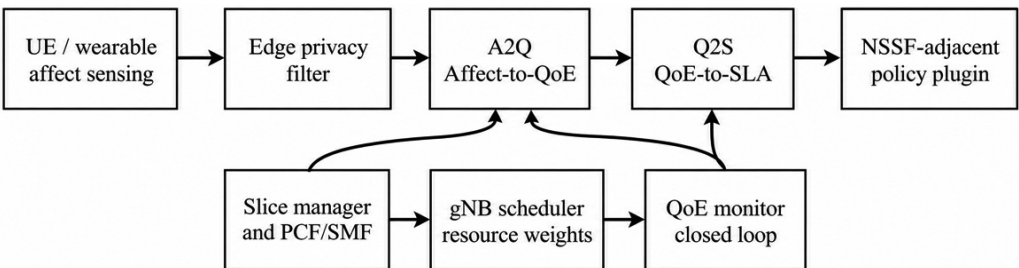


Fig. 1. End-to-end A2S control path from affect sensing to slice policy and closed-loop verification

ing before the descriptor reaches the operator domain. Raw biometric streams are not exported. The affect-to-QoE stage (A2Q) converts the descriptor into a protective target MOS, while the QoE-to-SLA stage (Q2S) transforms it into a service-specific tuple containing GBR, PDB, PER and scheduling weight. The resulting policy is passed to a slice selection policy plugin, extending standard slice-selection logic without creating a new emotional slice type.

The lower part of the architecture represents enforcement and verification. The slice manager and policy/session management functions translate the A2S policy into slice-level and Quality of Service (QoS) flow decisions, including resource guarantees, policy rules and session management parameters. These decisions shape base-station scheduler resource weights, where radio resources are prioritised by service class, affect-derived target and fairness constraint. The QoE monitor tracks bitrate, delay, packet loss and estimated MOS, allowing the controller to compare achieved QoE with the protective target and update the translation and mapping stages when the affect-risk residual remains positive. Thus, A2S acts above existing broadband, low-latency and massive machine-type communication services.

The framework is bidirectional: after the slice tuple is applied, the service-specific QoE model estimates the expected MOS from bitrate, delay budget and packet error rate. The difference between the protective target and predicted MOS forms an affect-risk residual, which can trigger higher scheduling weight or temporary boost.

This separation also limits control plane instability. Full slice reselection is slow and signalling-intensive, whereas scheduling weights, slice resource shares and QoS policy refinements can be updated more frequently. A2S is therefore treated as a near-real-time policy modifier.

Let $e_i(t) = [V_i(t), A_i(t)]$ denote the affect state of user i , where V_i is valence and A_i is arousal, both normalised in the interval $[-1, 1]$. A compact target-setting function is

$$(1) \quad M_i^*(t) = \text{clip} \left[M_0 + \frac{a_1(1 - V_i)}{2} + \frac{a_2(1 + A_i)}{2} + \frac{a_3(1 - V_i)(1 + A_i)}{4}, 1, 5 \right],$$

where $M_i^*(t)$ is the protective target MOS and M_0 is the neutral-state baseline; a_1, a_2, a_3 are tuning coefficients, and $\text{clip}[\dots, 1, 5]$ limits the MOS range. The low-valence/high-arousal term is a conservative control hypothesis rather than a universal psychological law [5–7]. Calibration is service-specific: M_0 is anchored to neutral MOS and a_1, a_2 and a_3 are bounded to reduce the affect-risk residual subject to the MOS range and fairness constraint in (4). Here they remain scenario parameters; deployment requires joint affect-QoE calibration. The target MOS is then mapped into a slice tuple

$$(2) \quad s_i(t) = [\text{GBR}_i(t), \text{PDB}_i(t), \text{PER}_i(t), w_i(t)] = g_k^{-1}(M_i^*(t), c_i),$$

where $s_i(t)$ is the SLA tuple, GBR_i is the guaranteed bitrate, PDB_i is the packet delay budget, PER_i is the packet error rate, w_i is the scheduling weight, $g_k^{-1}(\cdot)$ is the inverse QoE-to-SLA mapping, k denotes the service class, and c_i contains codec, device and session context. For adaptive video, the inverse model searches for minimum bitrate and buffering conditions that satisfy the target MOS [13]. For conversational voice, the ITU-T E-model is inverted to obtain admissible delay and packet-loss values [14]. For cloud gaming, latency and loss sensitivity are mapped to a stricter delay budget and higher scheduling weight [15, 16]. The resource weight used in scheduling is expressed as

$$(3) \quad w_i(t) = 1 + \lambda_w \sigma(\alpha_w A_i(t) - \beta_w V_i(t)),$$

where λ_w controls the maximum affect-driven increase of the scheduling weight, $\sigma(\cdot)$ avoids abrupt priority changes, and α_w and β_w determine the relative influence of arousal and valence, respectively. The formulation increases priority when arousal is high and valence is low, so the scheduler accounts for estimated affective vulnerability in addition to traffic class and radio conditions.

The mapping is service-specific because the same MOS target has different network implications. Adaptive video mainly needs sustainable bitrate and low rebuffering probability; conversational voice is dominated by one-way delay and packet loss; and cloud gaming combines video delivery with strict interactivity constraints. Q2S is therefore a family of inverse QoE models, not a universal conversion rule. The inverse mappings are implemented as service-specific lookup and constraint functions calibrated to the selected QoE models, so higher target MOS values impose stricter delay, loss and bitrate requirements.

The slice-level allocation maximises affect-weighted QoE utility under a total radio resource constraint, minimum slice guarantees and a fairness condition:

$$(4) \quad \max \sum_i w_i U(q_i), \quad \text{s.t.} \quad \sum_j \text{RB}_j \leq \text{RB}_{\text{total}}, \quad \text{RB}_j \geq \text{RB}_{\text{min},j}, \quad J \geq 0.85.$$

Here, q_i is the normalised QoE attainment and J is Jain's fairness index [17]. Equation (4) prevents severe starvation of users who are not in vulnerable affective states, while still allowing a temporary quality margin for users with higher subjective sensitivity.

Simulation setup. The evaluation follows a system-level 5G-LENA/ns-3 scenario calibrated for 3GPP New Radio studies [18, 19]. The topology is an urban macro deployment with three base-station sites, seven cells, 3.5 GHz carrier frequency, 100 MHz bandwidth and 30 kHz subcarrier spacing. It includes 150 user equipment units with heterogeneous traffic. Slice budgets are recomputed every 500 ms, enabling near-real-time adaptation while avoiding unstable sub-frame control (Table 1). Affect traces are smooth valence-arousal trajectories with a 30 s correlation time. Four schemes are compared: static slicing, QoS-aware proportional-fair slicing, load-aware dynamic slicing, and A2S.

T a b l e 1

Main simulation parameters

Parameter	Value
Simulator and radio model	ns-3 with 5G-LENA New Radio module
Topology	3 base-station sites, 7 cells, urban macro
Carrier/bandwidth	3.5 GHz/100 MHz, 30 kHz spacing
Users and duration	150 user equipment units, 300 s, 10 random seeds
Traffic mix	40% adaptive video, 25% voice, 20% cloud gaming, 15% background
Slice set and update period	broadband, low-latency, machine-type; 500 ms control loop

The affect traces are first-order autoregressive trajectories clipped to the normalised valence-arousal domain, providing smooth transitions between neutral and stress-sensitive regions. These trajectories serve as controlled policy inputs for evaluating the network-control mechanism. Validation of affect recognition using real affective datasets such as DEAP [11] is outside the scope of the present study. The traffic mix combines adaptive video, conversational voice, cloud gaming and background traffic to capture service-specific QoE sensitivity to bitrate, delay, packet loss, and short-term variation. Consequently, a single service-agnostic SLA rule would not adequately represent the requirements of all considered services.

The methodology isolates affect awareness by comparing fixed static slicing, conventional QoS-aware slicing, and a strong affect-blind load-aware baseline reacting to demand and radio state. The load-aware-to-A2S gap is therefore interpreted descriptively as the affect-to-SLA contribution. Results aggregate 10 seeds. In this context, paired seed-level outputs were not retained, so inferential significance is not claimed.

Results and discussion. The quantitative results in Table 2 show that affect-aware control improves both average and lower-tail experience. A2S achieves a mean MOS of 4.11, compared with 3.92 for the load-aware baseline and 3.61 for static slicing. The strongest improvement is in the 10th-percentile MOS, where A2S reaches 3.57, indicating that the method mainly protects low-QoE cases. The cumulative distribution function (CDF) in Fig. 2 confirms this behaviour by shifting the A2S curve to the right and compressing its low-MOS tail.

The latency gains are also relevant. The 99th-percentile latency decreases from 59.4 ms under load-aware slicing to 47.1 ms under A2S because affect-sensitive voice and gaming sessions receive a stricter delay budget when valence is low and arousal is high. The fairness index remains 0.87, above the 0.85 threshold, showing that A2S does not simply over-prioritise stressed users but balances protection and fairness. The additional signalling load is 18.6 kbps per base station, and the per-user inference time is 0.34 ms, compatible with the 500 ms orchestra-

T a b l e 2

Performance comparison of static, QoS-aware, load-aware and A2S slicing

Metric	Static	QoS-aware	Load-aware	A2S
Mean MOS	3.61	3.78	3.92	4.11
10th-percentile MOS	2.74	3.05	3.18	3.57
QoE variation, std. dev.	0.58	0.49	0.48	0.40
5th-percentile throughput, Mbps	18.9	20.0	21.3	23.5
Average latency, ms	31.7	27.9	24.8	21.6
99th-percentile latency, ms	84.0	67.3	59.4	47.1
Jain fairness index	0.82	0.91	0.83	0.87
Additional control signalling, kbps/base station	0	4.2	12.8	18.6

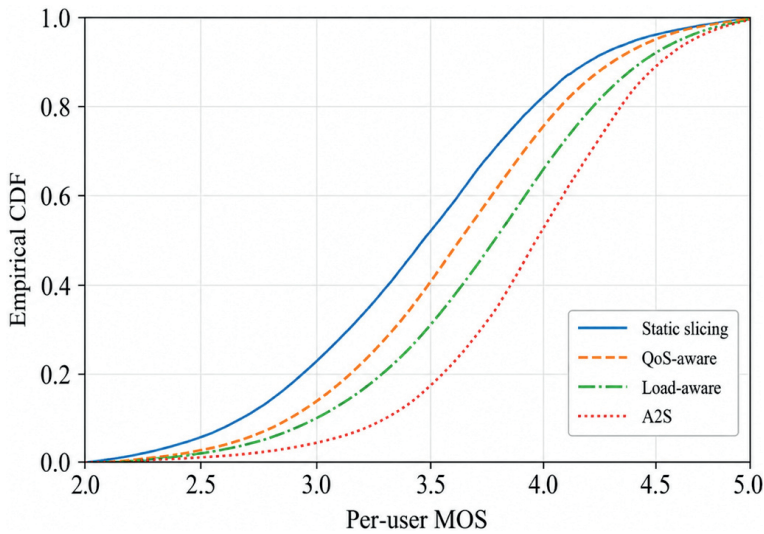


Fig. 2. Empirical MOS distribution for static, QoS-aware, load-aware and A2S slicing

tion cycle. From a QoE perspective, the important effect is the redistribution of quality rather than a uniform resource increase. Compared with the load-aware baseline, the average MOS gain is moderate, but the 10th-percentile MOS gain is larger, which means that the proposed mechanism mainly improves sessions close to unacceptable quality. This behaviour is consistent with the design objective of A2S – affect information is used to add protection where degradation is expected to be perceived more severely, not to raise all sessions equally.

The values indicate a trade-off. A2S adds more control signalling than the load-aware baseline, but the overhead remains small relative to user-plane traffic and is justified by lower-tail improvement. Reduced affect estimation accuracy produced graceful simulated degradation because the controller falls back to service/load information. Finally, this is a robustness check, not real-user affect validation.

The key result is the substantial improvement in lower-tail performance, while the increase in mean MOS is more moderate. Average values may hide short impairments affecting interactive users. By increasing 10th-percentile MOS and reducing 99th-percentile latency, A2S improves the most perception-sensitive operating points.

Deployment and validation outlook. A2S can be deployed as an opt-in operator-domain policy extension with edge-assisted inference. Raw biometric streams need not leave the device or edge, while only a compact descriptor reaches the slice-policy functions. However, the descriptor itself remains sensitive. Therefore, deployment should address spoofing, replay, inference manipulation, profiling and policy abuse. This can be achieved through authenticated and integrity-protected transport, freshness and provenance checks, confidence gating, least-privilege access, short retention periods and audit logs. In addition, missing consent or insufficient trust should trigger affect-blind QoS operation. Validation and lifecycle control can also be aligned with zero-touch principles [20]. The present results demonstrate the feasibility of A2S from a network-control perspective under simulation conditions. Validation of affect recognition using real-user data remains outside the scope of the current study. Future work will therefore incorporate joint affect-network datasets and UERANSIM/free5GC experiments, while retaining seed-level outputs for confidence intervals and paired statistical tests.

Conclusion. This paper presented A2S, a compact framework for affect-aware network slicing in 5G-Advanced and 6G networks. It links valence-arousal representation, target QoE and slice-level SLA parameters through an interpretable control chain. By using target MOS as an intermediate variable, the method avoids a direct and weakly justified mapping from affect to radio resources. The results primarily show lower-tail protection: 10th-percentile MOS increases and tail latency decreases, while the gain in mean MOS is more moderate and fairness remains above the imposed threshold. The main contribution is a formal cross-layer bridge between affect state, QoE target, and enforceable slice policy.

REFERENCES

- [1] ITU-R (2023) Framework and overall objectives of the future development of IMT for 2030 and beyond, Recomm. ITU-R M.2160-0.
- [2] DANG S., O. AMIN, B. SHIHADA et al. (2020) What should 6G be?, *Nat. Electron.*, **3**, 20–29.
- [3] 3GPP (2024) System architecture for the 5G System (5GS), Technical Specification TS 23.501, Release 18.
- [4] 3GPP (2024) Management and orchestration; concepts, use cases and requirements, Technical Specification TS 28.530.
- [5] HOSSAIN M. S., G. MUHAMMAD (2018) Emotion-aware connected healthcare big data towards 5G, *IEEE Internet Things J.*, **5**(4), 2399–2406.

- [6] CHEN M., Y. ZHANG, Y. LI et al. (2015) EMC: emotion-aware mobile cloud computing in 5G, *IEEE Netw.*, **29**(2), 32–38.
- [7] RUSSELL J. A. (1980) A circumplex model of affect, *J. Pers. Soc. Psychol.*, **39**(6), 1161–1178.
- [8] 3GPP (2024) 5G System; Network Slice Selection Services; Stage 3, Technical Specification TS 29.531.
- [9] 3GPP (2024) 5G Network Resource Model (NRM); Stage 2 and Stage 3, Technical Specification TS 28.541.
- [10] EBRAHIMI S., F. BOUALI, O. C. L. HAAS (2024) Resource management from single-domain 5G to end-to-end 6G network slicing: a survey, *IEEE Commun. Surv. Tutor.*, **26**(4), 2836–2866.
- [11] KOELSTRA S., C. MUHL, M. SOLEYMANI et al. (2012) DEAP: a database for emotion analysis using physiological signals, *IEEE Trans. Affect. Comput.*, **3**(1), 18–31.
- [12] ZHAO M., F. ADIB, D. KATABI (2016) Emotion recognition using wireless signals, *Proc. ACM MobiCom*, 95–108.
- [13] ITU-T (2017) Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services, *Recomm. ITU-T P.1203*.
- [14] ITU-T (2015) The E-model: a computational model for use in transmission planning, *Recomm. ITU-T G.107*.
- [15] JARSCHER M., D. SCHLOSSER, S. SCHEURING et al. (2013) Gaming in the clouds: QoE and the users' perspective, *Math. Comput. Model.*, **57**(11–12), 2883–2894.
- [16] SABET S. S., S. SCHMIDT, S. ZADTOOTAGHAI et al. (2020) Delay sensitivity classification of cloud gaming content, *Proc. ACM MMSys Workshop*, 25–30.
- [17] JAIN R., D.-M. CHIU, W. R. HAWK (1984) A quantitative measure of fairness and discrimination for resource allocation in shared computer systems, *DEC Research Report TR-301*.
- [18] PATRICIELLO N., S. LAGEN, B. BOJOVIC et al. (2019) An E2E simulator for 5G NR networks, *Simul. Model. Pract. Theory*, **96**, 101933.
- [19] KOUTLIA K., B. BOJOVIC, Z. ALI et al. (2022) Calibration of the 5G-LENA system level simulator in 3GPP reference scenarios, *Simul. Model. Pract. Theory*, **119**, 102580.
- [20] ETSI (2019) Zero-touch network and service management (ZSM); reference architecture, *ETSI GS ZSM 002*.

*Faculty of Telecommunications, Technical University of Sofia, 8 Kliment Ohridski Blvd,
1756 Sofia, Bulgaria*
e-mail: vstoynov@tu-sofia.bg