Ultra Packet Core CUPS Overview Guide, Release 21.26

First Published: 2021-12-24

CONTENTS

About this Guide

Note
Control and User Plane Separation (CUPS) represents a significant architectural change in the way StarOS-based products are deployed in the 3G, 4G, and 5G networks. This document provides information on the features and functionality specifically supported by this 3G/4G CUPS product deployed in a 3G/4G network. It should not be assumed that features and functionality that have been previously supported in legacy or non-CUPS products are supported by this product. References to any legacy or non-CUPS products or features are for informational purposes only. Furthermore, it should not be assumed that any constructs (including, but not limited to, commands, statistics, attributes, MIB objects, alarms, logs, services) referenced in this document imply functional parity with legacy or non-CUPS products. Please contact your Cisco Account or Support representative for any questions about parity between this product and any legacy or non-CUPS products.

Note
The documentation set for this product strives to use bias-free language. For purposes of this documentation set, bias-free is defined as language that does not imply discrimination based on age, disability, gender, racial identity, ethnic identity, sexual orientation, socioeconomic status, and intersectionality. Exceptions may be present in the documentation due to language that is hardcoded in the user interfaces of the product software, language used based on RFP documentation, or language that is used by a referenced third-party product.

This document provides an overview of Control and User Plane Separation (CUPS), an evolution of 3GPP Evolved Packet Core (EPC) architecture in which S-GW and P-GW are separated into their constituent User Plane and Control Plane functions.

-

Conventions Used

The following tables describe the conventions used throughout this documentation.

Notice Type	Description
Information Note	Provides information about important features or instructions.
Caution	Alerts you of potential damage to a program, device, or system.

Notice Type	Description
Warning	Alerts you of potential personal injury or fatality. May also alert you of potential electrical hazards.

Typeface Conventions	Description
Text represented as a `screen display`	This typeface represents displays that appear on your terminal screen, for example: `Login:`
Text represented as **commands**	This typeface represents commands that you enter, for example: **show ip access-list** This document always gives the full form of a command in lowercase letters. Commands are not case sensitive.
Text represented as a **command** *variable*	This typeface represents a variable that is part of a command, for example: **show card** *slot_number* *slot_number* is a variable representing the desired chassis slot number.
Text represented as menu or sub-menu names	This typeface represents menus and sub-menus that you access within a software application, for example: Click the **File** menu, then click **New**

CHAPTER 1

Overview

Control and User Plane Separation (CUPS) is the evolution of 3GPP Evolved Packet Core (EPC) architecture in which S-GW and P-GW are separated into their constituent User Plane and Control Plane functions. CUPS allows more flexibility and independent scalability that is suitable for implementing Software-defined Networking (SDN) and Network Function Virtualization (NFV). CUPS also helps maintain the mobility control provided by GPRS Tunneling Protocol (GTP), which is retained between the evolved nodes.

Cisco Ultra Services Platform (USP) with CUPS

The Cisco Ultra Packet Core (UPC) is the Virtual Packet Core (VPC) solution that brings together Cisco's UPC and Service Functioning Chaining (SFC) technology into a single, integrated virtual network function (VNF). It also brings a set of tools to automate and simplify the on-boarding, instantiation, and operation of this tightly coupled collection of technologies.

In compliance with 3GPP standard architectural enhancements, Cisco enhanced the operation of the EPC through the separation of Control and User Plane functions. As part of CUPS, Packet Gateway application is split into independent components—Control Plane and User Plane. Cisco CUPS solution leverages the SAEGW, which is an optimized combination of S-GW and P-GW. The SAEGW-C is the Cisco UPC CUPS Control Plane (CP) and SAEGW-U is the Cisco UPC CUPS User Plane (UP).

Advantages and Use Cases of CUPS

Following are the advantages and use cases of Cisco UPC CUPS:

- Ability to independently scale Control and User Planes per Mobile network requirements.

- Data center costs can be optimized by hosting Control and User Planes in different data centers, each catering to the specific SLA needs of CP and UP.

- Savings on backhaul cost by terminating the data at the edge.

- Flexibility to have specialized UPs for different applications.

- Multi-Gigabit throughput per PDN.

- Ability to cater to ultra-low latency use cases.

- Enables Mobile Edge Computing.

- 3GPP compliant CUPS solution.

- Flexible CUPS CP which can handle any mix of CUPS UPs.

- 5G-ready solution with ability to support 5G data rates, and aligned CUPS architecture.

- Full-fledge inline SPI/DPI feature capabilities.

- Simplified CUPS management with CUPS CP being the one stop function for all the configurations, statistics, logs, and Lawful Intercept management.

- Flexible CUPS Lawful Interception solution with option to locate X3 aggregator function either with CUPS CP, CUPS UP, or as an independent function.

- Automated method for performing IP Pool Management for various redundancy models and different use cases, and APN (VoLTE/Internet).

Platform Requirements

The Cisco UPC CUPS solution is based on the StarOS-based VPC Platforms:

- CUPS CP: Can be deployed on either VPC-Distributed Instance (VPC-DI) or VPC-Single Instance (VPC-SI)

- CUPS UP: Can be deployed on VPC-SI

The following table provides information about the minimum hardware requirement for each compute node that is part of the CUPS deployment regardless of whether VPC-DI or VPC-SI is used.

Table 1: Minimum Hardware Requirement

Hardware	
Host Server	UCS C220 M5
Processor	2 $\times$ Intel® Xeon® Gold 6148 (20 cores @ 2.4GHz)
NIC	2 $\times$ Intel® XL710 Dual Port 40G QSFP+
RAM	384GB
Local Disk Storage	2 $\times$ 1.6TB SSD RAID-1

Note For information on deploying CUPS, contact your Cisco Account representative.

VNF Tenant Networks

The following figure displays the types of networks typically required for CUPS.
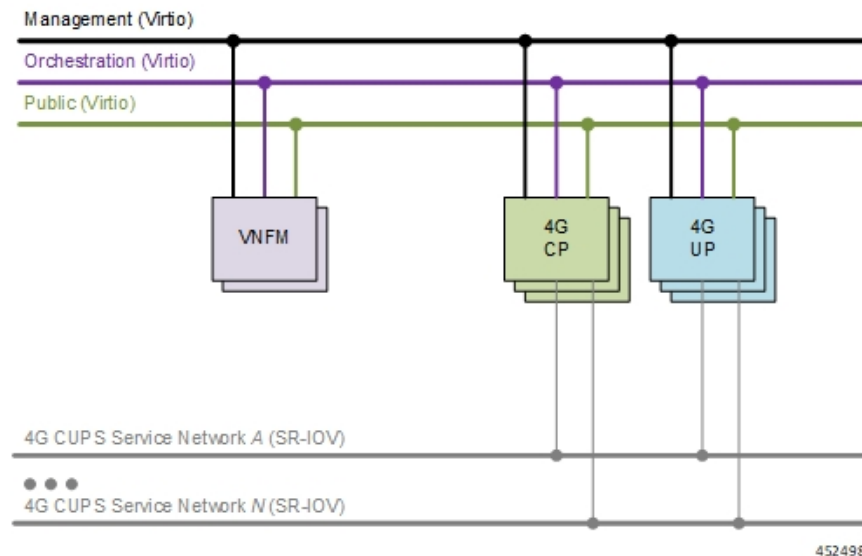
The VNFM networking requirements and the specific roles are described here:

- **Public Network**: Provides external connectivity to the VM. A floating IP from the public network is typically associated to the HA VIP. A floating IP from the public network can be assigned or associated to the virtual-link.

- **Management Network**: The VIRTIO/VLAN network used for management access to the VMs.

- **Orchestration Network**: The VIRTIO/VLAN network used by the VNFM and the VNF components for automation, monitoring, and orchestration.

- **UP Redundancy Network**: The SR-IOV flat network used by CUPS allow CP and UP in ICSR redundancy.

- **Additional networks**: The SR-IOV flat networks that are required by 4G VNF service interfaces.

Licensing Requirements

The Cisco UPC CUPS solution requires specific license(s). Contact your Cisco Account representative for detailed information on specific licensing requirements. For information on installing and verifying licenses, refer to the *Managing License Keys* section of the *Software Management Operations* chapter in the *System Administration Guide*.

Standards Compliance

The Cisco UPC CUPS solution complies with the following standards:

- 3GPP TS 23.401: General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access.

- 3GPP TS 23.214: "Architecture enhancements for control and user plane separation of EPC nodes; Stage 2".

- 3GPP TS 29.244: Interface between the Control Plane and the User Plane of EPC - Stage 3.

Architecture

The following diagram represents the Cisco UPC CUPS architecture.

Figure 1: CUPS Architecture
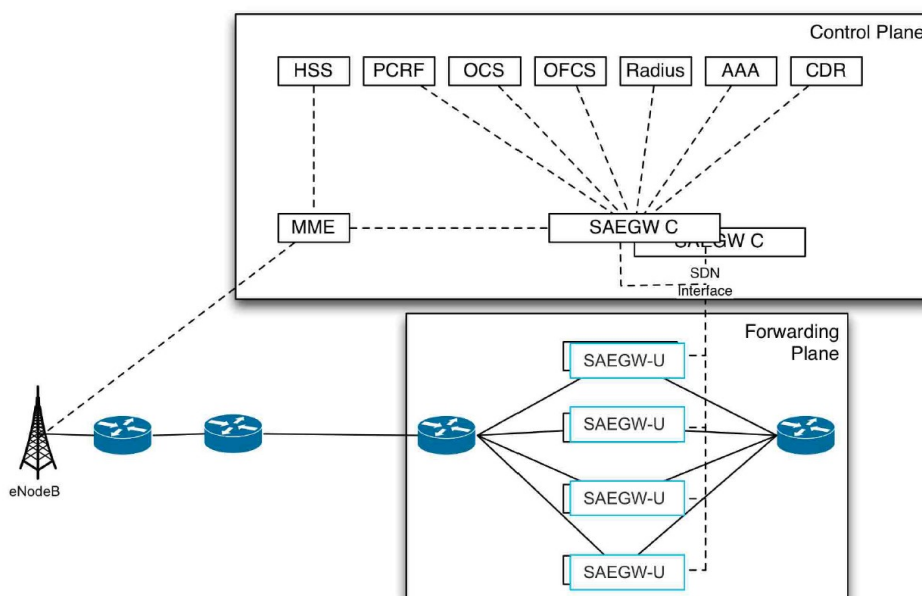
CPs and UPs in the Cisco UPC CUPS solution operate as independent VNFs that can be independently scaled up or down as needed.

CUPS CP (SAEGW-C)

The SAEGW-C can be hosted in Cisco USP platform on on hardware meeting the minimum requirements identified in the *Platform Requirements* section. The SAEGW-C can control multiple User Planes, irrespective of where they are located and what platform they are hosted on. The SAEGW-C can, therefore, control the mix of different types of SAEGW-Us. The SAEGW-C aggregates/consolidates statistics/bulkstats, logging, LI interception, pushes required node configuration (including predefined/static rules, APN profile information, ACL information, and others) for all the SAEGW-Us that it controls. Thus, the SAEGW-C helps in simplifying the complexity of CUPS solution management for operators.

The SAEGW-C supports 3GPP standards-based Sx protocol to interact with a User Plane. 3GPP Sx in its standard form doesn't cover all of the capabilities that a gateway needs. And so, private extensions are added

to support custom features with CUPS. For information about private extensions, refer the *Ultra Packet Core CUPS Sx Interface Administration and Reference Guide*.

CUPS CP Functionality

The CUPS CP has the following functionality:

- Session Subsystem

- IP Pool Subsystem

- AAA and Redundancy Subsystem

Session Subsystem

Functionality	Description
UP Association	Supports the Sx Node-level association procedures.
Packet Flow Description (PFD) Management	Currently, PFD Management uses custom information element (IE). Packet filter detection is a Sx Node-level procedure to program the UP with static and predefined rules of Enhanced Charging Service (ECS).
	The procedure is also used to push other configurations to program the UP, such as Access Control Lists (ACLs), and so on.
Session Management	Supports integration of Sx session-level procedures and programming of UP with the Sx parameters. Session Management also handles Sx reports from the UP.
UP Selection and Grouping	Supports selection of the UP for a particular session, grouping of applicable UPs into default or specific groups, IP Pool assignment, different redundancy schemes for different groups (for example, 1:1 redundancy for VoLTE/IMS, and N:M redundancy for internet).
Charging/Accounting	Supports the handling of usage reports from Sx and align them to charging interfaces, such as Gz/CDR, Gy, and so on.
LI	Supports LI functionality on CP for X1 and X2, and programming of UP through Sx interface for UP LI (X3).
CP Redundancy	Supports session recovery and ICSR for the CP.
Diameter, GTP	Provides Diameter stack and GTP stack functionality.
Sx-C	Provides the Sx CP functionality, that is the Node-level and session-level procedures as described in the 3GPP TS 29.244.
Sx-U	Provides the tunnel between the CP and UP. Sx-U functionality is used for Router advertisement (RA)/Router Solicitation (RS) exchange for IPv6.
Load/Overload	Supports detecting and handling of load/overload information received over the Sx interface.

IP Pool Subsystem

Functionality	Description
IP Pool Management	Supports an algorithm for distributing IP chunks across UPs and minimze the wastage of addresses incurred with respect to IP Pool as a resource. The algorithm automatically manages the IP chunks across multiple UPs.

AAA and Redundancy Subsystem

Functionality	Description
AAA	Provides the functionality for interfacing with AAA systems for authentication and accounting.
Redundancy	Supports CP redundancy, specifically the use of the Inter-chassis Session Redundancy Protocol (ICSR) used between CPs.
UP Redundancy	Supports storing of checkpoints from the UP and handling the logic for 1:1 and N:M UP redundancy.

For more information about CUPS CP (SAEGW-C) and supported features, see the *Ultra Packet Core CUPS Control Plane Administration Guide*.

CUPS UP (SAEGW-U)

The SAEGW-U VNF can be hosted in Cisco USP platform on hardware meeting the minimum requirements identified in the *Platform Requirements* section. The SAEGW-U can be co-located with SAEGW-C in the same data center or can be located remotely in a different data center. For more information see, *Ultra Packet Core CUPS User Plane Administration Guide.*

CUPS UP Functionality

The CUPS UP has the following functionality:

- Session Subsystem

- IP Pool Subsystem

- Redundancy Subsystem

Session Subsystem

Functionality	Description
Session Manager_User Plane (SM_U)	The SM_U supports Sx session-level programming instructions from the CP to create and manage user sessions. The SM_U also classifies the initial packets from the flow and programs the SM_P/VPP with instructions before offloading the flow to the SM_P/VPP.
Session Manager_Plugin (SM_P)	SM_P runs on VPP (multi-threaded) and supports the offloaded flows. To provide high throughput, the SM_P installs instructions in the form of flow/stream and its corresponding QoS, charging, and forwarding actions provided by the SM_U, and then executes the same.

Functionality	Description
FastPath Execution Engine and VPP OS	These provide the platform framework that is required for VPP and the Mobility features on which the SM_P fits in with its application logic.

IP Subsystem

Functionality	Description
IP Pool Management	Supports advertising the BGP routes for the IP Chunks received from the IP Pool management algorithm. It also validates the IP address amongst other functionalities.

Redundancy Subsystem

Functionality	Description
Redundancy	Supports the functionality for UP session recovery, and 1:1 and N:M redundancy aspects of UP.

CUPS UP Event Data Records (EDRs)

The CUPS UP node can be configured to generate EDRs. When enabled, the UP instance generates individual records for each IP flow of a subscriber. They can be configured to be generated at the end of a transaction (for example, HTTP GET/POST request) or when the IP flow is terminated. The information to be included in these records is also configurable. The generated records are temporarily stored on the UPs local storage which later uses SFTP to push the generated records to an external storage/server, as configured. For more information, refer *Event Data Records in CUPS* chapter in the *Ultra Packet Core CUPS User Plane Administration Guide*.
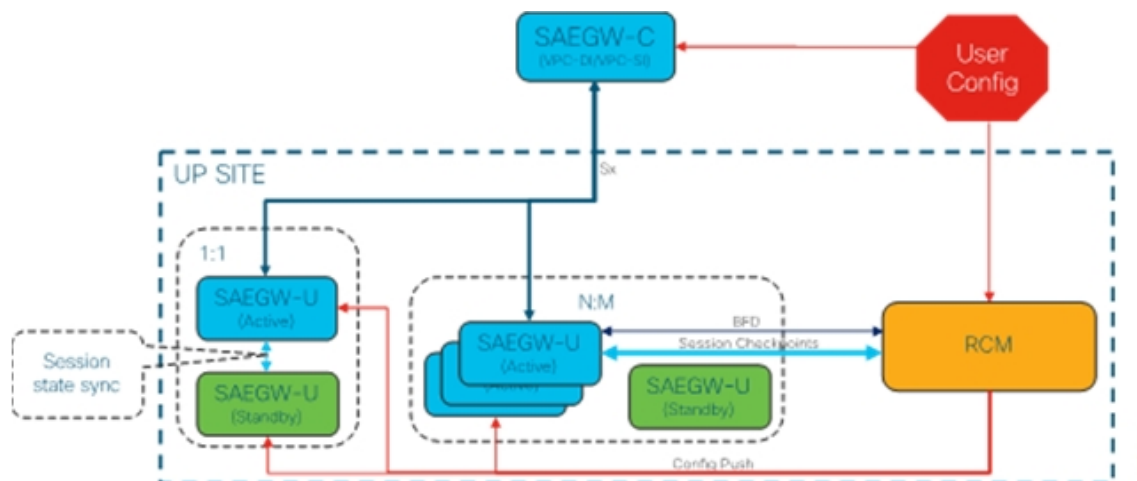
CUPS UP Redundancy

The following two types of redundancy are supported in CUPS UP:

- 1:1 UP Redundancy

- N:M UP Redundancy

The following diagram illustrates how UP nodes can be deployed redundantly in the network.

Figure 2: CUPS UP Redundancy Deployment Architecture

Redundancy and Configuration Manager

Redundancy and Configuration Manager (RCM) is a Cisco proprietary node that provides:

- Configuration management of UPs

- Session state storage of information from all UPs it serves

- UP monitoring that are deployed in N:M redundancy mode

- Failure detection and session state restoration of the standby UP

For more information about RCM, refer *Redundancy and Configuration Manager Configuration and Administration Guide*.

1:1 UP Redundancy

In 1:1 deployment mode, the UPs are deployed in hot-standby mode. In this mode, the RCM is required only for configuring the UPs. A propriety protocol over TCP (called Session Redundancy Protocol – SRP) is used between the two UPs which negotiates the Active-Standby state and also for monitoring each other.

After the state negotiation is done, a peer-to-peer TCP connection is established between the Session Managers on the active UP with the Session Managers on the standby UP for exchanging session state information. The information passed as part of session state includes:

- Call/Session ID

- Peer CP address

- User ID information (IMSI, MSISDN)

- Traffic Endpoint address (QCI, eNodeB address, and so on.)

- APN-MBR

- Rules installed for the subscriber

- Accounting/Usage information

- Statistics

For more information about 1:1 UP redundancy, refer the *1:1 User Plane Redundancy for 4G CUPS* chapter in the *Ultra Packet Core CUPS User Plane Administration Guide*.

Note Currently, it's recommended that 1:1 UP redundancy be used for UPs supporting VoLTE/IMS traffic whereas UPs supporting non-VoLTE/IMS data traffic can implement either 1:1 or N:M UP redundancy.

N:M UP Redundancy

In N:M deployment mode, the UPs are deployed in cold-standby mode. In this mode, the RCM is required for configuring and monitoring the UPs, state management of the UP instances, storing, and restoration of session state when an active UP fails.

The interface between the UP instance and the RCM is a propriety protocol over TCP, same as the one used in 1:1 redundancy. The Bidirectional Forwarding Detection (BFD) protocol is used to monitor and detect failure of the UP instances. A peer-to-peer TCP connection is established between the Session Managers on the active UP instance and the corresponding Checkpoint Managers in the RCM for exchanging session state information.

The session state information exchanged between the UP and the RCM is the same as the information exchanged in 1:1 redundancy mode. The Checkpoint Managers store the received session state information in memory. Upon detecting the failure of an active UP, the stored checkpoint information is pushed to the Session Managers on the standby UP and is used for restoring the session state.

For more information about N:M UP redundancy, refer the *N:M Redundancy* chapter in the *Ultra Packet Core CUPS User Plane Administration Guide*.

UPC CUPS Sx Interface

Sx is the interface between the Control Plane and User Plane in a split P-GW and S-GW architecture in an Evolved Packet Core (EPC) that provides Packet Forwarding Control Protocol (PFCP) service. One of the main tasks of the Sx interface is to enable the Control Plane function to instruct the User Plane function about how to forward user data traffic.

The features and functionality that are part of Sx includes:

- Heartbeat procedures

- Packet Detection Information (PDI) optimization

- Sx over IPsec

For more information, refer *Ultra Packet Core CUPS Sx Interface Administration and Reference Guide.*

UPC CUPS VPP for Mobility

The Vector Packet Processor for Mobility (VPPMOB) is a mobility-centric solution based on FD.io (Fast data – Input/Output) VPP. It utilizes Cisco Vector Packet Processing technology to provide a high performance, packet processing engine (forwarding plane) for virtualized deployments.

VPP is a software-based network processing unit (NPU) that supports an extensible framework that provides out-of-box production quality switch and router functionality. It is a high performance packet processing stack that can run on commodity CPUs.

Architecture

VPP is a set of directed graph nodes. Each one of the nodes in the graph performs a particular step in packet processing. Each node is optimized for a single function. For example, there is a graph node for IP forwarding information base (FIB) lookup, Ethernet processing, and so on. Therefore, the packet processing function is split a series of graph nodes. The unique functionality of VPP is that it processes a vector (also called a "frame") of up to 256 packets at a time. This implies that if the vector has 256 packets, these 256 packets run through IP FIB lookup, 256 packets run through User Datagram Protocol (UDP) processing, and so on.

At the beginning of processing, the driver (typically, Data Plane Development Kit (DPDK)) polls the wire and groups all the packets in a vector and passes it to the first step in the processing, the first graph node. For example, Ethernet processing could be the first graph node, likewise IPv4/IPv6, and so on. Based on the processing path, the packets move to the next processing stage. The packets move through the pipeline where the pipeline is divided according to the processing state. Thereby following a pipeline architecture.

With VPP as the core and VPPMOB as a set of VPP plugins, the solution runs in a single VM on subset of available cores. Packet processing is performed on the incoming packets from the service ports with some subset relayed to and from the kernel and StarOs (NPUSIM, SM_U, and so on).
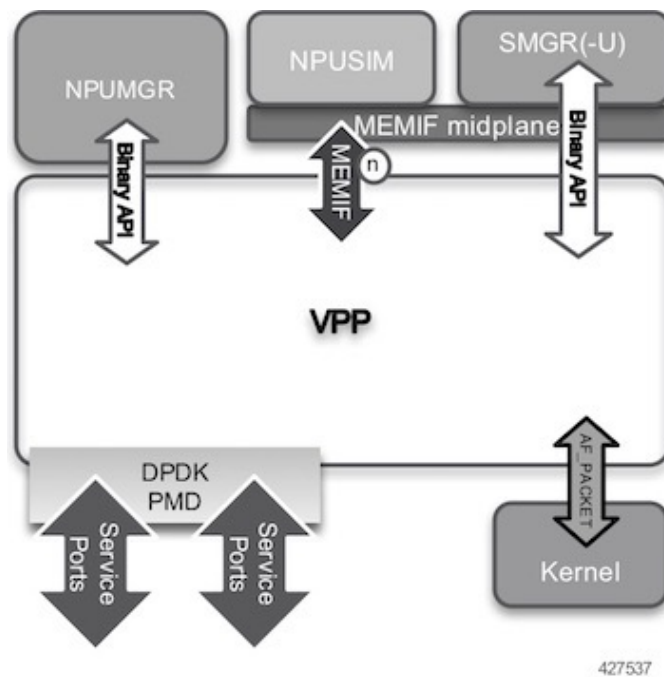
As depicted in the illustration, the service ports are connected to the DPDK drivers and the kernel is connected through an interface called "AF_Packet", and these (service ports and kernel) are connected to the VPP.

Other subsystems communicate with VPP using the Binary API and MEMIF interfaces. The NPUMGR controls VPP, NPUSIM is an adjunct processing layer to augment the functionality of VPP, and Session Manager is responsible for subscriber management.

Binary APIs not only support communcation between the components and VPP but also enable the exchange of hex data, which is the messaging format that is used for communication.

MEMIF is a shared memory-based memory packet interface that the NPUSIM and SM_U use to exchange packets with VPP. Each Session Manager application instance has a MEMFIF interface.

CUPS on VPC-SI

The Control Plane (CP) contains all the configurations, such as subscriber policies, traffic classification rules, and so on, which are running on a VPC-DI/VPC-SI instances.

The UP consists of one or more VPC-SI instances and can be co-located with the CP or deployed remotely at the edge of the network. Each VPC-SI instance contains a forwarding element that uses VPP and SM_U software tasks to handle the subscriber data sessions. The SM_U classifies the traffic and identifies the policies that need to be applied for the subscriber and programs the forwarding element before offloading data processing.

In this model, VPPMOB executes on a VPC-SI system that provides a mobility-forwarding solution, with both NPU-style flow processing and IP forwarding capabilities. It also executes a SM_P Fastpath thereby offloading SM_U. Both the NPUMGR and SM_U share the responsibility of configuring and managing VPPMOB, with SM_U responsible for Fastpath-related configuration.

For more information, refer *VPP Support* chapter in the *Ultra Packet Core CUPS Control Plane Administration Guide* or the *Ultra Packet Core CUPS User Plane Administration Guide*.

Deployment Scenarios

You can deploy the Cisco UPC CUPS SAEGW-C and SAEGW-U either as:

- P-GW only

- S-GW only

- SAEGW

The SAEGW-C and SAEGW-U can anchor any combination of following type of sessions:

- Pure S-GW (Pure-S)—When a UE uses S-GW part of SAEGW and the PDN connection is terminated at an external P-GW (not part of SAEGW).

- Pure P-GW (Pure-P)—When a UE uses an external S-GW (not part of SAEGW) and the PDN connection is terminated within P-GW part of SAEGW.

- Combined S-GW and P-GW—When a UE uses both S-GW and P-GW part of the same SAEGW service.

The Cisco UPC CUPS can be deployed in the following ways:

- Co-Located CUPS

- Remote CUPS

Co-Located CUPS

In Co-located CUPS, both the SAEGW-C and SAEGW-U (multiple) VNFs are located in the same data center, and are part of the same deployment instance.
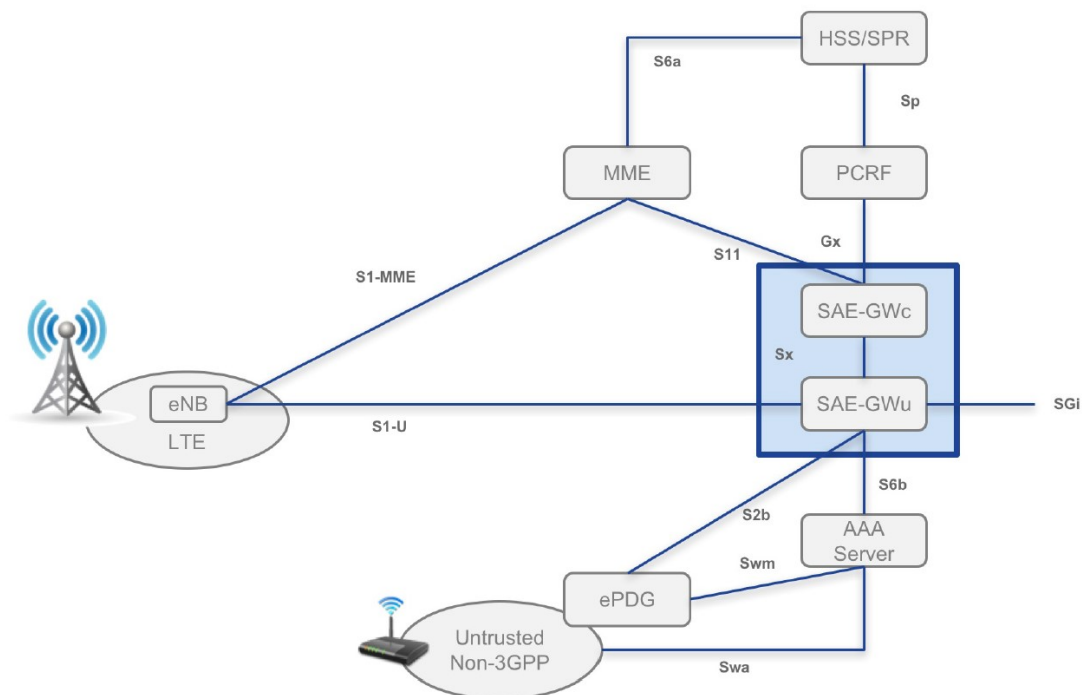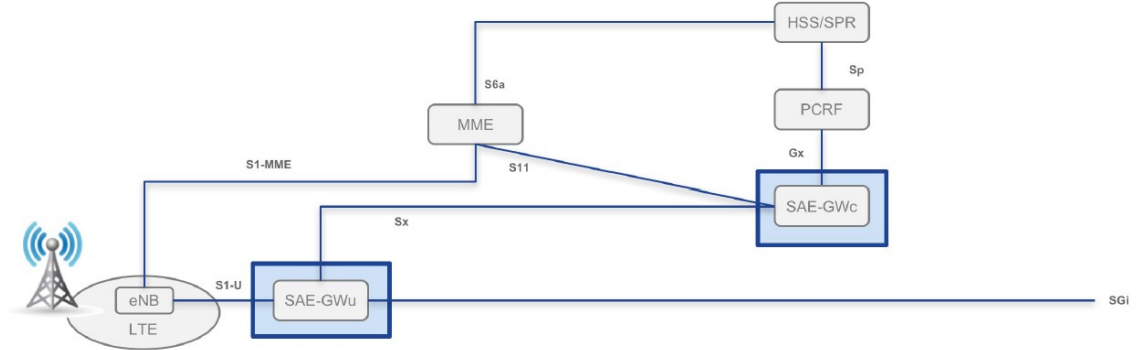
Remote CUPS

In Remote CUPS, the Control and User Planes are deployed in geographically separated data centers. Remotely deploying the User Plane function of the UPC CUPS architecture allows a mobile operator to continue to operate it as a single EPC Gateway, yet place the capability of the UP very close to traffic termination points. By locating an EPC Gateway User Plane at key aggregation or termination points, it significantly changes the amount of traffic the mobile operator must otherwise backhaul to a central location. This benefit becomes especially significant in enterprise, video content delivery, and IoT use cases whereby large volume of traffic terminates at given customer locations.

The CUPS CP can control both co-located and remotely located CUPS UPs, at the same time, providing flexibility to have a mix of different types of CUPS UP deployments controlled by the same CUPS CPs. Therefore, Mobile Edge Computing (MEC) requirements are addressed with Remote CUPS.

The following diagram shows the distribution of packet core functions in the remotely deployed user plane use case.

Benefits and Use Cases of Remote CUPS

The following are the benefits and use cases of Remote CUPS:

- Backhaul cost reduction
- Reduced latency and improved user experience
- Enhanced Enterprise/Corporate use cases
- Mobile Edge Computing use cases
- Operational efficiency