Wireless Priority Services

Feature Summary and Revision History

Summary Data

Table 1: Summary Data

Applicable Product(s) or Functional Area	SMF
Applicable Platform(s)	SMI
Feature Default Setting	Disabled – Configuration Required
Related Changes in this Release	Not Applicable
Related Documentation	Not Applicable

Revision History

Table 2: Revision History

Revision Details	Release
UPF Interaction while Deleting WPS Dynamic Rule	2021.01.0
SBI Message Priority Mechanism and Message-Prioritization based on Procedures are introduced.	2021.01.0

Revision Details	Release
The Wireless Priority Services feature is fully qualified in this release.	2020.03.0
First introduced. This feature is not fully qualified in this release. For more information, contact your Cisco Account representative.	2020.02.0

Feature Description

The Wireless Priority Services (WPS) feature is supported on the SMF+PGW-C over 5GC. The SMF+PGW-C validates prioritization of WPS services for session creation or modification and various handover scenarios. The SMF+PGW-C also evaluates the WPS services for Paging-Policy Differentiation for Network Triggered Service Request procedures.

Important With release 2021.02.0, SMF will not set MP flag in N4 message while deleting dynamic rule if no other existing rules ARP isn't matching wps-profile.

Use Cases

The WPS feature implements the 3GPP recommendations for wireless priority support for the following use cases in 5GS and EPS. The use cases are defined as per 3GPP TS 23.501 (sections 5.16.3, 5.16.4, 5.16.5, 5.16.6, 5.19, and 5.21).

WPS supports the following use cases:

- Multimedia Priority Services , on page 2
- DSCP Marking for N3, S5-U, or S2-B over PFCP, on page 7

Multimedia Priority Services

The Multimedia Priority Service (MPS) allows priority access to system resources to Service Users, creating the ability to deliver or complete sessions of a high priority nature. Service Users are government-authorized personnel, emergency management officials or other authorized users. MPS supports priority sessions on an "end-to-end" priority basis. MPS includes signalling priority and media priority.

MPS provides the ability to invoke, modify, maintain and release sessions with priority, and deliver the priority media packets under network congestion conditions.

All MPS-subscribed UEs get priority for QoS Flows (for example, used for IMS signalling) when established to the DN that is configured to have priority for a given Service User by configuring MPS-appropriate values in the QoS profile in the UDM. Service Users are treated as On Demand MPS subscribers and not On Demand MPS subscribers, based on regional or national regulatory requirements. On Demand service is based on Service User invocation or revocation explicitly and applied to the media QoS Flows being established. Not On Demand MPS service does not require invocation and provides priority treatment for all QoS Flows only to the DN that is configured to have priority for a given Service User after attachment to the 5G network.

Priority treatment for MPS includes priority message handling for Mobility Management procedures. Priority treatment for MPS session requires appropriate ARP and 5QI setting for QoS Flows according to the operator's policy.

MPS priority mechanisms can be classified as subscription-related mechanism and invocation-related mechanism. Subscription-related mechanisms can be applied as "always applied" and "conditionally applied".

Subscription-related mechanisms that are conditionally applied include:

- UDM—One or more ARP priority levels are assigned for prioritized or critical services. The ARP of the prioritized QoS Flows for each DN is configured to an appropriate ARP priority level.
- PCF—The "IMS Signalling Priority" information is configured for the subscriber in the UDM, and the PCF modifies the ARP of the QoS Flow used for IMS signalling.

Invocation-related mechanisms can be applied for mobile-originated SIP call or sessions, for mobile-terminated SIP call or sessions, and for Priority PDU connectivity services.

On-Demand MPS Service

The invocation-related priority mechanisms for prioritized services are based on communication with an Application Server and between the Application Server and the PCF over Rx or N5 interface (as described in 3GPP TS 23.228, clause 5.21, in the case of MPS using IMS).

Invocation-related mechanisms for Mobile Originations (for example, through SIP or IMS) are explained as follows:

- PCF:
  - When an indication for a session reaches over the Rx or N5 interface and the UE does not have priority for the signaling QoS Flow, the PCF derives the ARP and 5QI parameters plus associated QoS characteristics as appropriate, as per the Service Provider policy (specified in 3GPP TS 23.503, clause 6.1.3.11).
  - For MPS sessions, when establishing or modifying a QoS Flow as part of the session origination procedure, the PCF selects the ARP and 5QI parameters, and the associated QoS characteristics, as appropriate, to provide priority to the QoS Flows.
  - When all active sessions to a particular DN are released and the UE is not configured for priority treatment to that particular PDU session, the PCF downgrades the IMS Signaling QoS Flows from appropriate settings of the ARP and 5QI parameters and the associated QoS characteristics, as appropriate, to those entitled by the UE based on subscription.

Invocation-related mechanisms for Mobile Terminations (for example, through SIP or IMS) are explained as follows:

- PCF: When an indication for a session reaches over the Rx or N5 interface, the mechanisms as described above for Mobile Originations are applied.
- UPF: If an IP packet arrives at the UPF for a UE that is CM-IDLE, the UPF sends a "Data Notification" including the information to identify the QoS Flow for the DL data packet to the SMF (specified in 3GPP TS 23.502, clause 4.2.3.3.
- SMF: If the SMF receives the "Data Notification" message for a QoS Flow associated with an ARP priority level value for priority use, delivery of priority indication during the Paging procedure is provided by inclusion of the ARP in the N11 interface "N1N2MessageTransfer" message (specified in 3GPP TS 23.502, clause 4.2.3.3).

- AMF: If the AMF receives the "N1N2MessageTransfer" message containing an ARP priority level value for priority use, the AMF handles the request with priority. AMF also includes the "Paging Priority" IE in the N2 "Paging" message configured to a value assigned to indicate about an existing IP packet at the UPF requiring higher priority (specified in 3GPP TS 23.502, clause 4.2.3.3).
- SMF: For a UE that is not configured for a higher priority, upon receiving the "N7 Session Management Policy Modification" message from the PCF with an ARP priority level for priority use, the SMF sends an "N1N2MessageTransfer" to update the ARP for the Signaling QoS Flows (specified in 3GPP TS 23.502, clause 4.3.3.2).
- AMF: After receiving the "N1N2MessageTransfer" message from the SMF with an ARP priority level for priority use, the AMF updates the ARP for the Signaling QoS Flows (specified in 3GPP TS 23.502, clause 4.3.3.2).
- (R)AN: Inclusion of the "Paging Priority" in the N2 "Paging" message triggers priority handling of paging during congestion at the (R)AN (specified in 3GPP TS 23.502, clause 4.2.3.3).

Invocation-related mechanisms for the Priority PDU connectivity services:

- PCF:
  - If the state of the Priority PDU connectivity services is modified from disabled to enabled, the QoS Flows controlled by the Priority PDU connectivity services are established or modified to have the service appropriate configuration of the ARP and 5QI parameters and the associated QoS characteristics, using the PDU Session Modification procedure (specified in of 3GPP TS 23.502, clause 4.3.3).
  - If the state of Priority PDU connectivity services is modified from enabled to disabled, the QoS Flows controlled by the Priority PDU connectivity services are modified from service appropriate configuration of the ARP and 5QI parameters and the associated QoS characteristics, to those entitled by the UE as per subscription, using the PDU Session Modification procedure (specified in 3GPP TS 23.502 clause 4.3.3).

Message-Priority Indication over GTP-C

An overloaded node performs message prioritization when handling incoming messages during an overloaded condition. This condition is based on the relative GTP-C message priority signaled in the GTP-C header.

When message throttling is performed:

- GTP requests related to priority traffic (eMPS as described in 3GPP TS 22.153) and emergency have the highest priority. Depending on regional or national requirements and the network operator policy, these GTP requests are the last to be throttled when applying traffic reduction. The priority traffic is exempted from throttling due to GTP overload control up to the point where the requested traffic reduction cannot be achieved without throttling the priority traffic.
- For other types of sessions, message throttling considers the relative priority of the messages so that low priority messages are considered for throttling before the other messages. The relative priority of the messages is derived from the relative priority of the procedure for which the message is being sent (as specified in clause 12.3.9.3.2) or derived from the session parameters such as APN and ARP.

The high priority messages are given lower preference to throttle and low priority messages are given higher preference to throttle. An overloaded node also applies these message prioritization schemes when handling incoming initial messages during an overloaded condition, as part of the self-protection mechanism.

A sending GTP-C entity determines the relative message priority to signal in the message according to either procedure based or session parameters. If the message affects multiple bearers (for example, Modify Bearer Request), the relative message priority considers the highest priority ARP among all the bearers.

A GTP-C entity sets the same message priority in a Triggered message or Triggered Reply message as received in the corresponding Initial message or Triggered message respectively. For incoming GTP-C messages that do not have a message priority in the GTP-C header, the receiving GTP-C entity:

- Applies a default priority if the incoming message is an Initial message.
- Applies the message priority sent in the Initial message or Triggered message if the incoming message is a Triggered message or Triggered Reply message.

The nodes in the network homogenously support this feature to prevent an overloaded node to process initial messages received from the non-supporting nodes. These messages are received according to the default priority. The overloaded node processes initial messages that are received from the supporting nodes according to the message priority signaled in the GTP-C message.

Message-Prioritization based on Session Parameters

Message prioritization is also performed based on the session parameters, such as APN and ARP. The procedures and messages associated with the higher priority sessions are given lesser priority while throttling than the procedures and messages associated with the lower priority sessions. Within each group of sessions, the messages are further prioritized based on the category of the procedure for which the message is being sent.

Message Prioritization Based on Procedures

Message prioritization is performed based on the relative priority of the procedure for which the message is being sent. Procedures are grouped into various categories and each of these categories are assigned a priority. In addition, within a given category of procedures, messages can be further prioritized based on session parameters, such as APN, QCI, ARP or LAPI.

Messages with a high priority are given lower preference to throttle and messages with low priority are given higher preference to throttle. The grouping of the procedures isn't performed based on an individual GTP-C entity but while considering all the procedures in general. A GTP-C entity considers the procedures applicable to it and prioritizes message throttling based on the category of the procedure. The categories are listed in decreasing order of priority with category 1 having the highest priority. For each category, a nonexhaustive list of messages is provided. Any existing or newly defined message in future is considered based on the category of the procedure for which the message is sent. Following are the categories of a procedure:

1. UE session mobility within and across 3GPP or non-3GPP access—Procedures involving active or idle mode UE mobility, such that GTP-C signalling involved are classified under this category. Some examples are X2 or S1 based handover with or without an SGW change, TAU or RAU with a change of MMEor SGSN with or without an SGW change, and 3GPP access to trusted non-3GPP access handover. Throttling of these messages during the procedures related to UE session mobility results in the failure of the corresponding procedures. This failure can cause PDN disconnection or the interruption of the services. As a result, the following messages, when sent during the procedures belonging to this category, must be considered with the highest priority. Hence, these messages are given the lowest preference to throttle.
    - Create Session Request.
    - Create Session Request with "handover" indication bit set.
    - Modify Bearer Request.

- Modify Bearer Request with "handover" indication bit set.
- Modify Access Bearer Request.

2. Release of PDN connection or bearer resources—Procedures resulting in the deactivation of an existing PDN connection, the deactivation of bearers or of data forwarding tunnel of an UE leads to freeing up of the resources at the overloaded node. These procedures ease the overload situation as the freed up resources can be used for serving the remaining of the UEs. Hence, the following messages that belong to this category and cause the deactivation of PDN connection or bearers or data forwarding tunnels, must be treated with the next lower level of priority. Hence, these messages are given the corresponding preference whilst throttling:
   - Delete Session Request.
   - Delete Bearer Request.
   - Delete Bearer Command.
   - Delete Indirect Data Forwarding Tunnel Request.
3. Miscellaneous session management procedures—This category consists of the session management procedures, except the PDN connection creation and bearer creation or modification procedures. Some examples are location reporting, when it isn't combined with other mobility procedures and Service request and S1 release procedure. These procedures do not impact the ongoing service of the UE. Hence, the following messages when sent during the procedures identified under this category, must be treated with the next lower level of priority. Hence, these messages are given the corresponding preference whilst throttling.
   - Release Access Bearer Request.
   - Modify Bearer Request.
   - Change Notification.
   - Suspend Notification.
   - Resume Notification.
4. Request for new PDN Connection or bearer resources or modification of existing bearer resources—This category consists of the procedures requesting the creation of PDN connection, creation or modification of bearers, or creation of data forwarding tunnel. Throttling of the messages belonging to this category cause denial of new services while continuing with the existing services. In this overload condition, an overloaded node, due to lack of resources, isn't able to provide new services while trying to maintain the existing services. When the following messages are sent during the procedures belonging to this category are considered with the lowest level of priority. Hence, these messages are given highest preference to throttle:
   - Create Session Request during PDN connection request.
   - Create Bearer Request.
   - Update Bearer Request.
   - Bearer Resource Command.
   - Modify Bearer Command.

- Create Indirect Data Forwarding Tunnel Request.
- Downgrade the DSCP marking of the data packets for the session when quota exhausts.

Message-Priority Header for PFCP

When the message throttling is performed:

- PFCP requests related to priority traffic (that is, eMPS as described in 3GPP TS 22.153) and emergency have the highest priority. Depending on regional or national requirements and network operator policy, these PFCP requests are the last to be throttled when applying traffic reduction. Throttling exempts the priority traffic due to PFCP overload control up to the point where the requested traffic reduction cannot be achieved without throttling the priority traffic.
- For other types of sessions, the message throttling considers the relative priority of the messages so that the messages with low priority are first considered for the throttling. The relative priority of the messages is derived from the relative priority of the procedure for which the message is being sent or derived from the session parameters such as APN and ARP.

A PFCP entity determines whether to configure and use the message priority in PFCP signalling, based on operator policy. A sending PFCP entity determines the relative message priority to signal in the message which are derived from the session parameters, such as APN and ARP. If the message affects multiple bearers, the relative message priority is determined considering the highest priority ARP among all the bearers. A PFCP entity must configure the same message priority in a Response message as received in the corresponding Request message.

For incoming PFCP messages that do not have a message priority in the PFCP header, the receiving PFCP entity:

- Applies a default priority if the incoming message is a Request message.
- Applies the message priority sent in the Request message if the incoming message is a Response message.

The SMF and UPF functions in the network homogenously support this feature to prevent an overloaded node to process the Request messages received from the non-supporting nodes according to the default priority. With this support, an overloaded node does not need to process the Request messages received from supporting nodes according to the message priority signalled in the PFCP message.

DSCP Marking for N3, S5-U, or S2-B over PFCP

Transport Level Marking

Transport level marking is the process of marking traffic with a DSCP value based on the locally configured mapping from the QCI and optionally the ARP priority level. For EPC, the S-GW and PGW-C perform transport level marking on a per EPS bearer basis. For 5GC, the S-GW and PGW-C perform transport level marking on a per QoS flow basis.

The UPF performs transport level marking with a DSCP value based on the mapping from the 5QI, the Priority Level (if explicitly signaled), and optionally the ARP priority level configured at the SMF. The CP function controls transport level marking by providing the DSCP in the ToS or Traffic Class within the Transport Level Marking IE in the FAR (associated to the PDR matching the traffic to be marked).

The UP function performs transport level marking for the detected traffic and sends the marked packet to the peer entity. The CP function changes transport level marking by changing the Transport Level Marking IE in the related FAR.

WPS Profile Support

The SMF+PGW-C supports the WPS profile defined with ARP and DSCP marking value to be configured for GTP-C and PFCP Protocol IP-headers. Use the WPS profile to configure the message priority in the GTP-C and PFCP protocols.

The SMF+PGW-C allows a maximum of 64 WPS profiles and each WPS profile is associated in the DNN profile. For more information, see the Configuring Wireless Priority Services, on page 8 section.

How it Works

This section describes how Wireless Priority Service (WPS) feature works.

Standards Compliance

The Wireless Priority Services feature complies with the following standards:

- 3GPP TS 22.153
- 3GPP TS 23.228
- 3GPP TS 23.282
- 3GPP TS 23.379
- 3GPP TS 23.501
- 3GPP TS 23.502
- 3GPP TS 23.503
- 3GPP TS 24.301

Configuring Wireless Priority Services

This section describes how to configure the Wireless Priority Services feature.

Configuring the WPS Profile

Use the following sample configuration to configure the WPS profile.

```
config
  profile wps wps_profilename
    arp arp_value
    dscp [ n3 n3_value | message-priority { [ { gtpc | pfcp } [ arp | dscp
 ] ] }
    end
```

NOTES:

- **profile wps** *wps_profilename*: Accesses the Wireless Priority Services Profile configuration. *wps_profilename* must be an alphanumeric string of 1 to 63 characters.
- **arp** *arp_value*: Specifies the range of ARP levels. *arp_value* must be an integer from 1 to 15 separated either by "," or "-".
- **dscp [n3** *n3_value* **]** : Specifies the DSCP marking value for the N3 interface. The N3 value indicates the UP DSCP marking value within the range 0 to 0x3F.
- **message-priority { gtpc pfcp }**: Specifies the message priority for GTP-C and PFCP.

Verifying the WPS Profile Configuration

This section describes how to verify the WPS Profile configuration.

Run the **show running-config** command to view the configuration.

The following is an example of the **show running-config** command output.

```
show running-config profile wps wps1
   profile wps wps1
   arp 1,4-6,9
   dscp n3 10
   message-priority [ pfcp gtpc ]
   exit
```

Associating WPS Profile under DNN Profile

Use the following sample configuration to associate the WPS profile with the configured DNN profile.

```
config
  profile dnn profile_dnn_name
     wps-profile wps_profilename
     end
```

NOTES:

- **wps-profile** *wps_profilename*: Enables the Wireless Priority Services Profile configuration. This profile is configured under the existing DNN profile configuration.

Verifying WPS Profile under DNN Profile

This section describes how to verify the WPS profile configuration under the DNN profile.

Execute the **show running-config** command to view the configuration.

The following is an example of the **show running-config** command output.

```
show running-config profile dnn intershat
profile dnn intershat
network-element-profiles chf chf1
network-element-profiles amf amf1
network-element-profiles pcf pcf1
network-element-profiles udm udm1
charging-profile chgprf1
virtual-mac b6:6d:47:47:47:47
wps-profile wps1
```

```
ssc-mode 2 allowed [ 3 ]
session type IPV4 allowed [ IPV6 IPV4V6 ]
upf apn intershat
exit
```

Configuration Verification

To view the WPS parameters per subscriber session, use the **show subscriber** command.

The following is an example output of the **show subscriber** command.

```
show subscriber supi imsi-123456789012345
subscriber-details
{
  "subResponses": [
    {
      "status": true,
      "genericInfo": {
        "supi": "imsi-123456789012345",
        "pei": "imei-123456786666660",
        "pduSessionId": 5,
        "pduSesstype": "Ipv4PduSession",
        "accessType": "3GPP_ACCESS",
        "dnn": "intershat",
        "plmnId": {
          "mcc": "123",
          "mnc": "456"
        },
        "sScMode": 1,
        "uetimeZone": "UTC+12:00",
        "allocatedIp": "209.165.200.233",
        "nrLocation": {
          "ncgi": {
            "mcc": "123",
            "mnc": "456",
            "nrCellId": "123456789"
          },
          "tai": {
            "mcc": "123",
            "mnc": "456",
            "tac": "1820"
          }
        },
        "alwaysOn": "None",
        "dcnr": "None",
        "wps": "Wps Session",
        "ratType": "NR",
        "ueType": "NR Capable UE",
        "sessTimeStamp": "2021-05-28 12:46:11.165805357 +0000 UTC",
        "callDuration": "2.925145554s",
        "ipPool": "poolv4",
        "commonId": 11,
        "snssai": {
          "sd": "Abf123",
          "sst": 2
        },
        .
        .
        .
```

WPS OAM Support

SMF Session Gauge Counters

The "wps" label is introduced at the SMF service for session-level gauage counters that support WPS and non-WPS functionality.

For example:

```
smf_session_counters{always_on="disable",app_name="smf",cluster="smf",data_center="unknown",dnn="intershat",
instance_id="0",pdu_type="ipv4",rat_type="NR",service_name="smf-service",ssc_mode="ssc_mode_1",wps="non_wps"}
 10
smf_session_counters{always_on="disable",app_name="smf",cluster="smf",data_center="unknown",dnn="intershat",
instance_id="0",pdu_type="ipv4",rat_type="NR",service_name="smf-service",ssc_mode="ssc_mode_1",wps="wps"}
 20
```

N4 Interface Metrics

The N4 interface counters related to message priority include:

- SESSION_DELETION_REQUEST
- SESSION_ESTABLISHMENT_REQUEST
- SESSION_MODIFICATION_REQUEST

An example of the N4 interface metrics:

```
proto_pfcp_msg_total{app_name="SMF",cluster="Local",data_center="DC",instance_id="0",
message_direction="outbound",message_name="SESSION_DELETION_REQUEST",msgpriority=true,
service_name="protocol",status="accepted",transport_type="origin"} 4
proto_pfcp_msg_total{app_name="SMF",cluster="Local",data_center="DC",instance_id="0",
message_direction="outbound",message_name="SESSION_ESTABLISHMENT_REQUEST",msgpriority=true,
service_name="protocol",status="accepted",transport_type="origin"} 6
proto_pfcp_msg_total{app_name="SMF",cluster="Local",data_center="DC",instance_id="0",
message_direction="outbound",message_name="SESSION_MODIFICATION_REQUEST",msgpriority=true,
service_name="protocol",status="accepted",transport_type="origin"} 20
```

GTPv2 Metrics

The GTPv2 counters related to message priority include:

- NumCreateBearerSuccess
- NumRxCreateBearerRes
- NumTxCreateSessionReq

An example of the GTPv2 metrics:

```
gtpc_app_priority_events{app_name="SMF",cluster="Local",data_center="DC",
event_type="NumCreateBearerSuccess",instance_id="0",interface_type="S5",priority_msg="true",service_name="gtpc-ep"}
 2
gtpc_app_priority_events{app_name="SMF",cluster="Local",data_center="DC",
event_type="NumRxCreateBearerRes",instance_id="0",interface_type="S5",priority_msg="true",service_name="gtpc-ep"}
 2
gtpc_app_priority_events{app_name="SMF",cluster="Local",data_center="DC",
event_type="NumTxCreateSessionReq",instance_id="0",interface_type="S5",priority_msg="true",service_name="gtpc-ep"}
 2
```

KPIs

Following KPIs are supported for this feature:

```
sum(policy_dynamic_pcc_rules_total{pccrule_change_type="binding_param_change",event="attempted"})

sum(policy_dynamic_pcc_rules_total{pccrule_change_type="binding_param_change",event="success"})

sum(policy_dynamic_pcc_rules_total{pccrule_change_type="binding_param_change",event="failure"})
```

Table 3: Statistics for Tracking the Number of Times QCI or ARP is Modified

KPI Name	Type	Description or Formula	Label
policy_dynamic_pcc_rules_total	counter	Total number of dynamic pcc rules added, modified, or deleted as part of different procedures.	pccrule_change_type,status