|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change Request** | | | | | | | |
| **Document** | **ORAN-WG6.AAL-GAnP** | **ver** | **00.01.01** | **CR** | **NVD-002** | **rev** | 5 |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Title:*** | AAL Definitions – to align with Greenfield Agreements | | |
| ***Source to WG:*** | NVIDIA | | |
| ***Target WG :*** | **WG6** | | |
| ***Category:*** | **B** | ***CR Creation Date*** | October 25, 2021 |
|  | *Use one of the following* ***categories****:* ***A*** *(mirror corresponding to a change in an earlier release)* ***B*** *(addition of feature),* ***C*** *(functional modification of feature)* ***D*** *(editorial modification)* ***F*** *(correction)*  Detailed explanations of the above categories can be found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | | |

|  |  |
| --- | --- |
| ***Reason for Change:*** | To align AAL definitions with Greenfield Agreements and incorporate the decision made on (10/25/21) AAL call to generaize Accelerator type (beyond HW only accelerator) |
| ***Summary of change:*** | New text is proposed and can be reviewed by track change in the text below |
| ***Consequences if not aproved:*** | If not included, AAL Definitions will not be consistent with AAL Greenfield Agreements |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Clauses affected:*** | <list specific document sections impacted by the CR> | | | | |
|  | **Y** | **N** |  | |  |
| ***Other specs*** |  | **X** | Other core specifications: | <fill in related CRs if “Y”> | |
| ***affected:*** |  | **x** | Test specifications: | <fill in related CRs if “Y”> | |
| ***(show related CRs)*** |  | **X** | O&M Specifications: | <fill in related CRs if “Y”> | |
| ***Supporting material:***  ***Other comments:*** | <provide file name or URL of any material supporting this CR> | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Status:*** |  | ***CR Closed Date:*** |  |
| ***Outcome:*** |  | ***Duplication:*** |  |
| ***Outcome explanation:*** |  | | |

The proposed changes are indicated by Track Changes in the text below.

## Definitions and Abbreviations

### Definitions

For the purpose of this document the terms and definitions given in ETSI GS NFV-IFA 002 [4], ETSI GS NFV-IFA 004 [10] and the following apply:

**Hardware (HW) Accelerator** is a specialized HW implementation that can offload processing from application(s) running on General-Purpose Processor

NOTE 1: Examples of Hardware Accelerators include ASIC, FPGA, DSP and GPU.

NOTE 2: Throughout this document, the term “Accelerator” and “Hardware (HW) accelerator” are used interchangeably.

**Acceleration Abstraction Layer (AAL)** specifies a common and consistent interface between an application and underlying different types of HW accelerators within an O-Cloud instance.

**Acceleration Abstraction Layer Interface (AALI)** consists of a programming API and associated information models between an application and an AALI implementation within an O-Cloud instance.

**AALI Implementation** is a realization of an AAL interface including but not limited to the software libraries, drivers, and Hardware Accelerator

NOTE: AAL specification does not preclude AALI implementation of an AAL profile to be fully or partially SW based implementation. This version of the specification, however, is tailored towards HW based or hybrid (HW+SW) AALI implementation.

**Accelerated Function (AF)** is a representation of a workload building block that an accelerator processes on behalf of an application within an O-RAN Cloudified Network Function

**AAL Profile** specifies a set of Accelerated Functions that an accelerator processes on behalf of an application within an O-RAN Cloudified Network Function.

**AAL Profile API**s are a subset of the AALI that supports a specific set of Accelerated Functions defined by an AAL Profile.

**An Operation** is the action applied to input data which is processed in an AAL-LPU producing output data based on the AAL profile supported by the AAL-LPU.

**AAL Logical Processing Unit (AAL-LPU)** is a logical representation of resources within an instance of a HW Accelerator (example: there can be multiple processing units or subsystems on a hardware accelerator, or resource partitioning (hard – dedicated resources, soft – soft resources) and these can be logically represented as an AAL Logical Processing Unit)

* An AAL-LPU maps to a single HW Accelerator
* A HW Accelerator may support 1 to N AAL-LPU’s
* Each AAL-LPU shares the resources of the associated HW Accelerator with other AAL-LPU(s) mapped to the same HW Accelerator. AAL-LPU can also represent a hard partition of the HW accelerator where resources are dedicated to the partition.
* Mapping of HW Accelerator resources to AAL-LPU shall be configurable from O2 interface
* An AAL-LPU may support more than one AAL profile(s)

**AAL Queue** is part of specific AAL profile APIs and is defined as an abstract construct that is used by the application to group operations together and may access specific resources (compute, I/O) of an AAL-LPU supporting specific AAL profile(s).

* From the application point of view, each AAL-LPU supporting specific AAL profile(s) consists of one or more AAL queues
  + While an AAL-LPU may support multiple AAL profiles, an AAL queue supports only one type of AAL profile
  + AAL Queue optionally also supports priority, allowing the application/network function to schedule jobs of different priorities to the AAL-LPU

NOTE:

* An AAL queue can be used by an application/network function to share AAL-LPU resources between threads/cores belonging to the same process address space
* An application/network function may use multiple AAL queues to access different AAL profiles supported by an AAL-LPU

**AAL Queue ID** is a unique index used to designate the AAL Queue in functions exported by specific AAL profile APIs.

NOTE: An AAL Queue or an AAL Queue ID does not reflect a HW design or an AALI Implementation specification

**HW Accelerator Manager** is an acceleration management function, that provides management capabilities for the HW Accelerator(s) in the O-Cloud Node. Management capabilities include but not limited to lifecycle management, configuration, updates/upgrades and failure handling.