White Paper

Executive Summary

Application Management in the 5G Standalone Core

Authors

This whitepaper is jointly developed by SK Telecom and Intel Corporation.

SK Telecom — Core Network Dev. Team, 5GX Technology Group

DongJin Lee
5G Core Architect, dongjin@sk.com

SeongJun Lee
5G Core Architect, seoul.lee@sk.com

HyunJun Choi
5G Core Architect, chj@sk.com

Intel Corporation— Data Center Group

Andriy Glustsov
Software Architect
andriy.glustsov@intel.com

Peter McCarthy
Software Architect
peter.mccarthy@intel.com

Chetan Hiremath
Sr Principal Engineer
chetan.hiremath@intel.com

Khaled Qubaiah
Software Architect
khaled.qubaiah@intel.com

Rory Browne
Solutions Architect
rory.browne@intel.com

Introduction

As the mobile industry begins the next wave of 5G deployments, we see insatiable growth for services, end devices, and associated applications. Mobile networks continue to evolve and expand service mixes based on fixed wireless access (FWA), enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communication (URLLC). This evolution is coupled with vast application ecosystems from Apple, Google, and others. The sheer scale of the number of networking devices and applications continues to outpace our expectations.

For many years, there have been mechanisms in core networks to implement policy and charging for users and associated applications, but now massive user growth poses an operational problem for mobile core networks. Already, most networks in developed countries are oversubscribed. Analysts and vendors predict that a ramp in deployments of IoT and mMTC devices will add several billion endpoints in the next few years, and large application stores already host several million distinct applications. We are addressing this problem of how to design the core networks to scale to support this increasing demand while simplifying core network operations.

This paper outlines a key research and development solution from SK Telecom and Intel for application management in 5G core networks. We have engineered the user plane function (UPF) such that packet processing, QoS, and usage reporting is performed per application ID within an application group. This architecture provides a flexible way to map, group, apply policy, and bill based on the application traffic class and can be achieved at scale without high utilization or operational tax. This approach simplifies and scales the entire application management architecture in 5G core networks.

Overview

To address the problems of continuing user equipment (UE) and application growth as outlined, Intel and SK Telecom have partnered to deliver a 5G application management solution in order to simplify the complexity of configuration by introducing an application group hierarchy concept. This allows us to apply policies on a per-application group level while being able to classify and process traffic for a significant number of applications configured in the context of each protocol data unit (PDU) session.

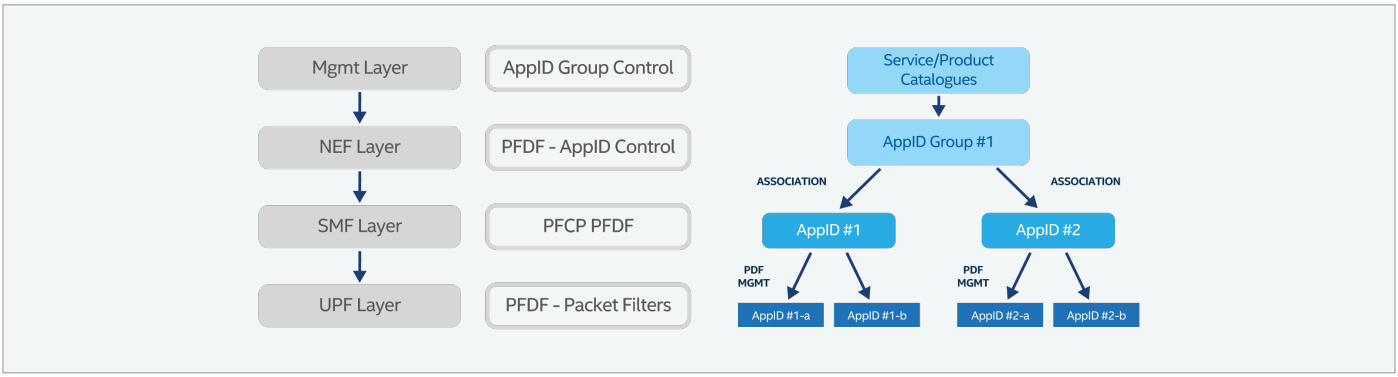

Figure 1. Application ID and Application Group Hierarchy

Figure 1 illustrates the overall concept of an application management hierarchy that is particularly focused on the relationship between applications and their groups. To aid in the understanding of this paper, the following terms are defined:

1. Application ID: this is an octet string identifier that can be mapped to a set of application detection rules based on either flow description, or URL or Domain Name. The application detection information is provisioned in the UPF using Packet Flow Descriptions (PFDs) and can be configured dynamically or predefined in a way to increase operational efficiency.
2. Application Group: This is an octet string that defines a group that has one or more application IDs as members. Core network functions such as UPF, Session Management Function (SMF), Network Exposure Function (NEF), and Policy Control Function (PCF) implement this feature so policies can be managed in groups. Each application in the group inherits that new configuration.
3. Application Detection and Control (ADC): This feature allows the UPF to match traffic to a particular application ID or group, and report to the SMF for various policy control purposes such as QoS control and usage monitoring.

Application traffic may be identified, classified, and have policy applied based on some service groups and application IDs. This is a hierarchical structure whereby each application can be selected for granular treatment in terms of QoS, bandwidth, latency, jitter etc.

Figure 2 illustrates the UPF packet processing architecture according to 3GPP TS 29.244. Packets of a particular traffic are matched with a Packet Detection Rule (PDR) that maps with the application ID and are processed according to various rule types, such as Forwarding Action Rule (FAR), QoS Enforcement Rule (QER), and Usage Reporting Rule (URR).

One challenge is optimizing thousands of different UE policies when they also use different types of application traffic. Also challenging, because it impedes scaling, is managing different UE profiles while maintaining different rules per each individual subscriber's PDU session. For example, a large number of applications and corresponding packet filters for packet detection could quickly increase the size of the PDR lookup information for each Packet Forwarding Control Protocol (PFCP) session, that in turn increases the memory footprint, memory bandwidth, and cache usage by a PDR lookup process. This architecture poses both an operational (provisioning) and a technical challenge (increased control plane traffic from the NEF/PCF and SMF for policy control) for network operators as the number of UEs grow within a network.

An ideal solution, developed by SK Telecom and Intel, is to use application IDs and their application group to allocate resources to UEs to identify their traffic profiles and policies. That is, subscribers are associated with one or more application groups. An application group, identified by an application group ID, could have one or more applications (identified by their application ID) contained within it.

The key concept is all applications that are part of a particular group have the same policy and charging rules associated with them (i.e., precedence, forwarding, QoS, usage, etc.), thereby substantially simplifying network operations and doing so within acceptable processing, throughput, and latency constraints.

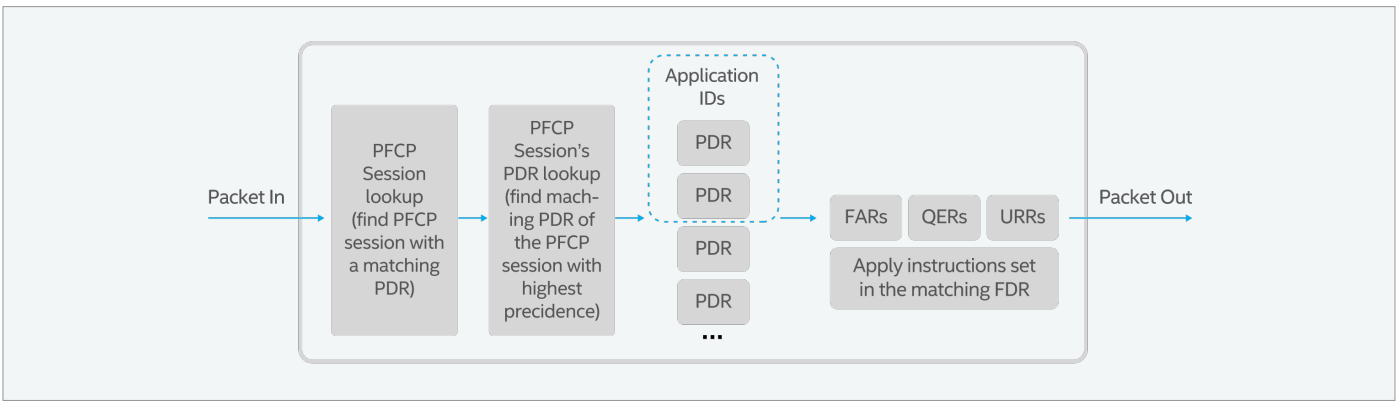

Figure 2. Packet Processing Flow in the User Plane Function

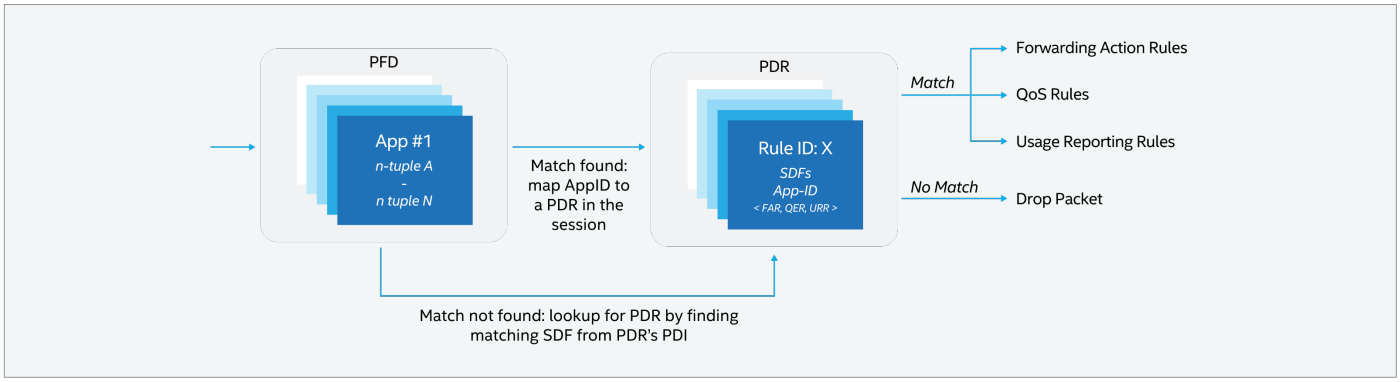

Figure 3. Application Detection Architecture

New Application Architecture

Application ID grouping assumes that one application group includes multiple applications, whereby the number of applications in the group may have a wide variation defined by the operator. A single Policy and Charging Control (PCC) rule associated with an application group carries policy and charging information for all applications within the group. In some cases, operators may have requirements to control specific aspects, such as usage reporting or QoS parameters on per-application basis, in addition to controlling these parameters at an application group level for applications with similarities such as messengers, video streaming, etc. This application group functionality can be achieved by deriving per-application rules from the application group/PCC rule. Expanding the application group into a list of application IDs and building corresponding PDRs could be performed at different levels.

Our proposed architecture closely follows the 3GPP TS 23.501, 23.502, and 23.503 technical specifications with an exception to the application ID and group, and relevant network functions (NFs), including UPF, SMF, NEF, and PCF. Our enhancements are shown in Figure 3.

User Plane Function (UPF)

PDRs per application group are delivered from the SMF to the UPF over the N4 interface. The UPF implements application group to application ID mapping by expanding the PDR for an application group into a PDR Set with the PDR for each application ID in the group using preconfigured mapping information.

As a result, the amount of information needed to be communicated over N4 is reduced because only a single PDR Set needs to be delivered from the SMF to the UPF per application group. If operators require an ADC or usage reporting at the application level, these operations can be performed by 1) using predefined FARs/QERs/URRs, 2) extracting FARs/QERs/URRs derived from the application group PDR, 3) or combining these two methods.

For application grouping, we assume most of the application traffic can be delivered using a default QoS flow, and the signaling of packet filters to a UE for such applications is not required. Packet filter information for selected applications that require special QoS handling (e.g., IMS application traffic) can still be signaled to the UE, assuming the SMF ensures the number of packet filters for all signalled QoS rules for a PDU session does not exceed the maximum number indicated by the UE.

Session Management Function (SMF)

The SMF is responsible for PDU session management and could support application group expansion, thereby creating a PDR Set for each application within application groups. For example, if the SMF identifies the PDU session for some UEs has to be configured with 10 application groups, with each group having 20 application IDs on average, the total number of PDRs to be signalled from the SMF to the UPF over N4 interface is about 400 if different PDRs are used for uplink/downlink directions.

This large number of PDRs requires multiple PFCP session modification request/response messages to be communicated over the N4 interface, in addition to the initial PDU session creation requests, resulting in the PDU session establishment/modification process taking extra time. Information about how to expand an application group into a set of application IDs is maintained at the SMF level. Such functionality is not defined by the 3GPP and could be implemented based on customer requirements. The UPF does not need to maintain information about application groupings, so any 3GPP-compliant UPF can work with such an SMF. Because of heavy N4 communication overhead resulting from the application group expansion in the SMF, the solution proposed in this document implements application group expansion functionality in the UPF.

Network Exposure Function (NEF)/ Policy Control Function (PCF)

The NEF performs the Packet Filter Description (PFD) management procedure, which delivers PFD context (identified as an application ID) to the SMF via the Nnef interface. Note the PDFs could be preconfigured in the SMF and UPF for known applications. The NEF's PFD can be shared and managed with other NFs and third-party nodes per the operator's policy.

The PCF performs the PDU session level procedure using Policy Control and Charging (PCC) rules via the Npcf interface, which is responsible for controlling how the application IDs or application groups should be executed, such as forwarding, QoS rating, and charging. It works with the NEFs to make sure the PCC rule names are mapped to application names. The PCC rule's service data flow (SDF) template may have a set of packet filters or an application ID that references the corresponding application detection filter. This remains unchanged in the proposed solution.

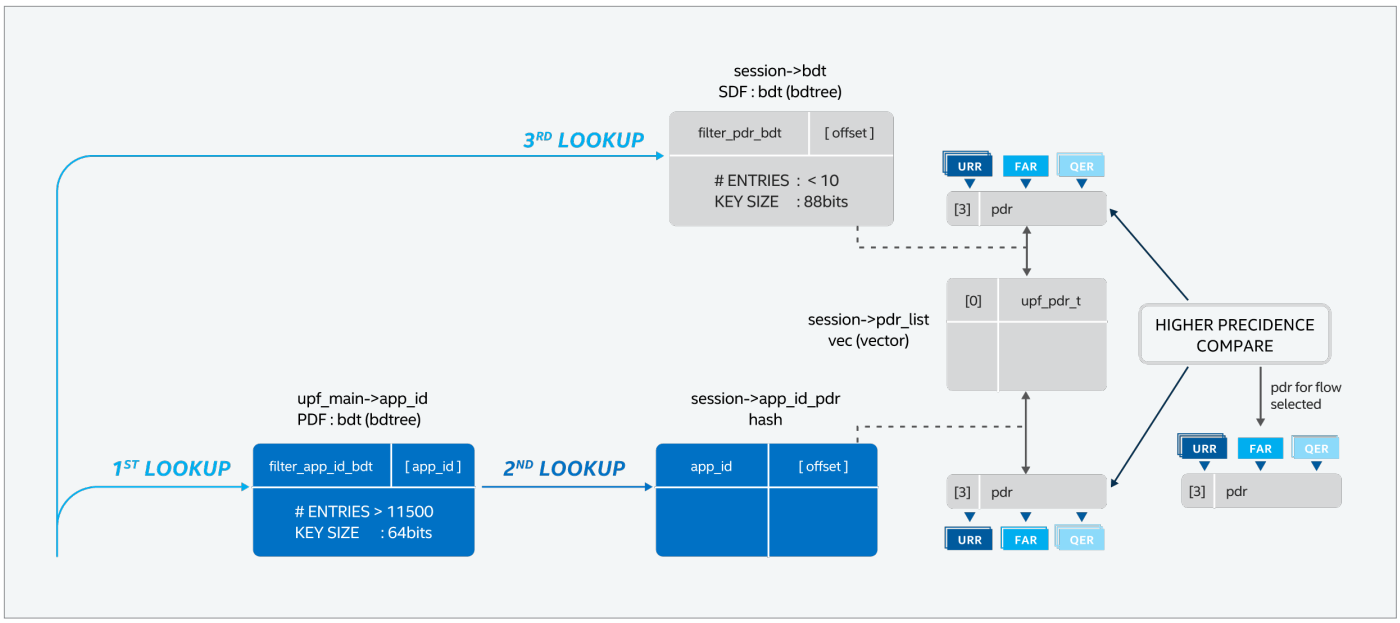

Figure 4. PDR Set Detection Logic for Matching Application ID Within an Application Group

Prototype Implementation

We used a preconfigured application group at the SMF and UPF level such that:

- The UPF employs a static configuration on the application ID grouping (map of application group into a set of application IDs).
- The SMF delivers PDRs for application groups (as well as independent PDRs for individual applications and PDRs with SDF filters) to the UPF over N4.
- The UPF derives information per application ID from PDR Sets and enables per application usage reporting (in addition to the per application group reporting configured by the SMF); the SMF supports usage reporting for individual applications and groups.

The solution for application group expansion required the following changes:

- Add support of application groups in the UPF
- Modify SMF functionality to support usage reporting on per application and per application group levels
- Modify the Application Detection stage within the UPF packet processing pipeline

To avoid overloading the interactions between the SMF and UPF on session establishment, the UPF takes on the job of expanding the application groups on session establishment. This expansion was achieved through a static mapping of applications within a group in the UPF. The rules (i.e., QER/FAR/URR and precedence) associated with a group were sent down over N4 interface, and the UPF expanded the group upon session establishment/modification. The Application Detection stage of the UPF pipeline was modified from what is defined in the standard (see Figure 4) as follows:

- First Lookup: Application ID detection. In this step, the UPF looks for the application ID in the global 'lookup database' (one per UPF) constructed from PFD information configured in the UPF. Note that we used numeric indexing of the strings to make UPF indexing more efficient and obviate the need for a string to integer conversion.
- Second Lookup: If an application ID is detected, the UPF looks for the PDR that references this application ID in the context of PFCP session, using a hash table which maps application ID to a PDR. These two lookups return the PDR associated with the application (or no matched PDR).
- Third Lookup: PDR lookup in the session context. In this step, the UPF looks for the PDR in the 'lookup database' constructed in the context of the current PFCP session from all PDRs having SDF filters as packet detection information.

If PDRs are found in both legs (i.e., first, second, and third lookups), the leg with the lowest precedence (i.e., highest priority) is chosen.

Test Setup and Measurement Results

We used five UPFs running on a single socket server, with every UPF configured to use 4 physical cores (which with hyperthreading enabled results in 8 packet processing threads per UPF instance). UPF #1 and #2 are set to handle PFCP sessions with 16 application groups configured in for each session, totaling 107 application IDs and more than 10,000 packet filters. UPF #3, #4, and #5 use lighter configuration with 10 application IDs configured in PFCP sessions. The system, based on an Intel® Xeon® Platinum 8280 processor (38.5M Cache, 2.70 GHz), emulated 50,000 UEs (10,000 per UPF) with a throughput of 100 Gbps using packet sizes as indicated in figure 5a, with a traffic profile of 25 Gbps uplink and 75 Gbps downlink.

As shown in Figure 5a, the UPF performs periodic usage reporting for each application and application group in every PDU session. PFDs for application detection are statically defined in the UPF or can be provisioned from the SMF. PDRs and corresponding QERs, URRs and FARs are dynamically configured per application or application group during PFCP session establishment procedure. Specifically, the number of application ID entries per group (varying per group), along with the application group ID proceed as follows:

- Addition of AppGroup.xml: Mapping the application group ID to the application ID is performed in the UPF, as there is no definition for it in 3GPP 29.244. A static configuration file was introduced to provide the mapping.
- PFCP Session Establishment/Modification: The processing of these messages was heavily modified to enable the expansion of a single PDR (with a PDR ID indicating a PDR group) into multiple PDRs (PDR Set) for the associated application IDs. If in the result of PDR expansion multiple resulting PDRs reference the same application the PDR with lowest precedence value should be selected.
- PDU Detect Node: The existing single binary decision tree (BDT) lookup is replaced with a three-stage lookup involving two BDT lookups and one hash lookup.

The application group logic takes a single PDR that is "tagged" as an application group PDR and expands it into one or more actual application ID PDRs based on the mapping in the AppGroup.xml file. All this expansion happens at the PFCP Session Establishment (or Modification) stage of processing. It is important to note the application group "layer" has been removed for that PFCP session by the time any packet processing happens. That is, only PDRs with legitimate application IDs that map to existing PDF rules will exist after the PFCP Session Establishment/Modification procedure is complete.

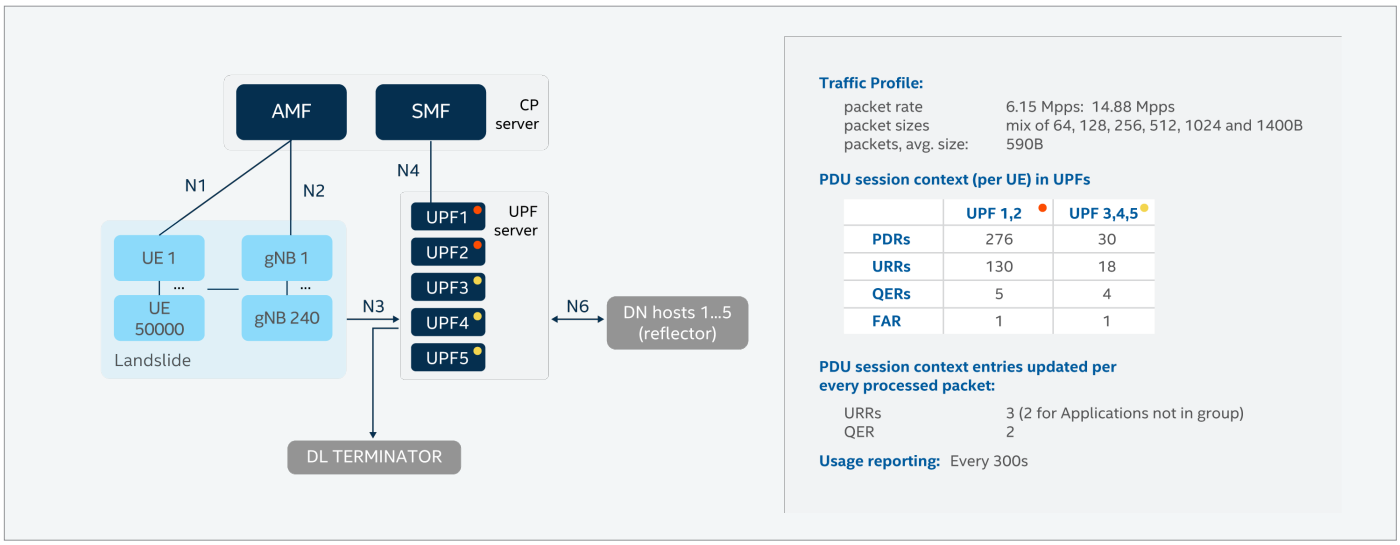

Figure 5a. Performance Test Setup for Application ID and Group

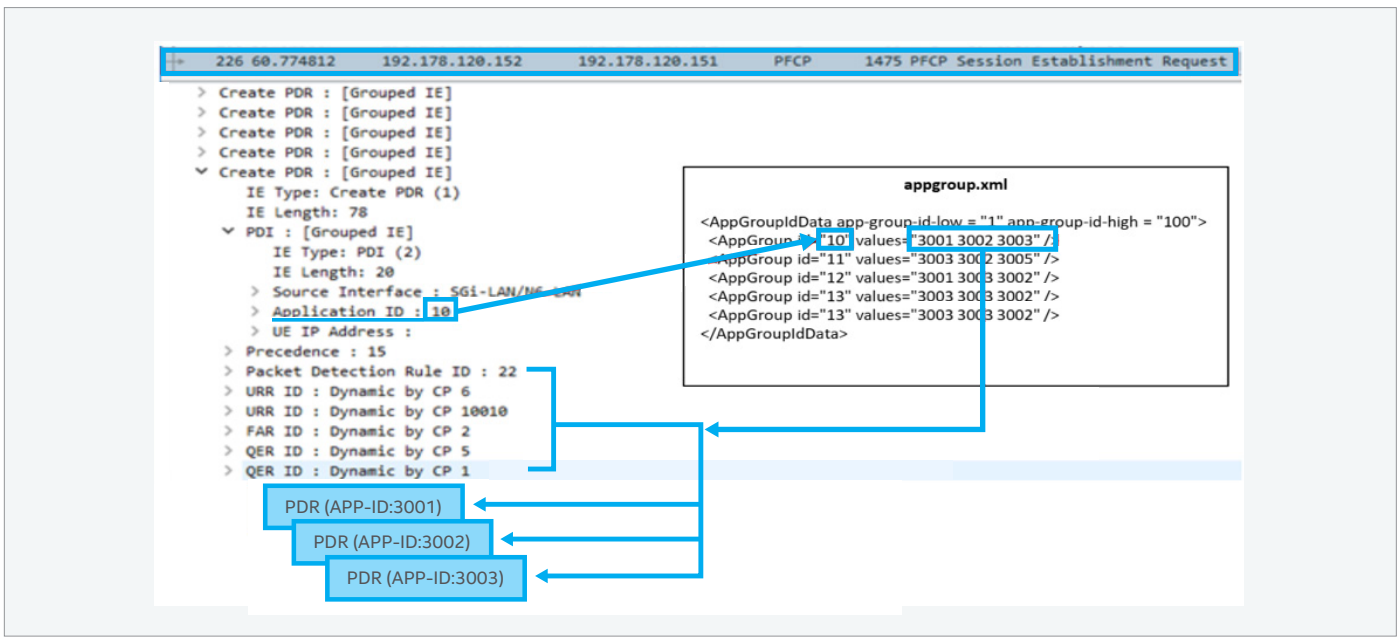

Figure 5b. PFCP Session Establishment - mapping Application Group to Application IDs

Once the Session establishment/modification has been fully processed, the first packets for a session arrive and flow identification occurs, along with the creation of a new entry in the flow table. At this point the PDR detect logic is exercised. Whereas before a single lookup of a session-based BDT was done, now there is a three-stage lookup that happens as shown in Figure 4. The application detection stage is executed for the first few UL and DL packets in the flow, the result of the operation is saved in the context of the flow cache entry and used for processing next packets of the same flow. We tested classification and application ID processing across several test servers and cases, measuring utilization with and without the application ID architecture.

For the development and testing environment, we used 5G core network reference stacks from ASTRI (Applied Science and Telecoms Research Institute - www.astri.org) to demonstrate a typical packet processing pipeline performance. During testing, we stressed the performance of the UPF. Other elements such as the UE, RAN, and DN were simulated by test equipment. We also used the Access and Mobility Management Function (AMF) and the SMF in order to test the user plane as sessions were established, deactivated, and moved due to mobility events.

For application ID testing, we used three Spirent Landslide test servers to run 24 test cases in total. Each test case had unique traffic profile configuration and simulated traffic from a subset of the UEs (with PDU sessions established over five UPF instances), with every UE generating 12 data message flows for several applications that could belong to different application groups.

As shown in Figure 6, application ID grouping was enabled on all UPF instances. UPF #1 and #2 had heavy configurations with 107 applications configured (that were visible from the number of PDRs in the sessions), while other UPFs ran lighter configurations with one application group (10 applications) and a couple standalone applications enabled. The results confirmed the architecture could handle configurations with a great number of configured applications and packet filters. Even with an aggressive application ID group hierarchy and configuration, the worker cores on UPF #1 and #2 only used about an additional seven percent core load to implement the architecture.

Furthermore, we greatly reduced the number of PDRs that must be signaled on the N4 interface. For UPFs 1 and 2, the number of PDRs was reduced from 276 (using a flat application ID architecture) to 38 using the hierarchical approach we described in this paper. This represents an 86 percent reduction in N4 signaling for the configuration above.

	PDRs SMF delivers to UPF per PDU session			PDU session configuration in UPF (post PDR expansion)			Number of applications UPF classifies			
	PDRs with AppID	PDRs with AppGroupID	PDRs with SDF	PDRs	URRs	QERs		Packet Rate	Bitrate	Average CPU utilization
UPF1	4	32	2	276	130	5	107	4.24 MPPS	21.2 Gbps	92.1%
UPF2	4	2	2	276	130	5	108	4.11 MPPS	20.6 Gbps	91.0%
UPF3	4	2	2	30	18	4	12	4.24 MPPS	21.2 Gbps	84.8%
UPF4	4	2	2	30	18	4	12	4.22 MPPS	21.1 Gbps	85.2%
UPF5	4	32	2	30	18	4	10	4.22 MPPS	21.1 Gbps	83.6%

Figure 6. System Load with Application ID Architecture

Conclusion

This paper identified and described a solution for key areas in the deployment of standalone (SA) core networks with an application ID hierarchy for 5G and with platform considerations that match the architecture of the SK Telecom 5G SA system. With respect to application IDs, we demonstrated that this flexible way of mapping and grouping application forwarding and billing policy can be delivered at scale in the 5G SA by Intel® architecture processors without a heavy utilization tax on the UPF processor resources, and while simultaneously reducing N4 signaling by up to 86 percent.

Thus, we decoupled the exponential rise in 5G subscriber application utilization with the associated network policy and billing. Additional improvement considerations were found during the implementation phase, which is left for further research. Looking forward, Intel and SK Telecom will continue to collaborate on the 5G and Beyond 5G Core with new software and hardware architectures, including processor and NIC technologies for optimal performance.

Acronyms

Term	Description
ADC	Application Detection and Control
AMF	Access and Mobility Management Function
eMBB	Enhanced Mobile Broadband
FAR	Forwarding Action Rule
mMTC	Massive Machine-type Communication
NEF	Network Exposure Function
PCC	Policy and Charging Control
PCF	Policy Control Function
PDR	Packet Detection Rule
PDU	Packet Data Unit
PFCP	Packet Forwarding Control Protocol
QER	QoS Enforcement Rule
SMF	Session Management Function
UPF	User Plane Function
URLLC	Ultra-reliable Low-latency Communication
URR	Usage Reporting Rule

© 2020 Intel Corporation Printed in USA 0920/BB/ICMCSW/PDF ♻ Please Recycle 344693-001 US