**Source: Tencent Cloud**

**Title: [EDGE] Triggering discussion on triggering events for edge servers**

**Agenda Item: 8.9**

**Document for: Discussion and Agreement**

# Introduction

The contribution S4aI221282 provided an extension to the M1 interface for provisioning edge servers. One of the parameters in the data resource in this contribution was the activationTrigger parameter which seems to provide conditions for triggering a new edge server.

This contribution proposes an event-based approach for activating the new edge server. The purpose of this contribution is to describe the use-case, the general approach and seek agreement on the approach. The exact solution would be provided depending on the discussion.

# Event-driven triggering of an edge server

The use case is simple. A running edge server may run out of resources to accommodate new traffic or application instances. The new traffic and/or application instances might be due to the increase of users and additional sessions, increase of the application computation complexity, or the networking and connectivity issues where the server is located.

In this cases, it makes sense to start a new edge server for additional traffic or for additional users to avoid increasing the load on the current server. Therefore the use-case can be summarized as:

1. Edge server E1 is running.
2. The traffic increases and the resource on E1 are reaching their limits, and an event is triggered.
3. The event triggers the start of a new edge server E2 with the characteristics that allow diverting additional traffic/load to that server.

This use-case is a prerequist for relocation or loadbalancing of an existing edge sever when the operational parameters requires switching or additional edge servers, such as when the workload reaches cloes to the edge server’s maximum capacity or the location of the users has changed.

# 5GMS AF role

In the content of 5GMS architecture, and considering the M1 provisioning API for EAS servers, the above use case can be translated to the following steps:

1. The Application Provider sets the conditions for generating events in a running edge server E1 during creation or updating the provisioning of the edge server through the M1 interface.
2. The Application Provider sets the condition of activating a new edge server E2 during the creation or updating the provisioning of the edge server through the M1 interface. The server is not activated yet.
3. The 5GMS AF subscribes to the events defined by E1 during its provisioning.
4. When any of the conditions in E1 reaches the set limits defined during step 1, the 5GMS AS/EAS issues an event.
5. Since the 5GMS AF is subscribed to the 5GMS AS/E1 events, it receives the event issued at step 4.
6. The 5GMS AF checks the received specific event meets any activation condition for E2. If that condition is met, the 5GMS AF activates the E2 server.

Note: the same process, particularly steps 3-6, can be driven by Media Session Handler, for the client-driven edge management.

# EAS Triggering event

The EAS triggering events can be divided into two classes:

1. EAS generic KPIs
2. Media related KPIs

The EAS profile defined in 29.558 defines a set of EAS generic service KPIs:

Table 8.1.5.2.4-1: Definition of type EASServiceKPI

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Attribute name | Data type | P | Cardinality | Description |
| maxReqRate | string | O | 0..1 | Maximum request rate from the Application Client supported by the EAS. |
| maxRespTime | Uinteger | O | 0..1 | The maximum response time, in the units of milliseconds, advertised for the AC’s service requests. This includes the round trip time of the request and response packet, the processing time at the EAS and time required by EAS to consume any 3GPP core network capabilities. |
| avail | string | O | 0..1 | Advertised percentage of time the server is available for the AC's use. |
| avlComp | string | O | 0..1 | The maximum compute resource available for the AC. |
| avlGraComp | string | O | 0..1 | The maximum graphical compute resource available for the AC. |
| avlMem | string | O | 0..1 | The maximum memory resource available for the AC. |
| avlStrg | string | O | 0..1 | The maximum storage resource available for the AC. |
| connBand | BitRate | O | 0..1 | The connection bandwidth in Kbit/s advertised for the AC's use. |

The attributes in this table seem suitable as triggering parameters. An event can be generated if one or more attributes of this table reach certain values.

Additionally, the following parameters can be considered for generating events:

1. Increase of the number of uses using 5GMS AS/EAS
2. Increase the number of requests to this EAS
3. Increase the number of users in the same geographical area

As for the media specific KPIs, there are parameters that are media specific. For instance, one parameter could be a new media function capability such as hardware assisted transcoder. Further investigation is needed for identifying such KPIs.

# Discussion

The proposed data resource for provisioning EAS through M1 is as the following:

Table 7.10.3-1: Definition of EdgeResourcesConfiguration resource

| Property name | Type | Cardinality | Description |
| --- | --- | --- | --- |
| *edgeResourcesConfigurationId* | ResourceId | 1..1 | An identifier for this Metrics Reporting Configuration that is unique within the scope of the enclosing Provisioning Session. |
|  edgeManagementMode | EdgeManagmentType | 1..1 | Indicates if the management of the edge resource session is client-driven or application provider-driven. |
|  activationTrigger | Activation‌Trigger‌Type | 0..1 | Condition to activate edge resources for this ProvisioningSession. |
|  profile | EASProfile | 1..1 | The EAS profile used by the 5GMS AF or by the EEC to discover and select one or more EAS instances to serve media streaming sessions.The format of the EASProfile is defined in table 8.1.5.2.3-1 of TS 29.558 [?]. |
|  application‌Context‌Relocation‌Requirements | ACR‌Requirements‌Type | 0..1 | Application Context Relocation tolerance and requirements. |

We would like to raise the following questions:

1. Does the activationTrigger property include the events to trigger activating this EAS?
2. Do we need a new property. e.g. triggeringEvents, for setting KPIs that generate events by this EAS?
3. Can the EVAX work item address provisioning, collection, and activation of the above events?

# Proposal

1. Agree whether we would like to address such use-case, i.e the need for an event driven API between 5GMS AS/EAS and 5GMS AF/EES
2. Send a liaison to SA5 and SA6, outlining the use-case of event-driven APIs for EAS and EES and whether they would consider to address it
3. Use of the highlevel media edge KPIs (8.1.5.2.4-1 + additional parameter) in M1 for edge provisioning.
4. Agree on an approach to solve address this use-case to bring proposals to the next meeting/adhoc calls