Cisco APIC OpenStack Plug-in Release Notes, Release 6.0(4)

Introduction

This document describes the features, bugs, and limitations for the Cisco Application Policy Infrastructure Controller (APIC) OpenStack Plug-in.

Cisco APIC OpenStack Plug-in allows policy deployment automation across Cisco Application Centric Infrastructure (ACI) and OpenStack, enabling a complete undercloud and overcloud visibility on Cisco ACI. The Cisco APIC OpenStack Plug-in allows dynamic creation of networking constructs to be driven directly from OpenStack, while providing extra visibility and control from the Cisco APIC.

Release notes are sometimes updated with new information about restrictions and bugs. See the following website for the most recent version of this document:

https://www.cisco.com/c/en/us/support/cloud-systems-management/application-policy-infrastructure-controller-apic/tsd-products-support-series-home.html

For more information about this product, see " Related Content."

Note: The documentation set for this product strives to use bias-free language. For the 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.

Date	Description
February 29, 2024	Release 6.0(4) became available.

New Software Features

Following are the new features introduced in the 6.0(4) plug-in release:

- The policy for an opflex-agent is currently only kept in the program's memory, and isn't preserved across agent restarts. This means that the agent must re- request its policy from the fabric. Certain OpenStack cloud operations, such as upgrades, cause all opflex agents to be restarted almost simultaneously, making them request their entire policy model at nearly the same time. An excessive policy request load on the fabric results in delayed policy resolutions, which leads to periods where the data plane is inconsistent, triggering outages. To avoid this, a policy persistence feature has been added to the opflex-agent. Before performing operations that trigger large scale agent restarts, a snapshot of the agent policy is made by the cloud administrator. This is then used as the intial policy for the agent, allowing it to render a consistent data plane and reduce the policy resolution load on the fabric. Once the data plane has been successfully rendered, the agent transitions to using the fabric for any new or refresh policy requests.

  The OpflexEnableStartupPolicy controls whether this feature is used at all. If the OpflexEnableStartupPolicy is set to false, or if there is no value configured for OpflexStartupPolicyFile, then the agent reverts to initial policy resolution with the fabric.

The OpflexStartupPolicyDuration should be long enough to ensure that the host data plane can be programmed from the persisted policy. The default value of 60 seconds is intended to cover most, if not all cases.

Two ansible playbooks have been included that help create and delete policy snapshots: opflex_snapshot.yaml and remove_opflex_policy.yaml. The opflex_snapshot.yaml should be used before triggering agent restarts, and the remove_opflex_policy.yaml should be used to delete the policy file after the agent has rendered the data plane and is resolving new policy from the fabric.

```
      OpflexEnableStartupPolicy:
        default: true
        description: Enable the use of an initial policy file on startup
        type: boolean
      OpflexStartupPolicyFile:
        default: /var/lib/opflex-agent-ovs/policy/pol.json
        description: File to use for initial policy
        type: string
      OpflexStartupPolicyDuration:
        default: 60
  description: Time to wait after connection before resolving policy from peer
        type: number
```

- A new extension has been created for the subnet resource in OpenStack. This extension is used to indicate whether a subnet in OpenStack should be used only for router gateway IPs. The extension has enable and disable flags:

  ```
  --apic_router_gw_ip_pool_enable
  --apic_router_gw_ip_pool_disable
  ```

  The extension is intended for use on subnets that belong to external networks in OpenStack. When a user connects a router gateway to an external network in OpenStack, the plugin checks to see if there is a subnet on the external network with this flag enabled. If so, it uses an IP from this subnet for the router gateway IP address. If there is no subnet with this flag enabled, then all subnets will be considered as candidates for the router gateway IP address.

- Support for enabling/disabling faults in tenant subscriptions. This configuration allows users to subscribe to tenant notifications (with or without faults). A new Tripleo parameter, `AciEnableFaultSubscription`, controls this behavior. The default configuration is to disable the faults while subscribing to the tenant.

  ```
  AciEnableFaultSubscription:
      type: boolean
      default: false
      description: >
         To subscribe to faults during the subscription for tenants with ACI.
         Not subscribing to the faults improves the performance.
  ```

- Support for Enhanced Prometheus metrics for NAT traffic. When this feature is enabled in opflex-agent-ovs.conf.in, the opflex-agent sends flow mod requests to the OVS switch and listens for the packet count and byte count for the NAT flows. When the switch sends the flow stat reply, the opflex-agent updates the NAT Statistics counters and exports the metrics to Prometheus along with the EP attributes, such as endpoint uuid, endpoint mapped IP, source and destination endpoint point group for ingress and egress flow, floating IP address. The metrics depend on OVS flow mods to get the current packet and byte count for ingress and egress flows. If the endpoint gets deleted, the metrics are deleted and counters are reset to zero. These metrics are identified as follows:

  - " opflex_endpoint_to_extnetwork_bytes"
  - " opflex_endpoint_to_extnetwork_packets"
  - " opflex_extnetwork_to_endpoint_bytes"
  - " opflex_extnetwork_to_endpoint_packets"

Note: Endpoints are referred to as virtual instances. For the metrics to show up, enable the NAT Statistics feature in opflex-agent-ovs.conf file and initiate traffic between the virtual instance and the external network.

For more information regarding the integration, please refer to the Upstream Opflex documentation.

The feature can be configured using the tripleo parameters:

```
OpflexStatisticsNatEnabled:
    default: false
    type: boolean
OpflexStatisticsNatInterval:
    default: 10000
    type: number
```

- Support for adding source and destination tenant ID of a sent packet to the droplogs, if it is available. Whenever a packet is dropped, and if that packet had information on it's source and/or destination EPG available, the droplog will include the tenant ID. The first tenant is the source, with the second being the destination. Sample shown below:

```
Int-POL_TABLE MISS prj_92b4def72fa949cdbb60aa33bc6cb51c
prj_92b4def72fa949cdbb60aa33bc6cb51c  SMAC=00:22:bd:f8:19:ff
DMAC=fa:16:3e:f6:be:1a ETYP=IPv4 SRC=40.40.40.8 DST=30.30.30.163 LEN=84 DSCP=0
TTL=63 ID=51999 FLAGS=2 FRAG=0 PROTO=ICMP TYPE=8 CODE=0 ID=30209 SEQ=4418
```

To enable tenant ID logging, add the following to the droplog configuration file:

```
“drop-log-print-tenant”: true
```

If the packet does not have information on it's source/destination, then the equivalent tenant ID will be indicated as N/A, as shown below:

```
Int-SOURCE_TABLE MISS N/A N/A  SMAC=a4:53:0e:a5:e1:c7 DMAC=01:00:0c:cd:cd:ce
ETYP=Qtag QTAG=105 76_unrecognized
```

Cisco ACI Virtualization Compatibility Matrix

For information about Cisco ACI and OpenStack, see the Cisco Virtualization Compatibility Matrix at the following URL:

https://www.cisco.com/c/dam/en/us/td/docs/Website/datacenter/aci/virtualization/matrix/virtmatrix.html

Supported Scale

For the verified scalability limits (except for CLI limits), see the Verified Scalability Guide for this release. For Kubernetes-based Integrations (including Docker, OpenShift, and Rancher), and OpenStack Platform Scale Limits, see the following table.

Note: The scalability information in the following table applies to Kubernetes or OpenStack resources integrated with OpFlex into the Cisco ACI fabric. It does not apply to Microsoft SCVMM hosts or Cisco ACI Virtual Edge instances.

Limit Type	Maximum Supported
Number of OpFlex hosts per leaf	120
Number of OpFlex hosts per port	20
Number of vPC links per leaf	40
Number of endpoints per leaf	10,000
Number of endpoints per host	400
Number of virtual endpoints per leaf	40,000

Notes:

- For containers, an endpoint corresponds to a pod’s network interface.
- For OpenStack, an endpoint corresponds to any of the following:
  - A virtual machine (VM) interface (also known as vnic)
  - A DHCP agent’s port in OpenStack (if in DHCP namespace on the network controller)
  - A floating IP address
- Total virtual endpoints on a leaf can be calculated as virtual endpoints / leaf = VPCs x EPGs, where:
  - VPCs is the number of VPC links on the switch in the attachment profile used by the OpenStack Virtual Machine Manager (VMM).
  - EPGs is the number of EPGs provisioned for the OpenStack VMM.

For the CLI verified scalability limits, see the *Cisco NX-OS Style Command-Line Interface Configuration Guide* for this release.

Known Limitations

This section lists the known limitations.

- Cisco ACI Unified Plug-in for OpenStack does not support the following features:
  - ESX hypervisor support
  - ASR1K edgeNAT support
  - GBP/NFP Service chaining

    - ML2 Network constraints

- Dual-stack operation requires that all IPv4 and IPv6 subnets - both for internal and external networks - use the same VRF in Cisco ACI. The one exception to this is when separate external networks are used for IPv4 and IPv6 traffic. In that workflow, the IPv4 and IPv6 subnets used for internal networks plus the IPv6 subnets used for external networks all belong to one VRF, while the subnets for the IPv4 external network belong to a different VRF. IPv4 NAT can then be used for external networking.
- For installations with B-series that use VXLAN encapsulation, Layer 2 Policies (for example, bridge domains) should each contain only one Policy Target Group (that is, Endpoint Group) to ensure a functional data plane.
- The Cisco ACI OpenStack Plug-in is not integrated with the Multi-Site Orchestrator. When deploying to a Multi-Site deployment, the Cisco ACI configurations implemented by the plug-in must not be affected by the Multi-Site Orchestrator.
- When you delete the Overcloud Heat stack, the Overcloud nodes are freed but the virtual machine manager (VMM) domain remains present in Cisco APIC. The VMM appears in Cisco APIC as a stale VMM domain along with the tenant unless you delete the VMM domain manually. Before you delete the VMM domain, verify that the stack has been deleted from the undercloud, and check that any hypervisors appearing under the VMM domain are no longer in the connected state. After both these conditions are met, you can safely delete the VMM domain Cisco APIC.
- Due to a bug in upstream Neutron, subport bindings are not cleaned up in trunk workflows. This has existed in earlier releases and is equally applicable to usage with Open vSwitch (OVS) reference implementation agents as well as OpFlex agents. For more information about the Neutron bug, see bug 1639111 on the Launchpad.net website.

Usage Guidelines

- When performing updates of the overlcoud, it is recommended to update compute nodes in small groups of 3 or less. This minimizes the policy request load by the opflex-agent on the fabric, and helps ensure that the data plane remains stable during the upgrade window. This is done using the "–limit" flag when running the "openstack overcloud update run" command.
- The OpflexDroplogConfig parameter added in the 5.2(6) plugin release allows configuration of the opflex-agent droplog feature across all hosts when deployed using OpenStack Platform (OSP) Director 16. The parameter requires a valid json blob, which is used for each host's opflex-agent droplog configuration file.
- There is a known issue related to connection tracking in Red Hat Enterprise Linux, which prevents communication between the LBaaS worker VMs and the LBaaS healthcheck service. See the solution for *What is the option "nf_conntrack_tcp_be_liberal" for* on the RedHat website.

  To avoid this, the ExtraSysctlSettings tripleo parameter can be used to add this tuning to deployments.
- The APIC SNAT subnet only extension is used to control IP address allocation from a subnet on a neutron external network. When setting the gateway on a neutron router, if no subnet or IP address is specified, neutron picks the subnet with the lowest UUID value, and allocates an IP address from that subnet to use for the router gateway port. In order to avoid exhausting IP addresses intended for SNAT, this extension can be enabled on subnets used for SNAT:

```
openstack subnet set --apic-snat-subnet-only-enable foosubnet
```

Once enabled, whenever a neutron router is attached to the external network that owns the SNAT subnet, that subnet will not be used to allocate gateway IP addresses. This extension can also be disabled, allowing allocations from the subnet:

```
openstack subnet set --apic-snat-subnet-only-disable foosubnet
```

The default value for the extension is False, which means existing workflows will behave the same as before. If a user tries specifying a subnet or IP address on a subnet with this extension *enabled* when setting a router gateway, that operation will fail.

- Logging of dropped packets on hosts can be enabled in Hat OpenStack Platform (RHOSP) Director 16 This is one using the following tripleo parameter specified in the /opt/ciscoaci-tripleo-heat-templates deployment/ deployment/opflex/opflex-agent-container-puppet.yaml template:

```
  OpflexEnableDroplog:
    default: false
    description: Enable droplog feature on hypervisors
    type: Boolean
```

Setting this parameter to true enables logging of dropped packets on the hypervisor.

- A new template has been added for RHOSP 16, in order to support simultaneous operation of hypervisors using both optimized and non-optimized DHCP and metadata. This template is found on the undercloud in /opt/ciscoaci-tripleo-heat-templates/deployment/neutron_opflex/neutron-opflex-agent-container-puppet-controller.yaml, and should only be used to deploy the neutron-opflex-agent service on controller nodes.
- We recommended that service VMs used in service function chaining (SFC) workflows use static IP addressing and not rely on DHCP. When the service VM becomes part of a service chain in OpenStack and correspondingly a service graph on Cisco ACI, the associated EPG is removed. Thus, services such as DHCP are not available for the endpoint. This is applicable with OVS reference implementation agents as well as OpFlex agents.
- When you run the host report ansible-playbook `(/opt/ciscoaci-tripleo-heat-templates/tools/report.yml`), the step to copy files from a running container may return an error, causing the host report to fail. If this happens, rerun the playbook until it succeeds. The failure is due to a known issue in Red Hat OpenStack Platform (OSP) 13 Director. For more information, see the Red Hat Bugzilla bug 1767289. You can find the related product note in the Red Hat Customer portal knowledge base article " docker cp command sometimes failed with invalid argument."
- If you are using Cisco ACI Virtual Edge with OpenStack or Kubernetes OpFlex on the same leaf, do not use Cisco APIC version 4.2(3), or you will encounter the bug CSCvs49419. if you have that configuration and need features from the Cisco APIC 4.2(x) release train, use the 4.2(2) or 4.2(4) version.
- JuJu charms users must first update the Charms before installing the updated plug-in.
- Newer RHEL installations limit the maximum number of multicast group subscriptions to 20. This is configured with the net.ipv4.igmp_max_memberships sysctl variable. Installations using VXLAN encapsulation for OpenStack VMM domains should set this value higher than the total number of endpoint groups (EPGs) that might appear on the node (one for each Neutron network with Neutron workflow, or one for each Policy Target Group with Group Based Policy workflow).

Note: Controller hosts running DHCP agents that are connected to OpFlex networks have an EPG for each network.

- When using the allowed address pair feature with the Cisco ACI plug-in, be aware of the following differences from upstream implementation:
  - As OpenStack allows the same allowed_address_pair to be configured on multiple interfaces for HA, the OpFlex agent requires that the specific VNIC that currently owns a specific allowed_address_pair to assert that address ownership using Gratuitous ARP.
  - When using the promiscuous mode, the vSwitch stops enforcing the port security check. To get reverse traffic for a different IP or MAC address, you still need to use the allowed-address-pair feature. If you are running tempest, you will see test_port_security_macspoofing_port fail in scenario testing, as that test does not use the allowed-address-pair feature.
- Keystone configuration update

  When the OpenStack plug-in is installed in the unified mode, the Cisco installer adds the required configuration for keystone integration with AIM. When not using unified mode, or when using your own installer, the configuration section must be provisioned manually:

  ```
  [apic_aim_auth]
  auth_plugin=v3password
  auth_url=http://<IP Address of controller>:35357/v3
  username=admin
  password=<admin_password>
  user_domain_name=default
  project_domain_name=default
  project_name=admin
  ```
- When using optimized DHCP, the DHCP lease times are set by the configuration variable apic_optimized_dhcp_lease_time under the [ml2_apic_aim] section.
  - This requires a restart of neutron-server to take effect
  - If this value is updated, existing instances will continue using the old lease time, provided their neutron port is not changed (e.g. rebooting the instance would trigger a port change, and cause it to get the updated lease time). New instances will however use the updated lease time.
- In upstream Neutron, the " advertise_mtu" option has been removed.

  Since the aim_mapping driver still uses this configuration, the original configuration which appeared in the default section should be moved to the aim_mapping section. For example:

  ```
  [aim_mapping]
  advertise_mtu = True
  ```

  It is set to True by default in the code (if not explicitly specified in the config file).
- The Unified Plug-in allows coexistence of GBP and ML2 networking models on a single OpenStack Cloud installation. However, they must operate on different VRFs. We recommend using a single model per OpenStack Project.
- If a default VRF is implicitly created for a tenant in ML2, it is not implicitly deleted until the tenant is deleted (even if it not being used anymore).

- Unified model impact of the transaction Model Updates in Newton.

  When GBP and ML2 co-exist, GBP implicitly created some neutron resources. In Newton, the neutron transaction model has been updated and has added various checks. Some of those checks spuriously see this nested transaction usage as an error and log and raise an exception. The exception is handled correctly by GBP and there is no functional impact but unfortunately the neutron code also logs some exceptions in neutron log file – leading to the impression that the action had failed.

  While most such exceptions are logged at the DEBUG level, occasionally you might see some exceptions being logged at the ERROR level. If such an exception log is followed by a log message which indicates that the operation is being retried, the exception is being handled correctly. One such example is the following:

  Delete of policy-target on a policy-target-group associated to a network-service-policy could raise this exception:

  ```
  2017-03-18 12:52:34.421 27767 ERROR neutron.api.v2.resource […] delete failed
  2017-03-18 12:52:34.421 27767 ERROR neutron.api.v2.resource Traceback …:
  2017-03-18 12:52:34.421 27767 ERROR neutron.api.v2.resource   File
  "/usr/lib/python2.7/site-packages/neutron/api/v2/resource.py", line 84, …
  ...
  2017-03-18 12:52:34.421 27767 ERROR neutron.api.v2.resource     raise …
  2017-03-18 12:52:34.421 27767 ERROR neutron.api.v2.resource ResourceClosedError: This
  transaction is closed
  ```

  Note: Cisco is working with the upstream community for further support on Error level logs.

- When a Layer 2 policy is deleted in GBP, some implicit artifacts related to it may not be deleted (resulting in unused BDs/subnets on Cisco APIC). If you hit that situation, the workaround is to create a new empty Layer 2 policy in the same context and delete it.

- If you use tempest to validate OpenStack, the following tests are expected to fail and can be ignored:

  ```
  tempest.scenario.test_network_basic_ops.TestNetworkBasicOps.test_update_router_admin_s
  tate
  ```

- Neutron-server logs may show the following message when DEBUG level is enabled:

  ```
  Timed out waiting for RPC response: Timeout while waiting on RPC response - topic:
  "<unknown>", RPC method: "<unknown>" info: "<unknown>"
  ```

  This message can be ignored.

- High Availability LBaaSv2 is not supported.

- OpenStack Newton is the last version to support non-unified plug-in. OpenStack Ocata and future releases will only be supported with the unified plug-in.

- For deployments running Cisco ACI version 4.1(2g) and using the Group Based Policy workflow and associated APIs, contract filters set to an EtherType of ARP can result in the filter being incorrectly set as "Unspecified" on the leaf. If an EtherType of ARP is required, then you must use a Cisco ACI release other than 4.1(2g).

- Some deployments require installation of an “allow” entry in IP Tables for IGMP. This must be added to all hosts running an OpFlex agent and using VXLAN encapsulation to the leaf. The rule must be added using the following command:

  ```
  # iptables -A INPUT -p igmp -j ACCEPT
  ```

  In order to make this change persistent across reboots, add the command either to `/etc/rc.d/rc.local` or to a cron job that runs after reboot.

- For deployments that use B-series servers, an additional service must be started on the hosts to ensure that connectivity is maintained with the leaf at all times. Complete the following steps:

Step 1. Install the Cisco APIC API package (python-apicapi for Debian packaging, apicapi for RPM packaging) for any servers running an OpFlex agent.

Step 2. Add the OpFlex uplink bond name to /etc/environments (that is, opflex_bondif=bond1).

This is needed if the interface is other than default (bond0).

Step 3. Enable the apic-bond-watch service using the following command:

```
sudo systemctl enable apic-bond-watch
```

Step 4. Start the apic-bond-watch service using the following command:

```
sudo systemctl start apic-bond-watch
```

For OpenStack Director installations using VXLAN encapsulation for VMM domains, two additional configuration items may be needed to handle large installations. The number of multicast groups should be configured to match the maximum number of endpoint groups for the host, and the maximum auxiliary memory for sockets needs to be increased for IPC. These are configured using the extra-config.yaml file, with the following parameters:

```
ControllerParameters:
  ExtraSysctlSettings:
    net.ipv4.igmp_max_memberships:
      value: 4096
    net.core.optmem_max:
      value: 1310720
ComputeParameters:
  ExtraSysctlSettings:
    net.ipv4.igmp_max_memberships:
          value: 1024
```

The IGMP max memberships value should be greater than or equal to the number of Neutron networks that the host has Neutron ports on. For example, if a compute host has 100 instances, and each instance is on a different Neutron network, then this number must be set to at least 100. Controller hosts running the neutron-dhcp-agent will need set this value to match the number of Neutron networks managed by that agent, which means this number will probably need to be higher on controller hosts than compute hosts.

- For installations not using OpenStack Director, the maximum allowed packet size for the database must be configured to support database transactions for tenants in AIM with large configurations. The default value installed by OpenStack director in `/etc/my.cnf.d/galera.cnf` is sufficient for most installations:

  ```
  [mysqld]
  …
  ```

```
max_allowed_packet = 16M
[mysqldump]
max_allowed_packet = 16M”
```

- After deploying Queens with Juju charms (18 or 19), sometimes a VM spawn fails. The failure is due to a neutron-opflex-agent failing to start on the host that the VM was scheduled to. The host can be determined using the neutron `agent-list` command: The neutron-opflex-agent is missing for the effected compute node.

  Restart of neutron-opflex-agent on the affected node fixes the problem and can be used as a workaround after a fresh deployment.

- When you do an upgrade involving Red Hat OSP13, the installer doesn’t delete the `/var/www/html/acirpo` directory. This causes problems when building new containers. When performing an upgrade using OSP13, be sure to manually delete this directory before installing the new RPM.

Open Issues

There are no known issues in this release.

Resolved Issues

Click the bug ID to access the Bug Search tool and see additional information about the bug.

Bug ID	Description
CSCwi58534	ACI OpenStack | Floating IP (FIP) is pointing to two private IPs.
CSCvt54179	Stale tenants entries after running tempest.
CSCwk04406	Simultaneous restart of opflex-agents results in 20 minute outage.
CSCwj30315	Exception when neutron port has no port binding.
CSCwk79521	When provisioning LBaaS service in Openstack with ACI ML2 plugin health check service doesnt work.

Known Issues

There are no known issues in this release.

Related Content

See the Cisco Application Policy Infrastructure Controller (APIC) page for the documentation.

The documentation includes installation, upgrade, configuration, programming, and troubleshooting guides, technical references, release notes, and knowledge base (KB) articles, as well as other documentation. KB articles provide information about a specific use case or a specific topic.

By using the " Choose a topic" and " Choose a document type" fields of the APIC documentation website, you can narrow down the displayed documentation list to make it easier to find the desired document.

You can watch videos that demonstrate how to perform specific tasks in the Cisco APIC on the Cisco Data Center Networking YouTube channel.

Documentation Feedback

To provide technical feedback on this document, or to report an error or omission, send your comments to apic-docfeedback@cisco.com. We appreciate your feedback.

Legal Information

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.