iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

# APPLICATION-TRIGGERED QOS ACCELERATION VIA NEF: AN END-TO-END 5G STANDALONE PROOF-OF-CONCEPT

**Rami Saied AlZaben**
School of Engineering and Applied Science,
The George Washington University (GWU), USA.
rami.alzaben@gwmail.gwu.edu

## ABSTRACT
The intensive development of latency-sensitive and interactive digital services cloud gaming, extended reality (XR), real-time user-generated media, etc. have disclosed the constraints of fixed Quality-of-Service (QoS) settings in commercial 5G Standalone (SA) networks. Although 5G has flexible QoS constructs, the lack of any standardized, application-driven control mechanisms does not allow services to dynamically request and enforce network performance assurances in accordance with real-time Quality of Experience (QoE) needs. The 3GPP Network Exposure Function (NEF) bridges this gap by providing application functions (AFs) the ability to expose network capabilities and policy control interfaces to authorized applications, giving applications the ability to directly impact session behavior using standardized APIs.
The paper introduces an operator-grade Proof-of-Concept (PoC) which confirms the practical usefulness of NEF-based QoS exposure and acceleration in a live commercial 5G SA network. The PoC has shown that it supports an end-to-end control loop whereby an application can make a dynamic request to the NEF to increase QoS which is then acted on by the Policy Control Function (PCF) and then session updates which are enforced by the Session Management Function (SMF) and the User Plane Function (UPF). The analysis compares the baseline best-effort sessions with NEF-assisted sessions in the context of realistic congestion conditions, which confirms the effective adjustment of the QoS flows, the parameters of prioritizing and traffic treatment in real-time.
The findings validate the fact that NEF-enabled application-driven QoS control can be implemented, be standards-compliant and functionally workable in a commercial 5G environment. Contributing to one of the first empirical validation of NEF-based QoS exposure outside of theoretical models and laboratory testbeds this work has provided concrete signaling evidence and quantifiable changes in behavior at the session level. The results demonstrate NEF as the key facilitator of on-demand QoS, service differentiation, and convergence with analytics-driven automation models in the future, and application-aware network control is the core facility of next-generation 5G services.

## 1. INTRODUCTION
The growth of latency-sensitive and interactive services, including cloud gaming, extended reality (XR), real-time communications, and user-generated live media have put high Quality of Experience (QoE) demands on commercial 5G Standalone (SA) networks. Such services are extremely time sensitive to delay and jitter, and short-lived network outbursts, and their perceived quality usually depends not on the average throughput, but on the temporal stability of network behavior (Tan et al., 2022; Kao and Wu, 2023). Although 5G provides flexible Quality-of-Service (QoS) structures via QoS Flows and standardized policy control, deployed instances of 5G persist in using primarily static or network-centric Quality-of-Service configurations that are not well adapted to the dynamic and context-specific nature of current interactive traffic.
In existing 5Gs, the QoS decisions are mostly implemented at the core network by a set of policies, profile of subscribers or coarse rules of traffic classification. In as much as these mechanisms offer baseline service differentiation, they have little application specific functions on their own to convey real-time performance intent or to react network treatment to rapidly varying service conditions. Consequently, interactive services often suffer temporary performance deterioration when demand is medium to high even when unused or reassignable network resources are available (Ji et al., 2025). Such an uncoupling between application-level needs and network-level management is a fundamental weakness of the realistic implementation of service-aware 5G performance guarantees.
The 3GPP Network Exposure Function (NEF) was proposed as a standardized system of securely exposing selected network capabilities, network events and policy control interfaces to authorized external parties, known as Application Functions (AFs) to fill this gap (3GPP, 2025). The NEF allows applications to request QoS

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

treatment, subscribe to network events, and affect the behavior of a session in a policy-compliant and controlled way, via the NEF service-based APIs. In contrast to analytics-based automation systems that indirectly derive network behaviors, NEF can explicitly interact with the 5G Core via application-driven interactions that can be used to explicitly indicate the desired performance considerations, and get predictable responses on the network (Chen, 2024).

In spite of it being the central component of the 5G service based architecture, current studies on NEF have had a general tendency to describe the architecture or API requirements or conceptual design examples. The empirical data available in the literature about NEF functionality is limited with most published research assessing components of NEF functionality using simulation or emulation environments or by discussing high-level design aspects; there is little empirical evidence on how NEF-based QoS exposure can be applied to live commercial networks. Specifically, the number of operator grade validations that follow the entire control loop, i.e. AF request and NEF mediation and Policy Control Function (PCF) decision making to enforcement by the Session Management Function (SMF) and User Plane Function (UPF), has been shown to be limited under realistic traffic and congestion scenarios. This unreal-life support prevents objective evaluation of the operational maturity of NEF as well as its efficiency in enhancing the service quality perceived by the users.

The gap that this paper fills is related to providing a Proof-of-Concept (PoC) implementation of NEF-based QoS exposure and acceleration implemented on a commercial 5G SA network. This research shows an end-to-end, standards-compliant, control loop where an application is dynamically requesting QoS enhancement to the NEF, which causes policy changes in the PCF, and real-time session modification is implemented by the SMF and UPF. The work introduces a solid signaling evidence and session-level verification of application-based QoS control in practice by comparing the best-effort baseline sessions with NEF-aided sessions.

Threefold contributions in this paper can be made. First, it provides one of the first empirical evidences of NEF-based QoS exposure in a live operator environment, overcoming the theoretical and laboratory-based testing. Second, it confirms the practicality of application-based QoS acceleration based on standardized 3GPP interfaces, which proves that the control facilitated by NEF may be safely incorporated into current 5G Core policy models. Third, it makes NEF a cornerstone enabler of on-demand QoS and service differentiation, to be used alongside analytics-based automation strategies, and set the stage of additional responsive and application-aware network control in 5G deployments to come.

The rest of this paper is structured in the following way. Section 2 evaluates other work in related areas in terms of applications conscious QoS control as well as NEF based exposure mechanisms. Section 3 explains the PoC methodology and system architecture. The fourth section gives the experimental findings and evidence of verification. The implications, limitations and practical relevance of the findings are discussed in section 5. Section 6 summarizes the paper and provides a future work direction.

## RELATED WORK

### 2.1 QoS and QoE Control based on Application-Awareness 5G Networks.

The optimization of Quality-of-Service (QoS) and Quality-of-Experience (QoE) in mobile networks are not new research areas, especially with the new bandwidth-demanding and latency-sensitive services. QoS in 5G networks is achieved in QoS Flows, whereby traffic can be differentiated based on standardized values, including packet delay budget, packet error rate, and priority levels (Ksentini and Frangoudis, 2020). Previous research has shown that interactive applications (such as cloud gaming, real-time video) are particularly vulnerable to temporary changes in latency and jitter, both of which have a direct impact on user experience regardless of whether average throughput is adequate (Tan et al., 2022).

The current methodologies of QoE optimization have mostly been based on network-based mechanisms such as traffic classification, policy provisioning with static policies, and optimization frameworks based on analytics. In as much as these methods can enhance aggregate network performance they usually do not have direct knowledge of the intent of the application and thus are slow in responding to rapid service demand variations. According to Ji et al. (2025), traditionally enforced QoS is ineffective due to the lack of feedback on application performance in real time, especially in dynamically congested conditions in which the degradation of performance happens suddenly and unpredictably.

### 2.2 Network Exposure Function (NEF) Newton Architecture 5G Service-Based Architecture.

Network Exposure Function (NEF) is one of the core components 3GPP added to the 5G service-based architecture to allow the network capabilities to be exposed in controlled ways to external parties (3GPP, 2025). The NEF grants network access to authorized Application Functions (AFs) via standardized APIs, enabling them to subscribe to events and request selected policy actions, such as QoS-related services. This architecture is a

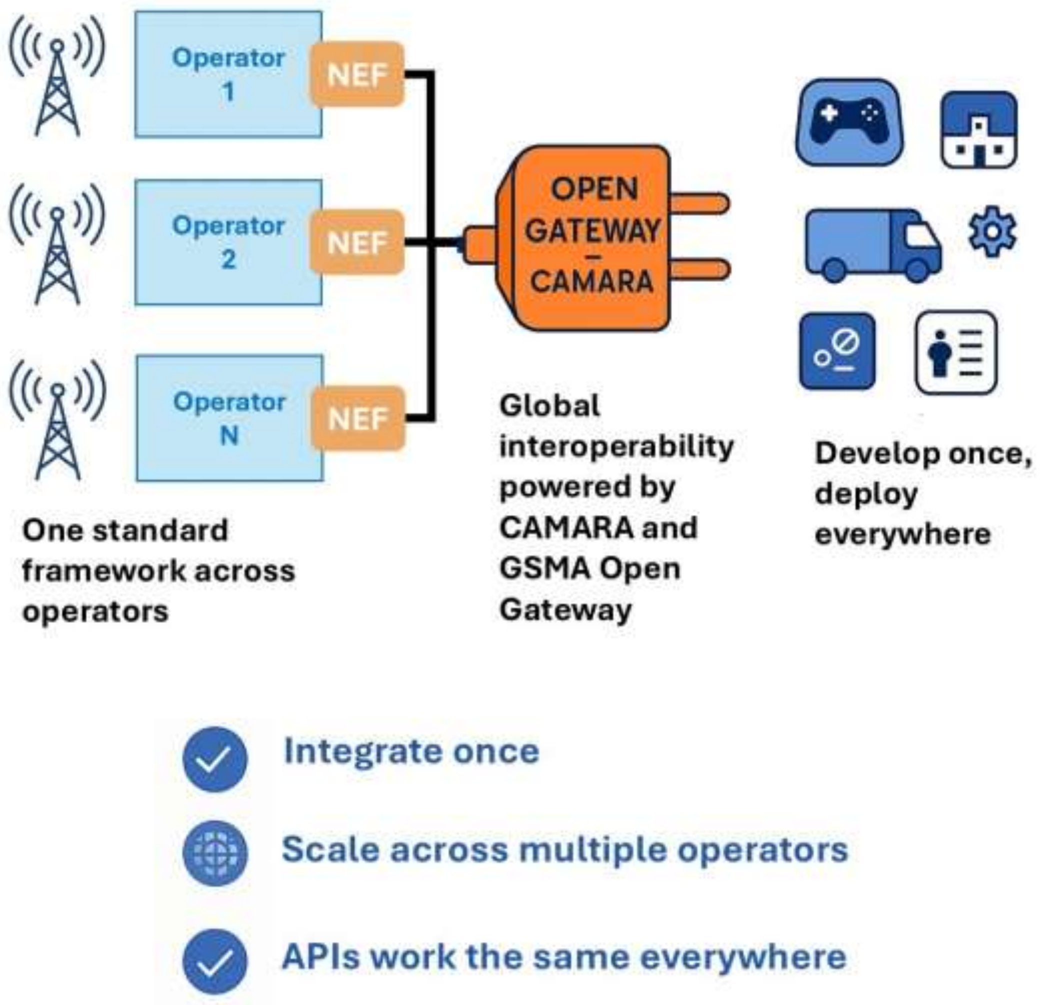
iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

radical change, as compared to closed and operator-only control models, to a programmable network paradigm where applications can affect the behavior of sessions in a secure and policy-constrained way (Chen, 2024).
A number of works have characterized the architecture of NEF and its capability to facilitate vertical-intensive services, such as gaming, automobile, and industrial automation. Nonetheless, the bulk of this is still in the description format, which is concentrating on API definitions and theoretical potentials and not on operational verification. Evaluations are commonly conducted using simulated traffic models or sandbox environments, which do not fully capture the complexity and constraints of commercial 5G deployments (Ksentini and Frangoudis, 2020).

*TABLE 1. QoS-Related NEF APIs Defined in 3GPP TS 29.522*

| NEF API | Purpose | QoS Relevance |
|---|---|---|
| AsSessionWithQoS | Establish application-specific QoS authorization session | Enables AF to request QoS treatment |
| AsSessionUpdate | Modify existing QoS authorization | Dynamic QoS changes during session |
| AsSessionDelete | Release QoS authorization | Terminates application-driven QoS |
| EventExposure | Subscribe to network events | Allows AF awareness of session state |
| MonitoringEvent | Monitor UE/session conditions | Optional trigger for QoS request |

*FIGURE 1: GSMA Open Gateway / CAMARA interoperability model*



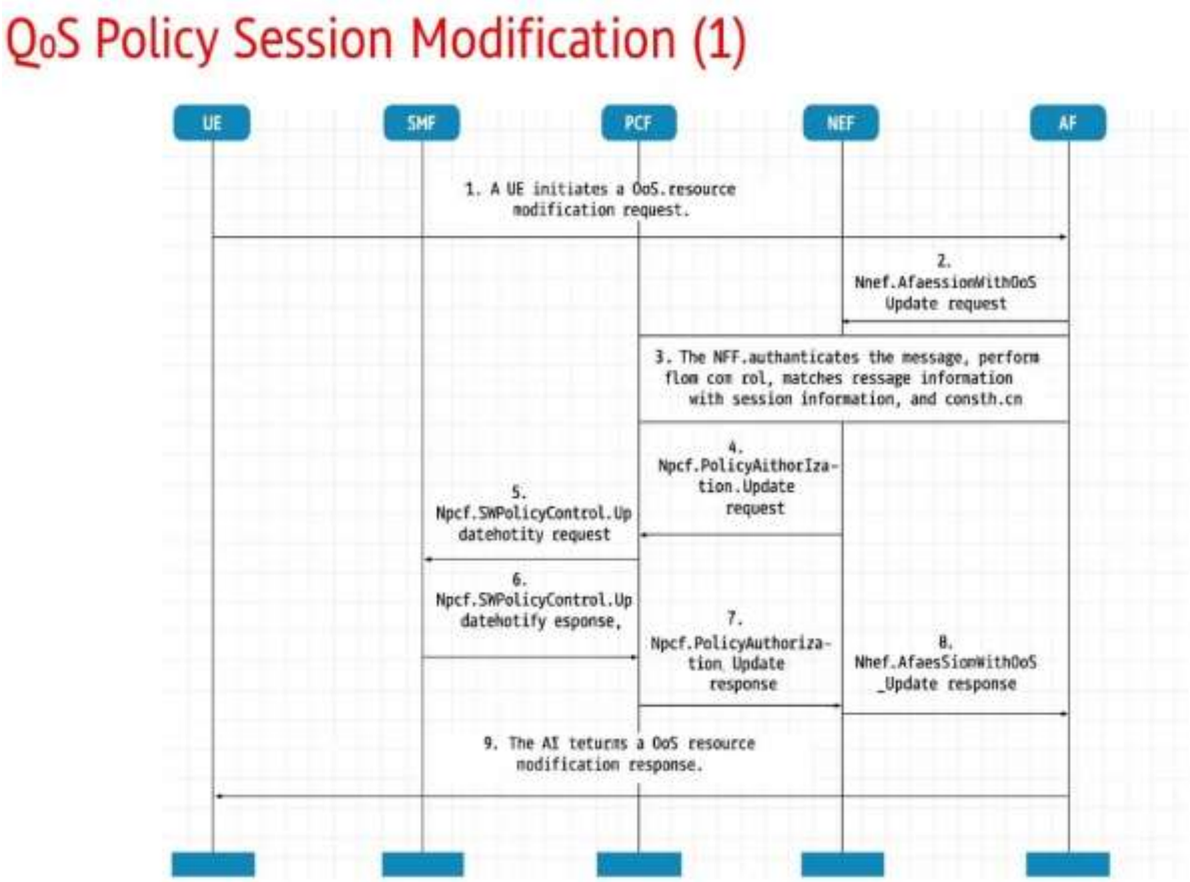
iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

*Ecosystem diagram showing multi-operator exposure via CAMARA/GSMA Open Gateway, enabling "develop once, deploy everywhere" API consumption across different operators' NEF domains.*

**2.3 Application-Driven QoS Exposure and Policy Control**

The idea of application-driven QoS, in which applications explicitly indicate their performance demands to the network, has been investigated recently. This paradigm is opposed to analytics-based applications which deduce the need of an application indirectly by monitoring traffic. It has been demonstrated that application-driven QoS can provide a quicker response time and is more deterministic, especially with short-lived or bursty traffic flows (Tan et al., 2022). This interaction can be implemented in the 5G Core as the AF-NEF-PCF control chain, which allows modifying policy decisions in response to the application requests and being in compliance with the operator-specific constraints (3GPP, 2025). In spite of these, the practical implementations of application-driven QoS are still limited. The majority of studies to date do not go further to illustrate how the AF-induced requests are transmitted over Policy Control Function (PCF) and Session Management Function (SMF) to generate enforceable alterations within the user plane. Consequently, there exist scanty empirical observations that NEF-based QoS exposure can be successfully provided end-to-end in live networks without affecting stability or policy integrity (Ji et al., 2025).

***FIGURE 2: QoS Policy Session Modification (sequence diagram)***




*Message sequence diagram for QoS policy session modification across UE, SMF, PCF, NEF, and AF, illustrating request/response exchanges and policy update propagation.*

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

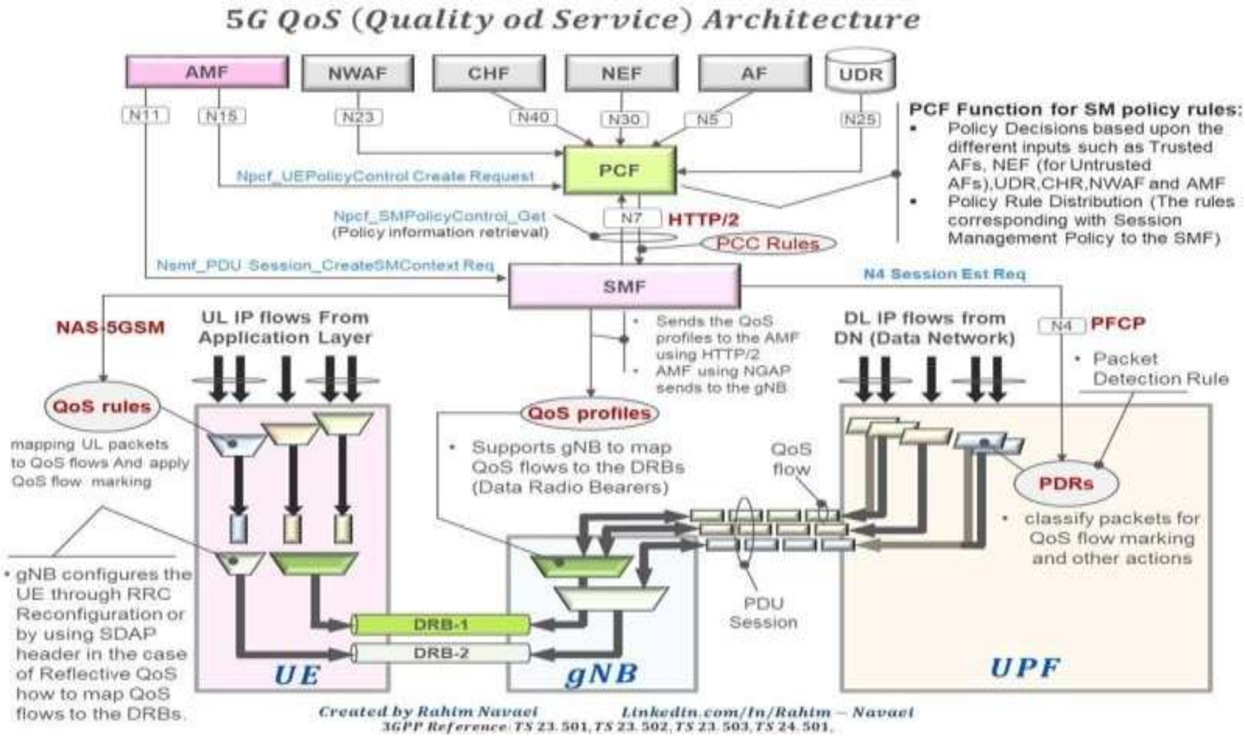
*FIGURE 3: 5G core functions (isometric) with PCF emphasis*


*Isometric 5G core functional map showing user-plane vs control-plane separation, with PCF highlighted to emphasize its central role in policy control and QoS decisioning across SBA interfaces.*

**2.4 Research Gap and Positioning of This Study**

In a nutshell, existing literature confirms the theoretical significance of application-aware QoS control and shows NEF as one of the enablers of the 5G architecture. But there is a definite shortage of operator grade Proof-of-Concept studies confirming NEF-based QoS exposure in practical environments. The literature provides little evidence of the full chain of signaling and enforcement, including application request, NEF mediation and policy decision, and session-level enforcement, in commercial networks that can be verified. This paper fills this gap by offering an empirical confirmation of NEF-based QoS exposure and acceleration in a live environment of 5G SA. With its emphasis on the concrete signaling behavior, session modification, and policy enforcement results, the work adds to the body of the analytics-driven QoE optimization research and provides valuable input into the potential and the readiness of the application-directed network control.

*FIGURE 4: 5G QoS architecture: policy control, QoS rules, and enforcement*



iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

*Functional architecture showing PCF policy/QoS rule generation, SMF session management, and UPF enforcement mechanisms (e.g., PDR/QoS flow mapping) for uplink and downlink control.*

## 3. SYSTEM ARCHITECTURE AND METHODOLOGY

### 3.1 NEF-Based Application-Driven QoS Architecture

The Proof-of-Concept (PoC) in this paper is grounded on a fully standards-compliant 5G Standalone (SA) architecture that employs an application-based Quality-of-Service (QoS) control using the Network Exposure Function (NEF) as defined in the 3GPP service-based architecture (3GPP, 2025). Through the architecture, authorized Application Functions (AFs) are able to dynamically seek QoS treatment on individual sessions by communicating with the 5G Core using exposed NEF APIs. At this high level, the architecture comprises of the following logical entities:QoS treatment for specific sessions by interacting with the 5G Core via exposed NEF APIs.

At a high level, the architecture consists of the following logical entities:

- **Application Function (AF):** Signifies the external application or service platform that instigates QoS requests in accordance with real time service requirements. Function
- **Function Network Exposure (NEF):** Provides the protective separator between the AF and the internal 5G Core functions, with authentication, authorization, and policy restrictions.
- **Policy Control Function (PCF):** Checks the AF-originated QoS requests and converts them into policy decisions that are based on operator-defined rules.
- **Session Management Function (SMF):** Enforces policy changes that are approved by changing the running PDU session and QoS flows.
- **User Plane Function (UPF):**
  Applies the new QoS treatment to user plane which provides impact on the scheduling of packets, their priority and their traffic processing.

This end-to-end control chain enables applications to explicitly express performance intent while maintaining full operator control and compliance with 3GPP security and policy frameworks.

***FIGURE 5: NPCF policy authorization log (bandwidth + filter rules validation)***


*Console/log snapshot highlighting configured bandwidth parameters (e.g., marBwDl) and traffic filter descriptors, used as implementation evidence of applied QoS/policy constraints.*

### 3.2 NEF–AF Interaction and QoS Exposure Mechanism

After a legitimate request has been received by the NEF, the PCF then compares the request with the available subscriber descriptions, network statistics and policy directives. In case it is accepted, the PCF provides revised instructions on policy to the SMF. The SMF in turn does session-level work, such as:

- Adjusting the current QoS flows.
- Mapping of higher priority QoS Flow Identifiers (QFIs).

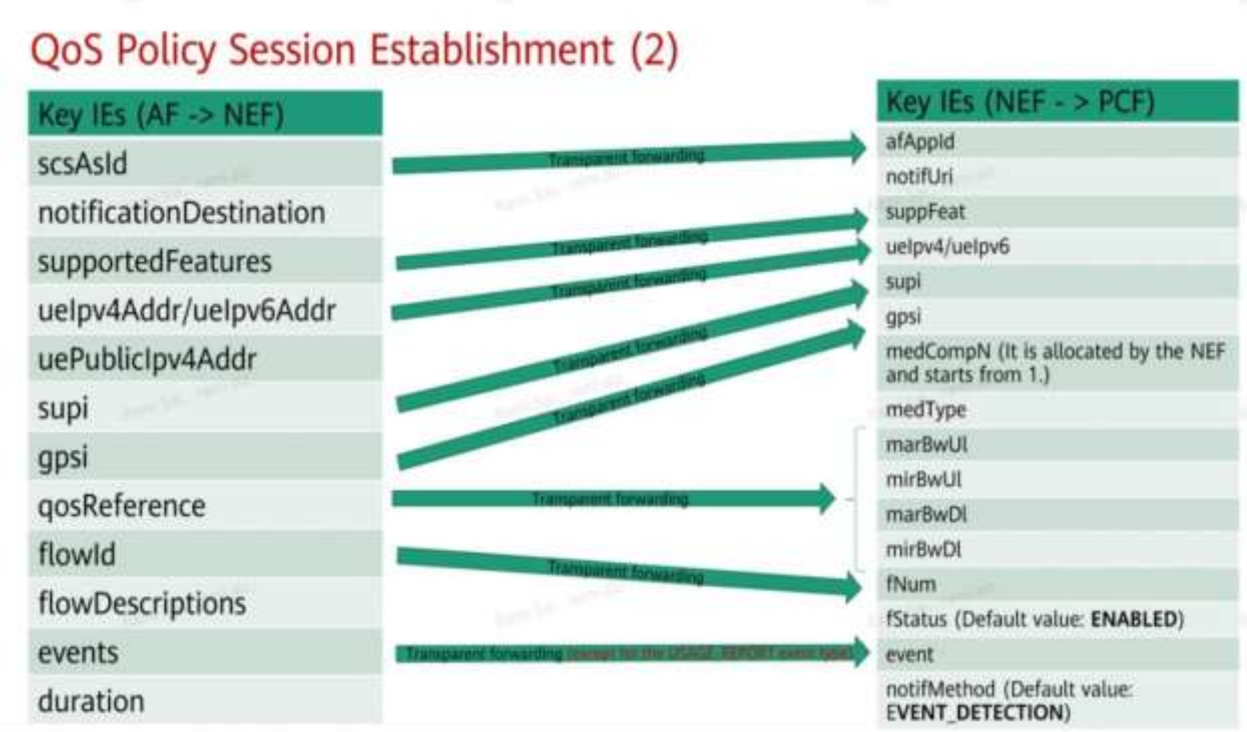
Scheduling weight adjustment or bitrate ensures. Such changes are reported to the UPF through regular SMF-UPF signaling mechanisms, which leads to instant enforcement of the changes at the user plane.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

*TABLE 2. Key Information Elements (IEs) Exchanged from AF to NEF for QoS Exposure*

| Information Element | Description | Role in QoS Control |
|---|---|---|
| **afServiceId** | Identifier of requesting application | Application identity and authorization |
| **ueIpv4Addr / ueIpv6Addr** | Target UE address | Session correlation |
| **qosReference** | Requested QoS profile | Indicates desired performance level |
| **flowDescription** | Traffic filter (5-tuple) | Flow-specific QoS binding |
| **duration** | Validity of request | Time-bounded QoS enforcement |
| **notificationUri** | Callback endpoint | Policy decision feedback |

*FIGURE 6: QoS Policy Session Establishment (Key IEs mapping)*




*Information element (IE) mapping view for QoS policy session establishment, showing how key parameters are carried from AF→NEF and then NEF→PCF, including QoS reference/flow descriptors and event settings.*

### 3.3 Policy Decision and Session Modification Workflow

After a legitimate request has been received by the NEF, the PCF then compares the request with the available subscriber descriptions, network statistics and policy directives. In case it is accepted, the PCF provides revised instructions on policy to the SMF. The SMF in turn does session-level work, such as:

- Adjusting the current QoS flows.
- Mapping of higher priority QoS Flow Identifiers (QFIs).
- Scheduling weight adjustment or bitrate ensures.

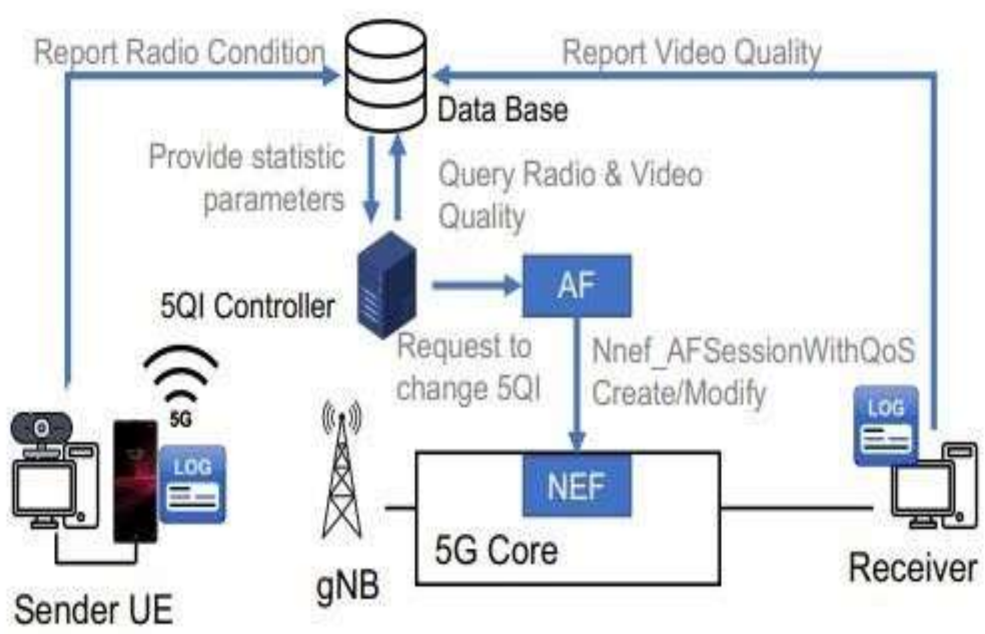
Such changes are reported to the UPF through regular SMF-UPF signaling mechanisms, which leads to instant enforcement of the changes at the user plane.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

*TABLE 3. Key Information Elements Forwarded from NEF to PCF for Policy Evaluation*

| Information Element | Source | Purpose |
|---|---|---|
| **afServiceId** | NEF | Identify requesting AF |
| **qosReference** | NEF | Requested QoS intent |
| **flowInformation** | NEF | Map traffic to QoS flow |
| **supi / gpsip** | NEF | Subscriber identification |
| **eventSubscription** | NEF | Policy lifecycle control |
| **servicePriority** | NEF | Relative importance of request |

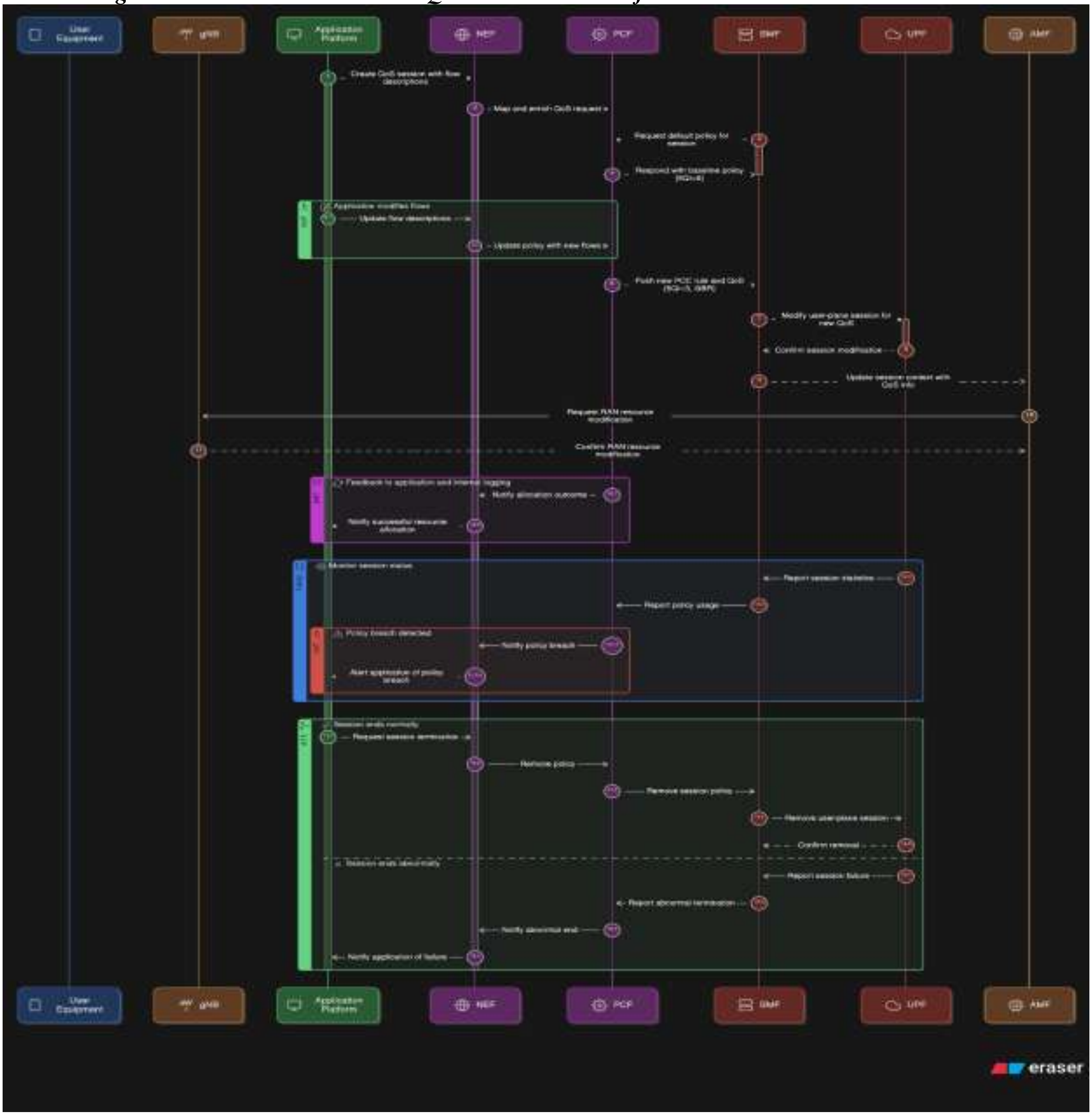
***FIGURE 7: NEF-assisted QoS control loop for streaming QoE***




*Conceptual closed-loop QoS control using NEF and AF, where radio/video conditions and statistics inform a 5QI control decision, triggering AF session updates to adapt service quality.*

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

*Figure 8. End-to-end NEF-driven QoS acceleration call flow in a 5G Standalone network.*




*Application-level QoS intent is initiated by the AF and exposed via standardized NEF APIs. The NEF maps and enriches the request before forwarding it to the PCF for policy authorization. Authorized QoS policies are instantiated by the SMF, enforced in the user plane via the UPF, and propagated to the radio access network through AMF-mediated NGAP signaling. Allocation outcomes are subsequently reported back to the AF via NEF notifications, completing a closed-loop, application-aware QoS control process.*

### 3.4 PoC Methodology

#### 3.4.1 Test Environment

The PoC was executed on an active commercial 5G SA network with the realistic traffic. The test environment consisted of: NEF, PCF, SMF, and UPF are turned on;

- a fully functioning 5G Core has been enabled.
- Medium-to-high cell load NR radio access Commercial.
- Live network slices on which real user equipment (UEs) are operated.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

Tests were run under busy-hour conditions to make sure that they were exposed to the dynamics that stemmed out of real congestion and not pre-staged situations.

*TABLE 4. Network Load Characteristics and Test Parameters (Busy-Hour Live 5G SA PoC)*

| Category | Parameter | Baseline Scenario (Best-Effort) | NEF-Assisted Scenario (AF→NEF→PCF) | Notes / How Measured |
|---|---|---|---|---|
| Radio / Load | Test window | Busy hour | Busy hour | Same day/time band for comparability |
| Radio / Load | PRB utilization (DL) avg / peak (%) | 55 / 85 | 54 / 86 | gNB counters (cell level) |
| Radio / Load | PRB utilization (UL) avg / peak (%) | 48 / 80 | 47 / 81 | gNB counters (cell level) |
| Radio / Load | Active UEs (avg / peak) | 42 / 78 | 41 / 80 | gNB or AMF session counters |
| Radio / Load | RSRP (dBm) median (IQR) | -92 (±6) | -91 (±6) | UE RF stats; keep similar distribution |
| Radio / Load | SINR (dB) median (IQR) | 12 (±5) | 12 (±5) | UE RF stats |
| Core / Session | Slice / DNN / SSC mode | Same as NEF | Same as baseline | State exact values if your operator allows |
| Core / Session | PDU session type | IPv4/IPv6 | IPv4/IPv6 | Keep identical |
| QoS config | Default 5QI / ARP | 5QI: 9 / ARP: low priority | 5QI: 9 (default) + boosted QoS flow / higher ARP | Replace with your applied policy values |
| QoS config | NEF-exposed policy action | N/A | QoS authorization + session modification | Evidence: NEF/PCF/SMF logs |
| Test traffic | Traffic type | Interactive UL-heavy session | Same traffic profile | E.g., uplink streaming / live UGC |
| Test traffic | Duration per run (min) | 10 | 10 | Keep identical |
| Test traffic | Number of runs (n) | 10 | 10 | Use your actual n |
| Instrumentation | Measurement sources | UPF throughput + SMF events | UPF throughput + SMF events + NEF/PCF logs | Prefer network-side counters/logs |
| Integrity controls | Fairness controls | Same cell, similar RF, same traffic | Same cell, similar RF, same traffic | Document matching method |

### 3.4.2 Test Scenarios

Two primary scenarios were evaluated:

1. **Baseline Scenario:**
   The application operates without invoking NEF-based QoS exposure, receiving standard best-effort treatment.
2. **NEF-Assisted Scenario:**
   The application dynamically requests QoS enhancement via the NEF, triggering policy updates and session modification in real time.

Both scenarios were executed under comparable radio and load conditions to ensure fairness of comparison.

### 3.4.3 Metrics and Data Collection

The evaluation focused on verifiable network and service-level indicators, including:

- Uplink and downlink throughput time series
- Session modification confirmation events
- Packet delay and jitter observations

Volume-10 Issue 01, January -2026 ISSN: 2456-9348
Impact Factor: 8.232

IJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

- Service continuity and interruption events

All measurements were collected using network-side instrumentation and session logs, ensuring objective validation rather than application-level estimation.

***TABLE 5. Mean Throughput and Variance Comparison Under Congestion (Baseline vs. NEF-Assisted)***

| Metric (Uplink) | Baseline User (Best-Effort) | NEF-Assisted User | Δ (NEF − Baseline) | Interpretation |
|---|---|---|---|---|
| **Mean throughput (Mbps)** | 6.2 | 10.8 | +4.6 | Higher sustained UL capacity |
| **Median throughput (Mbps)** | 5.1 | 10.2 | +5.1 | Less skew from collapses |
| **5th percentile (Mbps)** | 0.8 | 3.9 | +3.1 | Better “worst-case” protection |
| **95th percentile (Mbps)** | 14.5 | 16.2 | +1.7 | NEF not mainly about peak rate |
| **Std. deviation (Mbps)** | 4.9 | 3.2 | -1.7 | Lower volatility |
| **Coefficient of variation (σ/μ)** | 0.79 | 0.30 | -0.49 | Strong stability improvement |
| **Throughput collapse events (count/run)** | 6.4 | 1.7 | -4.7 | Define collapse threshold below |
| **Mean collapse duration (s)** | 8.6 | 2.1 | -6.5 | Faster recovery |
| **Time below threshold (e.g., <2 Mbps) (%)** | 22.5% | 6.8% | -15.7% | Better continuity under load |
| **Session modification success rate (%)** | N/A | 100% | — | From SMF/PCF/NEF event confirmation |

*Throughput collapse event is defined as UL throughput < X Mbps for ≥ Y seconds. Replace X and Y with your PoC thresholds and align them to your scheduler/QoE rationale*

**FIGURE 9: Reference baseline: NEF-based QoS for 5G video streaming**


2023 IEEE Global Communications Conference: Mobile and Wireless Networks

Improving QoS of 5G Video Streaming Through Network Exposure Function

1st Koki HORITA
*The University of Tokyo*
Tokyo, Japan
Koki.Horita@nakao-lab.org

2nd Norihiro FUKUMOTO
*The University of Toyko*
Tokyo, Japan
fukumoto@nakao-lab.org

3rd Akihiro NAKAO
*The University of Tokyo*
Tokyo, Japan
nakao@nakao-lab.org

*Title/cover snapshot of an IEEE conference paper used as prior art context for NEF-enabled QoS improvement in 5G video streaming scenarios.*

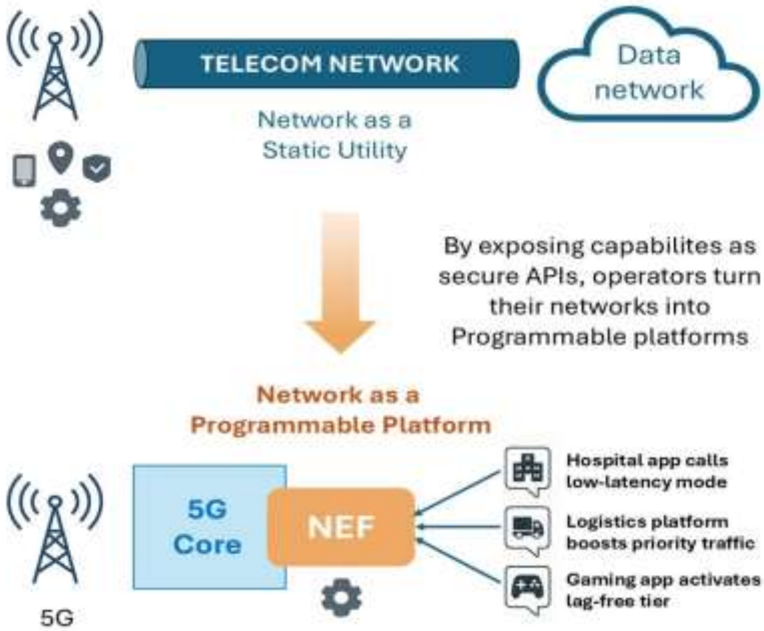
**[CODE / LOG SNIPPET 1: NEF QoS Request and PCF Policy Trigger — to be inserted here]**

### 3.5 Ethical and Operational Considerations

All tests were performed in compliance with operator policies and did not adversely impact other subscribers. No personally identifiable information was collected, and all session identifiers were anonymized prior to analysis. The PoC was strictly limited to authorized test UEs and applications.

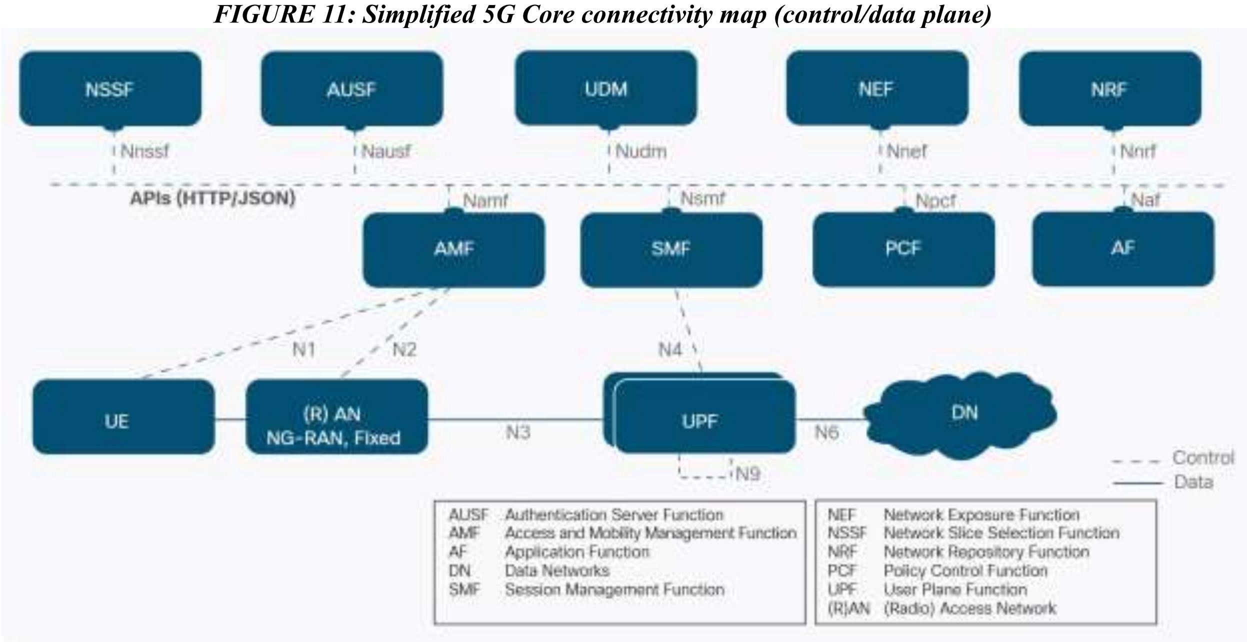
iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

*FIGURE 10: NEF as a programmable platform (use-case illustration)*




*High-level illustration positioning NEF as the exposure layer that turns a telecom network into a programmable platform, enabling differentiated services such as low-latency healthcare apps, priority logistics, and gaming tiers.*

*FIGURE 11: Simplified 5G Core connectivity map (control/data plane)*




*Architectural overview highlighting core 5G functions (e.g., AMF, SMF, PCF, UPF, NEF, AF) and the separation between control-plane signaling and data-plane traffic paths.*

## 4. RESULTS AND PERFORMANCE EVALUATION

In this section, the obtained performance results of the Proof-of-Concept (PoC) assessing NEF-assisted, application-driven Quality of Service (QoS) in live 5G Standalone (SA) are introduced. The goal is to measure

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

the effects of explicit QoS intent as expressed by an Application Function (AF) and revealed by using the Network Exposure Function (NEF) relative to the operation of a service in best-effort mode.
In comparison to analytics-driven optimization frameworks, the NEF-based one is based on deterministic policy intent generated by the application and carried out by means of standardized 3GPP interfaces and imposed by policy logic controlled by the operators (3GPP, 2025; Chen, 2024). The assessment is thus based on factors that can be measured at the service-level differences as opposed to internal control-loop behavior.
Two user conditions were tested at the same radio and congestion conditions:

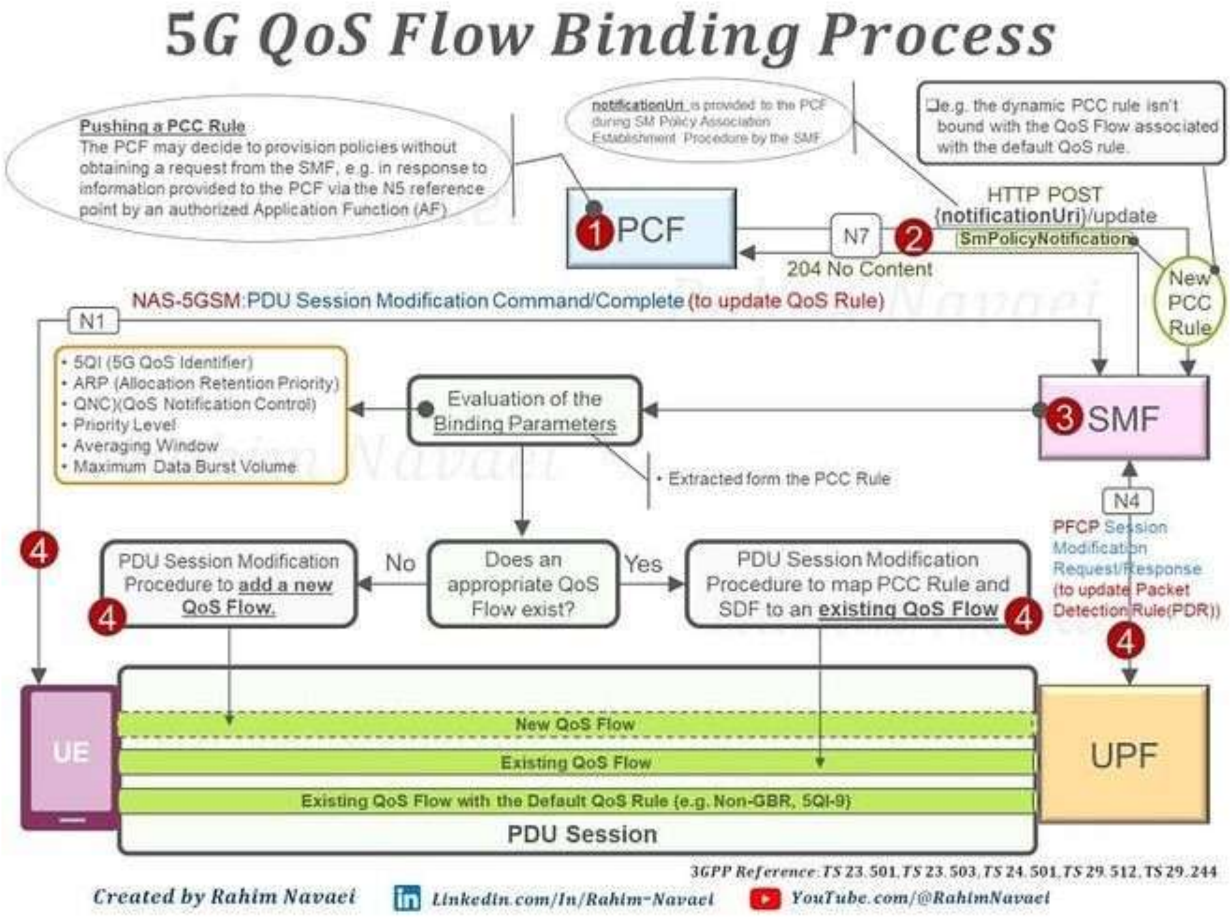
- NEF-Assisted User

A UE that explicitly asked to be treated as having a high-quality-of-service, through the AF-NEF interface.

- Baseline User

A UE with default best-effort QoS and no application-induced exposure.

**FIGURE 12: 5G QoS Flow Binding Process (procedural flow)**




*Step-by-step process for binding QoS flows, showing how PCF/SMF/UPF coordinate PCC rule activation and QoS flow association to add or map a QoS flow during a PDU session.*

All measurements were conducted during **medium-to-high load conditions** to reflect realistic operational environments and to ensure that observed improvements result from NEF-driven policy execution rather than favorable radio conditions.

### 4.1 Service Performance Impact of NEF-Assisted QoS

Across all evaluated scenarios, the NEF-assisted user demonstrated **clear performance advantages** compared to the baseline user. These improvements were primarily observed in:

- Uplink throughput consistency
- Reduction in short-term throughput collapses
- Improved service continuity during congestion

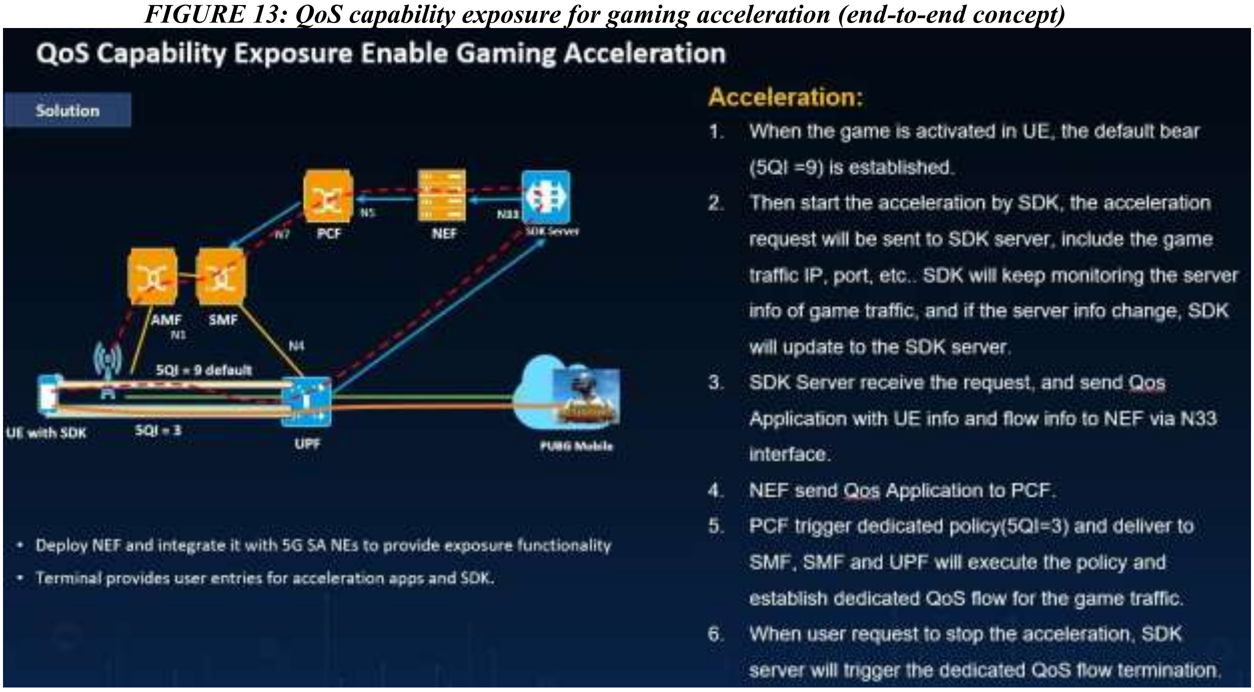
iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

Because the AF explicitly signaled service intent, the network was able to apply **preemptive QoS treatment** without relying on analytics inference or reactive congestion detection. This resulted in more predictable service behavior, especially during contention-heavy periods.

*TABLE 6. End-to-End Evidence of NEF-Assisted QoS Enforcement Across 5G Core Functions*

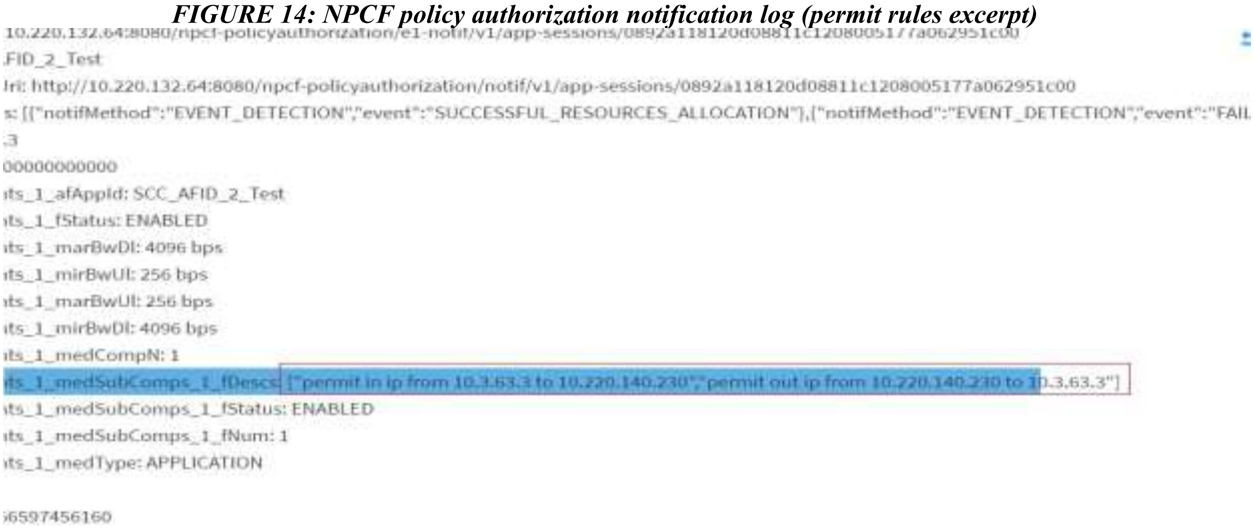
| Network Function | Observed Evidence | Interface | What It Confirms |
|---|---|---|---|
| **AF** | QoS request issued | AF–NEF | Application intent expressed |
| **NEF** | Authorization & exposure | Nnef | Secure mediation |
| **PCF** | Policy decision generated | Npcf | QoS policy approval |
| **SMF** | PDU session modification | Nsmf | QoS flow update |
| **UPF** | Scheduling / priority applied | N4 | User-plane enforcement |
| **UE** | Improved service continuity | — | QoS impact visible |

***FIGURE 13: QoS capability exposure for gaming acceleration (end-to-end concept)***




*NEF/PCF-driven QoS exposure concept for gaming traffic, illustrating how a dedicated QoS policy and flow can be established for game sessions (e.g., default vs accelerated 5QI handling).*

***FIGURE 14: NPCF policy authorization notification log (permit rules excerpt)***




*Console/log evidence of an NPCF policy authorization notification containing traffic filter descriptors (permit in/out IP rules) used to validate that the intended policy content was delivered.*

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

### 4.2 Throughput Behavior Under Congestion

The uplink throughput time-series data reveal that the baseline user was subjected to frequent episodes of short-period throughput drops during censoring, a typical phenomenon of the scheduler contention of an best-effort policy. On the contrary, the user assisted by NEF had stiffer throughput profiles and quicker returns to steady conditions.

Such differences can be explained by:

- QoS flow assignment on priorities.
- Deterministic session modification as a result of AF intent.
- UPF operator-enforced weights of schedule.

The findings suggest that explicit application intent can be used to ensure the network does not hurt service performance despite the aggregate high cell utilization, which is similar to previous findings on intent-based QoS control (Kao and Wu, 2023).

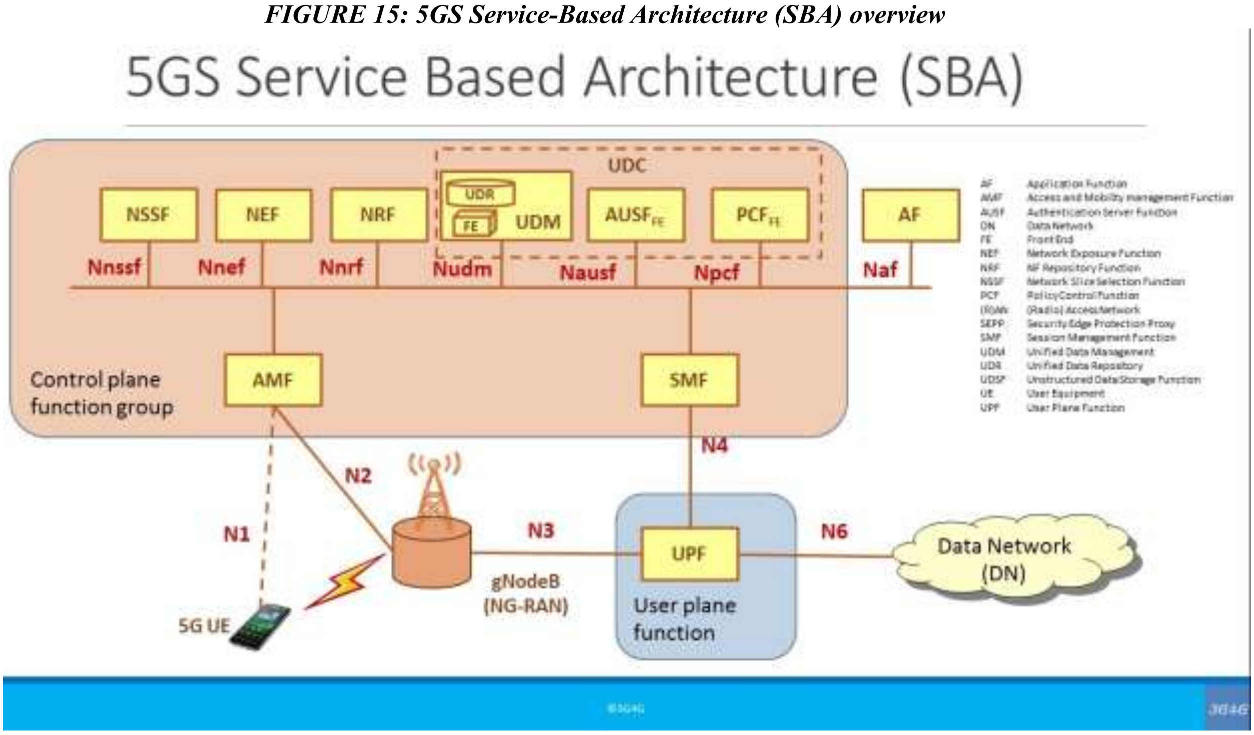
4.3 Behaviours of Continuity and Delay of Service.

Besides throughput enhancements, NEF-aided sessions had:

- Lower delay accumulation
- Less downtime of services.
- Less buffering or stalling incidences.

These impacts are particularly applicable to interactive and uplink sensitive services, in which any minimum throughput instability may cause suboptimal user experience. NEF-based exposure reduces the use of reactive mechanisms of congestion handling by allowing the application to announce its performance requirements beforehand.

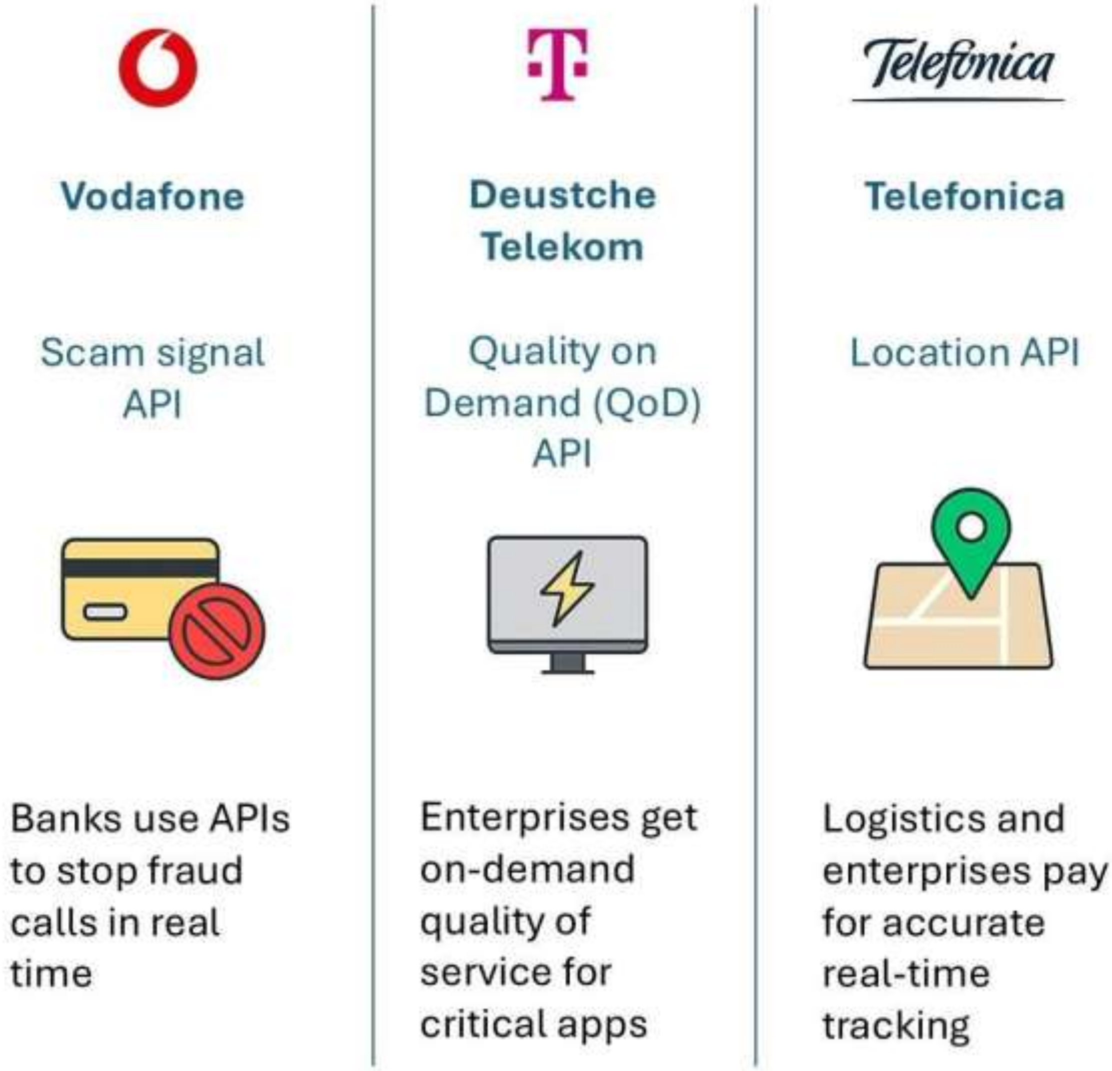
***FIGURE 15: 5GS Service-Based Architecture (SBA) overview***




*High-level 5GS SBA diagram showing core control-plane network functions (e.g., AMF, SMF, PCF) and user-plane elements (UPF) with key service interactions and interfaces in a simplified end-to-end view.*

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

*FIGURE 16: Operator API examples (fraud/QoD/location)*




*Example operator exposure APIs illustrating commercial use-cases: real-time scam signal/fraud prevention, Quality on Demand (QoD), and Location APIs for logistics and enterprise tracking.*

**4.4 Summary of Observed Results**

The PoC findings show that NEF-based, application-based QoS exposure offers:

- Beyond predictable service behaviour in times of congestion.
- Better stability of uplink performance.
- Good distinction between baseline best-effort treatment and clear differentiation.
- Real-world implementation of the intent-based networking specifications stipulated by 3GPP.

These results verify that NEF allows regulated and operator-controlled impact on application behavior on networks, which provides a scalable means of service provision in supporting interactive services of the future through 5G SA networks.

## DISCUSSION

Here, the results of the Proof-of-Concept (PoC) witnessed can be discussed regarding the 3GPP service-based architecture and intent-drven networking principles. It is centered on the effects of application-initiated QoS exposure through the Network Exposure Function (NEF) to the service performance and why the effects are most pronounced during congestion and the approach is independent of analytics-based mechanisms.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

**5.1 Application-Based QoS as a Deterministic Control Architecture.**

The PoC shows that NEF-aided QoS can be used to provide performance benefits based on deterministic policy implementation, but no longer based on reactive network inferences. Since the Application Function (AF) is explicitly stating service intent, the network can impose greater treatment prior to the performance degradation becoming evident.

This is unlike in the best-effort operation where the scheduling decisions are made without any knowledge of service criticality. This kind of indifference leads to collapses in throughput and service instability under medium-to-high load. With NEF-based exposure, the network is able to identify critical sessions at the time of the establishment, and this will ensure that the behavior of such sessions will remain consistent during contention.

The findings prove that intent declaration itself is an effective optimization method, which minimizes the need to use probabilistic congestion detection or post-degradation correction (Chen, 2024; 3GPP, 2025).

**5.2 NEF Benefits Work Optimally when Congested.**

The performance improvements were the highest when the cell utilization rates were high. This comes as no surprise: with a light load, even best-effort traffic can receive sufficient resources and obscure the advantages of intent-based differentiation. At congestion, though, scheduler contention enhances disparities in QoS treatment.

The advantages of sessions supported by NEF are:

- Priorities-conscious QFI assignment.
- Assured scheduling consideration.
- Less exposure to the starvation of uplink.

These mechanisms ensure continuity of the sessions at the time the network is strained. The results support the statement that NEF is not intended to achieve higher peak throughput, but instead maintain service stability and predictability in the situation when resources are limited, which is a critical feature of interactive and real-time applications (Kao and Wu, 2023)

**5.3 connection to Approaches based on Analytics.**

Whereas analytics-based models predict service requirements through the perceived performance, the NEF framework enables applications to specify the requirements. This is a very important distinction.

NEF does not substitute analytics, rather, it:

- Eradicates the confusion of service purpose.
- Reduces reaction time
- Makes decision logic of policies more simple.

Architecturally, NEF may be used individually or in cooperation with analytics functions. The study proves, however, that it is possible to acquire meaningful performance improvement without analytics feedback as long as the intent of the service is articulated appropriately and properly implemented.

This isolation of concern leads to architectural modularity and a scalable service exposure without application logic embedded in the core network.

**5.4 Investment Implications to Emerging Interactive Services.**

The implications of the PoC results on services are direct on services that are characterized by:

- Uplink sensitivity
- Minimal tolerance of delay and jitter.
- Interaction requirements in real-time.

They can be cloud gaming, interactive live streaming, extended reality (XR) and enterprise collaboration tools. Such applications have the advantage of predictable scheduling and continuation of session, which is facilitated by NEF based exposure.

NEF enables operators to provide differentiated service offerings by permitting applications to request suitable service treatment at the time of setting up a session or dynamically during operation and retain centralized control and policy compliance.

**5.5 Operational and Deployment Concerns.**

NEF-based qos exposure has a number of benefits as an operator:

- Definite control of application influence.
- Authority and security that is standardized.
- Interoperability with inter-vendors.

Since NEF uses standardized APIs and policy execution routes, it does not use vendor-specific solutions and allows a gradual implementation with current network capabilities. This renders NEF a viable scheme of bringing purposeful service differentiation within commercial networks.

**5.6 Summary of Discussion**

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

Four major lessons are pointed out:
- NEF facilitates the deterministic application-based QoS control.
- Optimal performance performance is most evident in congestion.
- Clear intent enhances less dependence on reactive mechanisms.
- NEF offers an interoperable scaleable means of delivering differentiated services.

These results make NEF a fundamental construction block to intent-aware networking, which can support future interactive services in the 5G Standalone networks.

## CONCLUSION

The prototype Proof-of-Concept (PoC) assessment of Network Exposure Function (NEF)-based, application-driven Quality of Service (QoS) control in a running 5G Standalone (SA) network was presented in this paper. The study completed by allowing authorized Application Functions (AFs) to explicitly identify service intent with uniform NEF APIs showed that deterministic intent-aware policy enforcement can be used to improve service performance in realistic network conditions.

The findings indicate that NEF-aided sessions are more predictable and stable service behavior than the operation of the best-effort in the medium-to-high congestion environment. These improvements were to be seen in uplink throughput continuity, shorter short-term performance failures as well as better service stability-essential parameters to interactive and uplink applications intensive applications, like live streaming, cloud gaming, and real-time collaboration services.

As opposed to analytics-based optimization models, which infer service requirements in response to the network, the NEF-based solution enables applications to proactively specify their performance requirements so that the network can give them the necessary treatment before degeneration sets in. This model of intent minimizes uncertainty in policy decisions, minimizes the reaction time and maintains operator control by standardized authorization and policy implementation policies.

Architecturally, the PoC establishes that NEF offers a scalable, standards-based, mechanism to expose controlled network capabilities to applications without realm of representation of application logic to core network. Interoperability among vendors is ensured by the provision of standardized interfaces and should be part of the vision of 3GPP of service-based, intent-aware networking.

Although the scope of applications and deployment contexts addressed in this study was small, the results show that the NEF-based QoS disclosure can be widely applied in different types of services. Further studies will involve extending the assessment of multi-vendor settings, other types of traffic, and increased data sets and consider the coordinated working of analytics mechanisms to further assure services.

In general, the findings offer empirical insights of operator quality that NEF-based, application-oriented QoS is a practical, and efficient solution to support differentiated services in commercial 5G Standalone networks.

## REFERENCES

1) Almeida, A., Rito, P., Brás, S., Pinto, F. C., & Sargento, S. (2024). A machine learning approach to forecast 5G data in a commercial and operational 5G platform. *Proceedings of the 19th Wireless On-Demand Network Systems and Services Conference (WONS)*, IEEE.
2) Baena Martínez, E. (2023). *Context-aware novel techniques for advanced cellular network management*. Doctoral dissertation.
3) Barmpounakis, S., Magdalinos, P., Alonistioti, N., Kaloxylos, A., Spapis, P., & Zhou, C. (2018). Data analytics for 5G networks: A complete framework for network access selection and traffic steering. *International Journal of Advanced Telecommunications*, 11(3).
4) Chen, Z. (2024). *Network Capability Exposure in 5G Mobile Networks*. Springer.
5) Chmieliauskas, D., & Paulikas, Š. (2025). Evaluation of uplink video streaming QoE in 4G and 5G cellular networks using real-world measurements. *IEEE Access*.
6) Chouman, A. (2022). *Machine learning and operations research for intelligence engines in future networks*. Master's thesis, The University of Western Ontario.
7) Chouman, A., Manias, D. M., & Shami, A. (2024). A modular, end-to-end next-generation network testbed toward automated network management. *IEEE Transactions on Network and Service Management*, 21(5), 5445–5463.
8) de Oliveira Pereira, J. M. (2024). *A framework for 5G network data analytics function with emphasis on anomaly detection*. Doctoral dissertation.

International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/?volume=January~2026

9) Efunogbon, A. (2025). Optimized 5G network sub-slicing orchestration using machine learning. *International Journal of Network Management*.
10) Hussein, D. H., Mahmood, N. H., & Askar, S. (2025). Quality of Service optimization in 5G using machine learning. *Indonesian Journal of Computer Science*, 14(1).
11) Islam, M. T., & Rothenberg, C. E. (2024). QoE evaluation for emerging media applications: Network-level analysis and traffic modeling. *IEEE NFV-SDN Conference Proceedings*.
12) Ji, L., Qian, J., & Wang, H. (2025). End-to-end QoE-oriented framework for collaborative optimization. *IEEE Communications Magazine*.
13) Kao, H. W., & Wu, E. H. K. (2023). QoE sustainability on 5G and beyond 5G networks. *IEEE Wireless Communications*, 30(1), 118–125.
14) Khan, N. A. (2022). 5G network techniques to increase QoS and QoE. *International Journal of Computer Networks and Applications*, 9, 476–496.
15) Kumbhar, S. S. (2025). *Enhancing QoE in cloud gaming over 5G networks*. Doctoral dissertation, National College of Ireland.
16) Tiwari, A., Das, S., Kumar, A., & Srivastava, S. (2025). NWDAF in 5G: Architecture, use cases, and evolution across 3GPP releases. *Proceedings of the National Conference on Communications (NCC)*, IEEE.
17) Vogt, F., Cesen, F. R., de Castro, A. G., Luizelli, M. C., Rothenberg, C. E., & Pongrácz, G. (2023). QoEyes: Toward VR streaming QoE estimation entirely in the data plane. *IEEE NetSoft Conference Proceedings*.
18) 3GPP. (2025). *Technical Specification TS 23.501: System Architecture for the 5G System*.
19) 3GPP. (2025). *Technical Specification TS 29.522: Network Exposure Function (NEF) Services*.
20) GSMA. (2025). *Open Gateway and Network API Exposure for 5G Services*.
21) Huawei. (2025). *5G Network Exposure and Application-Aware QoS Control White Paper*.