OpenShift Container Platform 4.2

Networking

Configuring and managing cluster networking

Last Updated: 2019-10-31

OpenShift Container Platform 4.2 Networking

Configuring and managing cluster networking

Legal Notice

Abstract

This document provides instructions for configuring and managing your OpenShift Container Platform cluster network, including DNS, ingress, and the Pod network.

Table of Contents

CHAPTER 1. UNDERSTANDING NETWORKING

Kubernetes ensures that Pods are able to network with each other, and allocates each Pod an IP address from an internal network. This ensures all containers within the Pod behave as if they were on the same host. Giving each Pod its own IP address means that Pods can be treated like physical hosts or virtual machines in terms of port allocation, networking, naming, service discovery, load balancing, application configuration, and migration.

1.1. OPENSHIFT CONTAINER PLATFORM DNS

If you are running multiple services, such as front-end and back-end services for use with multiple Pods, environment variables are created for user names, service IPs, and more so the front-end Pods can communicate with the back-end services. If the service is deleted and recreated, a new IP address can be assigned to the service, and requires the front-end Pods to be recreated to pick up the updated values for the service IP environment variable. Additionally, the back-end service must be created before any of the front-end Pods to ensure that the service IP is generated properly, and that it can be provided to the front-end Pods as an environment variable.

For this reason, OpenShift Container Platform has a built-in DNS so that the services can be reached by the service DNS as well as the service IP/port.

CHAPTER 2. CLUSTER NETWORK OPERATOR IN OPENSHIFT CONTAINER PLATFORM

The Cluster Network Operator (CNO) deploys and manages the cluster network components on an OpenShift Container Platform cluster, including the Container Network Interface (CNI) Software Defined Networking (SDN) plug-in selected for the cluster during installation.

2.1. CLUSTER NETWORK OPERATOR

The Cluster Network Operator implements the **network** API from the **operator.openshift.io** API group. The Operator deploys the OpenShift SDN plug-in, or a different SDN plug-in if selected during cluster installation, using a DaemonSet.

Procedure

The Cluster Network Operator is deployed during installation as a Kubernetes **Deployment**.

1. Run the following command to view the Deployment status:

```
$ oc get -n openshift-network-operator deployment/network-operator

NAME               READY   UP-TO-DATE   AVAILABLE   AGE
network-operator   1/1     1            1           56m
```

2. Run the following command to view the state of the Cluster Network Operator:

```
$ oc get clusteroperator/network

NAME      VERSION   AVAILABLE   PROGRESSING   DEGRADED   SINCE
network   4.2.0     True        False         False      50m
```

The following fields provide information about the status of the operator: **AVAILABLE**, **PROGRESSING**, and **DEGRADED**. The **AVAILABLE** field is **True** when the Cluster Network Operator reports an available status condition.

2.2. VIEWING THE CLUSTER NETWORK CONFIGURATION

Every new OpenShift Container Platform installation has a **network.config** object named **cluster**.

Procedure

- Use the **oc describe** command to view the cluster network configuration:

```
$ oc describe network.config/cluster

Name:         cluster
Namespace:
Labels:       <none>
Annotations:  <none>
API Version:  config.openshift.io/v1
Kind:         Network
Metadata:
  Self Link:           /apis/config.openshift.io/v1/networks/cluster
```

```
Spec: 1
  Cluster Network:
    Cidr:         10.128.0.0/14
    Host Prefix:  23
  Network Type:   OpenShiftSDN
  Service Network:
    172.30.0.0/16
Status: 2
  Cluster Network:
    Cidr:               10.128.0.0/14
    Host Prefix:        23
  Cluster Network MTU:  8951
  Network Type:         OpenShiftSDN
  Service Network:
    172.30.0.0/16
Events:  <none>
```

1. The **Spec** field displays the configured state of the cluster network.
2. The **Status** field displays the current state of the cluster network configuration.

2.3. VIEWING CLUSTER NETWORK OPERATOR STATUS

You can inspect the status and view the details of the Cluster Network Operator using the **oc describe** command.

Procedure

- Run the following command to view the status of the Cluster Network Operator:

```
$ oc describe clusteroperators/network
```

2.4. VIEWING CLUSTER NETWORK OPERATOR LOGS

You can view Cluster Network Operator logs by using the **oc logs** command.

Procedure

- Run the following command to view the logs of the Cluster Network Operator:

```
$ oc logs --namespace=openshift-network-operator deployment/network-operator
```

2.5. CLUSTER NETWORK OPERATOR CUSTOM RESOURCE (CR)

The cluster network configuration in the **Network.operator.openshift.io** custom resource (CR) stores the configuration settings for the Cluster Network Operator (CNO). The Operator manages the cluster network.

You can specify the cluster network configuration for your OpenShift Container Platform cluster by setting the parameters for the **defaultNetwork** parameter in the CNO CR. The following CR displays the default configuration for the CNO and explains both the parameters you can configure and valid parameter values:

Cluster Network Operator CR

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
  clusterNetwork: 1
  - cidr: 10.128.0.0/14
    hostPrefix: 23
  serviceNetwork: 2
  - 172.30.0.0/16
  defaultNetwork: 3
    ...
  kubeProxyConfig: 4
    iptablesSyncPeriod: 30s 5
    proxyArguments:
      iptables-min-sync-period: 6
      - 30s
```

1. A list specifying the blocks of IP addresses from which Pod IPs are allocated and the subnet prefix length assigned to each individual node.
2. A block of IP addresses for services. The OpenShift SDN Container Network Interface (CNI) plug-in supports only a single IP address block for the service network.
3. Configures the software-defined networking (SDN) for the cluster network.
4. The parameters for this object specify the Kubernetes network proxy (kube-proxy) configuration.
5. The refresh period for **iptables** rules. The default value is **30s**. Valid suffixes include **s**, **m**, and **h** and are described in the Go time package documentation.
6. The minimum duration before refreshing **iptables** rules. This parameter ensures that the refresh does not happen too frequently. Valid suffixes include **s**, **m**, and **h** and are described in the Go time package

2.5.1. Configuration parameters for OpenShift SDN

The following YAML object describes the configuration parameters for OpenShift SDN:

```
defaultNetwork:
  type: OpenShiftSDN 1
  openshiftSDNConfig: 2
    mode: NetworkPolicy 3
    mtu: 1450 4
    vxlanPort: 4789 5
```

1. The Software Defined Networking (SDN) plug-in being used. OpenShift SDN is the only plug-in supported in OpenShift Container Platform 4.2.
2. OpenShift SDN specific configuration parameters.

3. The network isolation mode for the OpenShift SDN CNI plug-in.
4. MTU for the VXLAN overlay network. This value is normally configured automatically.
5. The port to use for all VXLAN packets. The default value is **4789**.

2.5.2. Cluster Network Operator example CR

A complete CR for the CNO is displayed in the following example:

Cluster Network Operator example CR

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
  clusterNetwork:
  - cidr: 10.128.0.0/14
    hostPrefix: 23
  serviceNetwork:
  - 172.30.0.0/16
  defaultNetwork:
    type: OpenShiftSDN
    openshiftSDNConfig:
      mode: NetworkPolicy
      mtu: 1450
      vxlanPort: 4789
  kubeProxyConfig:
    iptablesSyncPeriod: 30s
    proxyArguments:
      iptables-min-sync-period:
      - 30s
```

CHAPTER 3. DNS OPERATOR IN OPENSHIFT CONTAINER PLATFORM

The DNS Operator deploys and manages CoreDNS to provide a name resolution service to pods, enabling DNS-based Kubernetes Service discovery in OpenShift.

3.1. DNS OPERATOR

The DNS Operator implements the **dns** API from the **operator.openshift.io** API group. The operator deploys CoreDNS using a DaemonSet, creates a Service for the DaemonSet, and configures the kubelet to instruct pods to use the CoreDNS Service IP for name resolution.

Procedure

The DNS Operator is deployed during installation as a Kubernetes **Deployment**.

1. Use the **oc get** command to view the Deployment status:

```
$ oc get -n openshift-dns-operator deployment/dns-operator
NAME           READY     UP-TO-DATE   AVAILABLE   AGE
dns-operator   1/1       1            1           23h
```

   ClusterOperator is the Custom Resource object which holds the current state of an operator. This object is used by operators to convey their state to the rest of the cluster.

2. Use the **oc get** command to view the state of the DNS Operator:

```
$ oc get clusteroperator/dns
NAME      VERSION     AVAILABLE   PROGRESSING   DEGRADED   SINCE
dns       4.1.0-0.11  True        False         False      92m
```

   AVAILABLE, **PROGRESSING** and **DEGRADED** provide information about the status of the operator. **AVAILABLE** is **True** when at least 1 pod from the CoreDNS DaemonSet is reporting an **Available** status condition.

3.2. VIEW THE DEFAULT DNS

Every new OpenShift Container Platform installation has a **dns.operator** named **default**. It cannot be customized, replaced, or supplemented with additional **dnses**.

Procedure

1. Use the **oc describe** command to view the default **dns**:

```
$ oc describe dns.operator/default
Name:         default
Namespace:
Labels:       <none>
Annotations:  <none>
API Version:  operator.openshift.io/v1
Kind:         DNS
...
Status:
```

```
Cluster Domain:  cluster.local 1
Cluster IP:      172.30.0.10 2
...
```

1 The Cluster Domain field is the base DNS domain used to construct fully qualified Pod and Service domain names.

2 The Cluster IP is the address pods query for name resolution. The IP is defined as the 10th address in the Service CIDR range.

2. To find the Service CIDR of your cluster, use the **oc get** command:

```
$ oc get networks.config/cluster -o jsonpath='{$.status.serviceNetwork}'
[172.30.0.0/16]
```

NOTE

Configuration of the CoreDNS Corefile or Kubernetes plugin is not supported.

3.3. DNS OPERATOR STATUS

You can inspect the status and view the details of the DNS Operator using the **oc describe** command.

Procedure

View the status of the DNS Operator:

```
$ oc describe clusteroperators/dns
```

3.4. DNS OPERATOR LOGS

You can view DNS Operator logs by using the **oc logs** command.

Procedure

View the logs of the DNS Operator:

```
$ oc logs --namespace=openshift-dns-operator deployment/dns-operator
```

CHAPTER 4. INGRESS OPERATOR IN OPENSHIFT CONTAINER PLATFORM

The Ingress Operator implements the **ingresscontroller** API and is the component responsible for enabling external access to OpenShift Container Platform cluster services. The operator makes this possible by deploying and managing one or more HAProxy-based Ingress Controllers to handle routing. You can use the Ingress Operator to route traffic by specifying OpenShift Container Platform **Route** and Kubernetes **Ingress** resources.

4.1. THE INGRESS CONFIGURATION ASSET

The installation program generates an asset with an **Ingress** resource in the **config.openshift.io** API group, **cluster-ingress-02-config.yml**.

YAML Definition of the Ingress resource

```
apiVersion: config.openshift.io/v1
kind: Ingress
metadata:
  name: cluster
spec:
  domain: apps.openshiftdemos.com
```

The installation program stores this asset in the **cluster-ingress-02-config.yml** file in the **manifests/** directory. This **Ingress** resource defines the cluster-wide configuration for Ingress. This Ingress configuration is used as follows:

- The Ingress Operator uses the domain from the cluster Ingress configuration as the domain for the default Ingress Controller.
- The OpenShift API server operator uses the domain from the cluster Ingress configuration as the domain used when generating a default host for a **Route** resource that does not specify an explicit host.

4.2. VIEW THE DEFAULT INGRESS CONTROLLER

The Ingress Operator is a core feature of OpenShift Container Platform and is enabled out of the box.

Every new OpenShift Container Platform installation has an **ingresscontroller** named default. It can be supplemented with additional Ingress Controllers. If the default **ingresscontroller** is deleted, the Ingress Operator will automatically recreate it within a minute.

Procedure

- View the default Ingress Controller:

```
$ oc describe --namespace=openshift-ingress-operator ingresscontroller/default
```

4.3. VIEW INGRESS OPERATOR STATUS

You can view and inspect the status of your Ingress Operator.

Procedure

Procedure

- View your Ingress Operator status:

```
$ oc describe clusteroperators/ingress
```

4.4. VIEW INGRESS CONTROLLER LOGS

You can view your Ingress Controller logs.

Procedure

- View your Ingress Controller logs:

```
$ oc logs --namespace=openshift-ingress-operator deployments/ingress-operator
```

4.5. VIEW INGRESS CONTROLLER STATUS

Your can view the status of a particular Ingress Controller.

Procedure

- View the status of an Ingress Controller:

```
$ oc describe --namespace=openshift-ingress-operator ingresscontroller/<name>
```

4.6. SETTING A CUSTOM DEFAULT CERTIFICATE

As an administrator, you can configure an Ingress Controller to use a custom certificate by creating a **Secret** resource and editing the **IngressController** custom resource (CR).

Prerequisites

- You must have a certificate/key pair in PEM-encoded files, where the certificate is signed by a trusted certificate authority and valid for the Ingress domain.
- You must have an **IngressController** CR. You may use the default one:

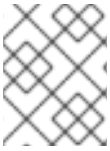
```
$ oc --namespace openshift-ingress-operator get ingresscontrollers
NAME      AGE
default   10m
```

WARNING

If the default certificate is replaced, it **must** be signed by a public certificate authority already included in the CA bundle as provided by the container userspace.

Procedure

The following assumes that the custom certificate and key pair are in the **tls.crt** and **tls.key** files in the current working directory. Substitute the actual path names for **tls.crt** and **tls.key**. You also may substitute another name for **custom-certs-default** when creating the **Secret** resource and referencing it in the **IngressController** CR.



NOTE

This action will cause the Ingress Controller to be redeployed, using a rolling deployment strategy.

1. Create a **Secret** resource containing the custom certificate in the **openshift-ingress** namespace using the **tls.crt** and **tls.key** files.

```
$ oc --namespace openshift-ingress create secret tls custom-certs-default --cert=tls.crt --
key=tls.key
```

2. Update the **IngressController** CR to reference the new certificate secret:

```
$ oc patch --type=merge --namespace openshift-ingress-operator ingresscontrollers/default \
  --patch '{"spec":{"defaultCertificate":{"name":"custom-certs-default"}}}'
```

3. Verify the update was effective:

```
$ oc get --namespace openshift-ingress-operator ingresscontrollers/default \
  --output jsonpath='{.spec.defaultCertificate}'
```

The output should look like:

```
map[name:custom-certs-default]
```

The certificate secret name should match the value used to update the CR.

Once the **IngressController** CR has been modified, the Ingress Operator will update the Ingress Controller's deployment to use the custom certificate.

4.7. SCALING AN INGRESS CONTROLLER

Manually scale an Ingress Controller to meeting routing performance or availability requirements such as the requirement to increase throughput. **oc** commands are used to scale the **IngressController** resource. The following procedure provides an example for scaling up the default **IngressController**.

Procedure

1. View the current number of available replicas for the default **IngressController**:

```
$ oc get -n openshift-ingress-operator ingresscontrollers/default -o
jsonpath='{$.status.availableReplicas}'
2
```

2. Scale the default **IngressController** to the desired number of replicas using the **oc patch** command. The following example scales the default **IngressController** to 3 replicas:

```
$ oc patch -n openshift-ingress-operator ingresscontroller/default --patch '{"spec":{"replicas":
3}}' --type=merge
ingresscontroller.operator.openshift.io/default patched
```

3. Verify that the default **IngressController** scaled to the number of replicas that you specified:

```
$ oc get -n openshift-ingress-operator ingresscontrollers/default -o
jsonpath='{$.status.availableReplicas}'
3
```

NOTE

Scaling is not an immediate action, as it takes time to create the desired number of replicas.

4.8. CONFIGURING INGRESS CONTROLLER SHARDING BY USING ROUTE LABELS

Ingress Controller sharding by using route labels means that the Ingress Controller serves any route in any namespace that is selected by the route selector.

Ingress Controller sharding is useful when balancing incoming traffic load among a set of Ingress Controllers and when isolating traffic to a specific Ingress Controller. For example, company A goes to one Ingress Controller and company B to another.

Procedure

1. Edit the **router-internal.yaml** file:

```
# cat router-internal.yaml
apiVersion: v1
items:
- apiVersion: operator.openshift.io/v1
  kind: IngressController
  metadata:
    name: sharded
    namespace: openshift-ingress-operator
  spec:
    domain: <apps-sharded.basedomain.example.net>
    nodePlacement:
      nodeSelector:
        matchLabels:
          node-role.kubernetes.io/worker: ""
    routeSelector:
      matchLabels:
        type: sharded
  status: {}
kind: List
metadata:
  resourceVersion: ""
  selfLink: ""
```

2. Apply the Ingress Controller **router-internal.yaml** file:

```
# oc apply -f router-internal.yaml
```

The Ingress Controller selects routes in any namespace that have the label **type: sharded**.

4.9. CONFIGURING INGRESS CONTROLLER SHARDING BY USING NAMESPACE LABELS

Ingress Controller sharding by using namespace labels means that the Ingress Controller serves any route in any namespace that is selected by the namespace selector.

Ingress Controller sharding is useful when balancing incoming traffic load among a set of Ingress Controllers and when isolating traffic to a specific Ingress Controller. For example, company A goes to one Ingress Controller and company B to another.

Procedure

1. Edit the **router-internal.yaml** file:

```
# cat router-internal.yaml
apiVersion: v1
items:
- apiVersion: operator.openshift.io/v1
  kind: IngressController
  metadata:
    name: sharded
    namespace: openshift-ingress-operator
  spec:
    domain: <apps-sharded.basedomain.example.net>
    nodePlacement:
      nodeSelector:
        matchLabels:
          node-role.kubernetes.io/worker: ""
    routeSelector:
      matchLabels:
        type: sharded
  status: {}
kind: List
metadata:
  resourceVersion: ""
  selfLink: ""
```

2. Apply the Ingress Controller **router-internal.yaml** file:

```
# oc apply -f router-internal.yaml
```

The Ingress Controller selects routes in any namespace that is selected by the namespace selector that have the label **type: sharded**.

CHAPTER 5. CONFIGURING NETWORK POLICY WITH OPENSHIFT SDN

5.1. ABOUT NETWORK POLICY

In a cluster using a Kubernetes Container Network Interface (CNI) plug-in that supports NetworkPolicy, network isolation is controlled entirely by NetworkPolicy objects. In OpenShift Container Platform 4.2, OpenShift SDN supports using NetworkPolicy in its default network isolation mode.

NOTE

The Kubernetes **v1** NetworkPolicy features are available in OpenShift Container Platform except for egress policy types and IPBlock.

By default, all Pods in a project are accessible from other Pods and network endpoints. To isolate one or more Pods in a project, you can create NetworkPolicy objects in that project to indicate the allowed incoming connections. Project administrators can create and delete NetworkPolicy objects within their own project.

If a Pod is matched by selectors in one or more NetworkPolicy objects, then the Pod will accept only connections that are allowed by at least one of those NetworkPolicy objects. A Pod that is not selected by any NetworkPolicy objects is fully accessible.

The following example NetworkPolicy objects demonstrate supporting different scenarios:

- Deny all traffic:
  To make a project deny by default, add a NetworkPolicy object that matches all Pods but accepts no traffic:

```
kind: NetworkPolicy
apiVersion: networking.k8s.io/v1
metadata:
  name: deny-by-default
spec:
  podSelector:
  ingress: []
```

- Only allow connections from the OpenShift Container Platform Ingress Controller:
  To make a project allow only connections from the OpenShift Container Platform Ingress Controller, add the following NetworkPolicy object:

```
apiVersion: networking.k8s.io/v1
kind: NetworkPolicy
metadata:
  name: allow-from-openshift-ingress
spec:
  ingress:
  - from:
    - namespaceSelector:
        matchLabels:
          network.openshift.io/policy-group: ingress
```

```
  podSelector: {}
  policyTypes:
  - Ingress
```

- Only accept connections from Pods within a project:
  To make Pods accept connections from other Pods in the same project, but reject all other connections from Pods in other projects, add the following NetworkPolicy object:

```
kind: NetworkPolicy
apiVersion: networking.k8s.io/v1
metadata:
  name: allow-same-namespace
spec:
  podSelector:
  ingress:
  - from:
    - podSelector: {}
```

- Only allow HTTP and HTTPS traffic based on Pod labels:
  To enable only HTTP and HTTPS access to the Pods with a specific label (**role=frontend** in following example), add a NetworkPolicy object similar to:

```
kind: NetworkPolicy
apiVersion: networking.k8s.io/v1
metadata:
  name: allow-http-and-https
spec:
  podSelector:
    matchLabels:
      role: frontend
  ingress:
  - ports:
    - protocol: TCP
      port: 80
    - protocol: TCP
      port: 443
```

NetworkPolicy objects are additive, which means you can combine multiple NetworkPolicy objects together to satisfy complex network requirements.

For example, for the NetworkPolicy objects defined in previous samples, you can define both **allow-same-namespace** and **allow-http-and-https** policies within the same project. Thus allowing the Pods with the label **role=frontend**, to accept any connection allowed by each policy. That is, connections on any port from Pods in the same namespace, and connections on ports **80** and **443** from Pods in any namespace.

5.2. EXAMPLE NETWORKPOLICY OBJECT

The following annotates an example NetworkPolicy object:

```
kind: NetworkPolicy
apiVersion: extensions/v1beta1
metadata:
  name: allow-27107 (1)
```

```
spec:
  podSelector: 2
    matchLabels:
      app: mongodb
  ingress:
  - from:
    - podSelector: 3
        matchLabels:
          app: app
    ports: 4
    - protocol: TCP
      port: 27017
```

1. The **name** of the NetworkPolicy object.
2. A selector describing the Pods the policy applies to. The policy object can only select Pods in the project that the NetworkPolicy object is defined.
3. A selector matching the Pods that the policy object allows ingress traffic from. The selector will match Pods in any project.
4. A list of one or more destination ports to accept traffic on.

5.3. CREATING A NETWORKPOLICY OBJECT

To define granular rules describing Ingress network traffic allowed for projects in your cluster, you can create NetworkPolicy objects.

Prerequisites

- A cluster using the OpenShift SDN network plug-in with **mode: NetworkPolicy** set. This mode is the default for OpenShift SDN.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster.

Procedure

1. Create a policy rule:
   a. Create a **<policy-name>.yaml** file where **<policy-name>** describes the policy rule.
   b. In the file you just created define a policy object, such as in the following example:

```
kind: NetworkPolicy
apiVersion: networking.k8s.io/v1
metadata:
  name: <policy-name> 1
spec:
  podSelector:
  ingress: []
```

1. Specify a name for the policy object.

2. Run the following command to create the policy object:

```
$ oc create -f <policy-name>.yaml -n <project>
```

In the following example, a new NetworkPolicy object is created in a project named **project1**:

```
$ oc create -f default-deny.yaml -n project1
networkpolicy "default-deny" created
```

5.4. DELETING A NETWORKPOLICY OBJECT

You can delete a NetworkPolicy object.

Prerequisites

- A cluster using the OpenShift SDN network plug-in with **mode: NetworkPolicy** set. This mode is the default for OpenShift SDN.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster.

Procedure

- To delete a NetworkPolicy object, run the following command:

```
$ oc delete networkpolicy -l name=<policy-name> 1
```

1 Specify the name of the NetworkPolicy object to delete.

5.5. VIEWING NETWORKPOLICY OBJECTS

You can list the NetworkPolicy objects in your cluster.

Prerequisites

- A cluster using the OpenShift SDN network plug-in with **mode: NetworkPolicy** set. This mode is the default for OpenShift SDN.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster.

Procedure

- To view NetworkPolicy objects defined in your cluster, run the following command:

```
$ oc get networkpolicy
```

5.6. CONFIGURING MULTITENANT ISOLATION USING NETWORKPOLICY

You can configure your project to isolate it from Pods and Services in other projects.

Prerequisites

- A cluster using the OpenShift SDN network plug-in with **mode: NetworkPolicy** set. This mode is the default for OpenShift SDN.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster.

Procedure

1. Create the following files containing NetworkPolicy object definitions:
   a. A file named **allow-from-openshift-ingress.yaml** containing the following:

```
apiVersion: networking.k8s.io/v1
kind: NetworkPolicy
metadata:
  name: allow-from-openshift-ingress
spec:
  ingress:
  - from:
    - namespaceSelector:
        matchLabels:
          network.openshift.io/policy-group: ingress
  podSelector: {}
  policyTypes:
  - Ingress
```

   b. A file named **allow-from-openshift-monitoring.yaml** containing the following:

```
apiVersion: networking.k8s.io/v1
kind: NetworkPolicy
metadata:
  name: allow-from-openshift-monitoring
spec:
  ingress:
  - from:
    - namespaceSelector:
        matchLabels:
          network.openshift.io/policy-group: monitoring
  podSelector: {}
  policyTypes:
  - Ingress
```

2. For each policy file, run the following command to create the NetworkPolicy object:

```
$ oc apply -f <policy-name>.yaml \ 1
  -n <project> 2
```

1. Replace **<policy-name>** with the filename of the file containing the policy.
2. Replace **<project>** with the name of the project to apply the NetworkPolicy object to.

3. Optional: Confirm that the NetworkPolicy object exists in your current project by running the following command:

```
$ oc get networkpolicy <policy-name> -o yaml
```

In the following example, the **allow-from-openshift-ingress** NetworkPolicy object is displayed:

```
$ oc get networkpolicy allow-from-openshift-ingress -o yaml

apiVersion: extensions/v1beta1
kind: NetworkPolicy
metadata:
  name: allow-from-openshift-ingress
  namespace: project1
spec:
  ingress:
  - from:
    - namespaceSelector:
        matchLabels:
          network.openshift.io/policy-group: ingress
  podSelector: {}
  policyTypes:
  - Ingress
```

CHAPTER 6. CREATING DEFAULT NETWORK POLICIES FOR A NEW PROJECT

As a cluster administrator, you can modify the new project template to automatically include NetworkPolicy objects when you create a new project.

6.1. MODIFYING THE TEMPLATE FOR NEW PROJECTS

As a cluster administrator, you can modify the default project template so that new projects are created using your custom requirements.

To create your own custom project template:

Procedure

1. Log in as a user with **cluster-admin** privileges.
2. Generate the default project template:

```
$ oc adm create-bootstrap-project-template -o yaml > template.yaml
```

3. Use a text editor to modify the generated **template.yaml** file by adding objects or modifying existing objects.
4. The project template must be created in the **openshift-config** namespace. Load your modified template:

```
$ oc create -f template.yaml -n openshift-config
```

5. Edit the project configuration resource using the web console or CLI.
   - Using the web console:
     i. Navigate to the **Administration → Cluster Settings** page.
     ii. Click **Global Configuration** to view all configuration resources.
     iii. Find the entry for **Project** and click **Edit YAML**.
   - Using the CLI:
     i. Edit the **project.config.openshift.io/cluster** resource:

```
$ oc edit project.config.openshift.io/cluster
```

6. Update the **spec** section to include the **projectRequestTemplate** and **name** parameters, and set the name of your uploaded project template. The default name is **project-request**.

Project configuration resource with custom project template

```
apiVersion: config.openshift.io/v1
kind: Project
metadata:
  ...
```

```
spec:
  projectRequestTemplate:
    name: <template_name>
```

7. After you save your changes, create a new project to verify that your changes were successfully applied.

6.2. ADDING NETWORK POLICY OBJECTS TO THE NEW PROJECT TEMPLATE

As a cluster administrator, you can add network policy objects to the default template for new projects. OpenShift Container Platform will automatically create all the NetworkPolicy CRs specified in the template in the project.

Prerequisites

- A cluster using the OpenShift SDN network plug-in with **mode: NetworkPolicy** set. This mode is the default for OpenShift SDN.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster with a user with **cluster-admin** privileges.
- You must have created a custom default project template for new projects.

Procedure

1. Edit the default template for a new project by running the following command:

```
$ oc edit template <project_template> -n openshift-config
```

   Replace **<project_template>** with the name of the default template that you configured for your cluster. The default template name is **project-request**.

2. In the template, add each NetworkPolicy object as an element to the **objects** parameter. The **objects** parameter accepts a collection of one or more objects.
   In the following example, the **objects** parameter collection includes several NetworkPolicy objects:

```
objects:
- apiVersion: networking.k8s.io/v1
  kind: NetworkPolicy
  metadata:
    name: allow-from-same-namespace
  spec:
    podSelector:
    ingress:
    - from:
      - podSelector: {}
- apiVersion: networking.k8s.io/v1
  kind: NetworkPolicy
  metadata:
    name: allow-from-openshift-ingress
  spec:
```

```
    ingress:
    - from:
      - namespaceSelector:
          matchLabels:
            network.openshift.io/policy-group: ingress
    podSelector: {}
    policyTypes:
    - Ingress
...
```

3. Optional: Create a new project to confirm that your network policy objects are created successfully by running the following commands:

   a. Create a new project:

   ```
   oc new-project <project>
   ```

   Replace **<project>** with the name for the project you are creating.

   b. Confirm that the network policy objects in the new project template exist in the new project:

   ```
   oc get networkpolicy
   NAME                           POD-SELECTOR   AGE
   allow-from-openshift-ingress   <none>         7s
   allow-from-same-namespace      <none>         7s
   ```

CHAPTER 7. MULTIPLE NETWORKS

7.1. UNDERSTANDING MULTIPLE NETWORKS

In Kubernetes, container networking is delegated to networking plug-ins that implement the Container Network Interface (CNI).

OpenShift Container Platform uses the Multus CNI plug-in to allow chaining of CNI plug-ins. During cluster installation, you configure your *default* Pod network. The default network handles all routine network traffic for the cluster. You can define an *additional network* based on the available CNI plug-ins and attach one or more of these networks to your Pods. You can define more than one additional network for your cluster, depending on your needs. This gives you flexibility when you configure Pods that deliver network functionality, such as switching or routing.

7.1.1. Usage scenarios for an additional network

You can use an additional network in situations where network isolation is needed, including data plane and control plane separation. Isolating network traffic is useful for the following performance and security reasons:

Performance

You can send traffic on two different planes in order to manage how much traffic is along each plane.

Security

You can send sensitive traffic onto a network plane that is managed specifically for security considerations, and you can separate private data that must not be shared between tenants or customers.

All of the Pods in the cluster still use the cluster-wide default network to maintain connectivity across the cluster. Every Pod has an **eth0** interface that is attached to the cluster-wide Pod network. You can view the interfaces for a Pod by using the **oc exec -it <pod_name> -- ip a** command. If you add additional network interfaces that use Multus CNI, they are named **net1**, **net2**, ..., **netN**.

To attach additional network interfaces to a Pod, you must create configurations that define how the interfaces are attached. You specify each interface by using a Custom Resource (CR) that has a **NetworkAttachmentDefinition** type. A CNI configuration inside each of these CRs defines how that interface is created.

7.1.2. Additional networks in OpenShift Container Platform

OpenShift Container Platform provides the following CNI plug-ins for creating additional networks in your cluster:

- **bridge**: Creating a bridge-based additional network allows Pods on the same host to communicate with each other and the host.
- **host-device**: Creating a host-device additional network allows Pods access to a physical Ethernet network device on the host system.
- **macvlan**: Creating a macvlan-based additional network allows Pods on a host to communicate with other hosts and Pods on those hosts by using a physical network interface. Each Pod that is attached to a macvlan-based additional network is provided a unique MAC address.
- **ipvlan**: Creating an ipvlan-based additional network allows Pods on a host to communicate with

other hosts and Pods on those hosts, similar to a macvlan-based additional network. Unlike a macvlan-based additional network, each Pod shares the same MAC address as the parent physical network interface.

- **SR-IOV**: Creating a SR-IOV based additional network allows Pods to attach to a virtual function (VF) interface on SR-IOV capable hardware on the host system.

7.2. ATTACHING A POD TO AN ADDITIONAL NETWORK

As a cluster user you can attach a Pod to an additional network.

7.2.1. Adding a Pod to an additional network

You can add a Pod to an additional network. The Pod continues to send normal cluster related network traffic over the default network.

Prerequisites

- The Pod must be in the same namespace as the additional network.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster.

Procedure

To add a Pod to an additional network, complete the following steps:

1. Edit the Pod resource definition. If you are editing an existing Pod, run the following command to edit its definition in the default editor. Replace **<name>** with the name of the Pod to edit.

```
$ oc edit pod <name>
```

2. In the Pod resource definition, add the **k8s.v1.cni.cncf.io/networks** parameter to the Pod **metadata** mapping. The **k8s.v1.cni.cncf.io/networks** accepts a comma separated string of one or more NetworkAttachmentDefinition Custom Resource (CR) names:

```
metadata:
  annotations:
    k8s.v1.cni.cncf.io/networks: <network>[,<network>,...] 1
```

1 Replace **<network>** with the name of the additional network to associate with the Pod. To specify more than one additional network, separate each network with a comma. Do not include whitespace between the comma. If you specify the same additional network multiple times, that Pod will have multiple network interfaces attached to that network.

In the following example, two additional networks are attached to the Pod:

```
apiVersion: v1
kind: Pod
metadata:
  name: example-pod
  annotations:
    k8s.v1.cni.cncf.io/networks: net1,net2
```

```
spec:
  containers:
  - name: example-pod
    command: ["/bin/bash", "-c", "sleep 2000000000000"]
    image: centos/tools
```

3. Optional: Confirm that the annotation exists in the Pod CR by running the following command. Replace **<name>** with the name of the Pod.

```
$ oc describe pod <name>
```

In the following example, the **example-pod** Pod is attached to the **net1** additional network:

```
$ oc get pod example-pod -o yaml
apiVersion: v1
kind: Pod
metadata:
  annotations:
    k8s.v1.cni.cncf.io/networks: macvlan-bridge
    k8s.v1.cni.cncf.io/networks-status: |- (1)
      [{
          "name": "openshift-sdn",
          "interface": "eth0",
          "ips": [
              "10.128.2.14"
          ],
          "default": true,
          "dns": {}
      },{
          "name": "macvlan-bridge",
          "interface": "net1",
          "ips": [
              "20.2.2.100"
          ],
          "mac": "22:2f:60:a5:f8:00",
          "dns": {}
      }]
  name: example-pod
  namespace: default
spec:
  ...
status:
  ...
```

(1) The **k8s.v1.cni.cncf.io/networks-status** parameter is a JSON array of objects. Each object describes the status of an additional network attached to the Pod. The annotation value is stored as a plain text value.

7.3. REMOVING A POD FROM AN ADDITIONAL NETWORK

As a cluster user you can remove a Pod from an additional network.

7.3.1. Removing a Pod from an additional network

You can remove a Pod from an additional network.

Prerequisites

- You have configured an additional network for your cluster.
- You have an additional network attached to the Pod.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster.

Procedure

To remove a Pod from an additional network, complete the following steps:

1. Edit the Pod resource definition by running the following command. Replace **<name>** with the name of the Pod to edit.

```
$ oc edit pod <name>
```

2. Update the **annotations** mapping to remove the additional network from the Pod by performing one of the following actions:

   - To remove all additional networks from a Pod, remove the **k8s.v1.cni.cncf.io/networks** parameter from the Pod resource definition as in the following example:

```
apiVersion: v1
kind: Pod
metadata:
  name: example-pod
  annotations: {}
spec:
  containers:
  - name: example-pod
    command: ["/bin/bash", "-c", "sleep 2000000000000"]
    image: centos/tools
```

   - To remove a specific additional network from a Pod, update the **k8s.v1.cni.cncf.io/networks** parameter by removing the name of the NetworkAttachmentDefinition for the additional network.

3. Optional: Confirm that the Pod is no longer attached to the additional network by running the following command. Replace **<name>** with the name of the Pod.

```
$ oc describe pod <name>
```

   In the following example, the **example-pod** Pod is attached only to the default cluster network.

```
$ oc describe pod example-pod

Name:               example-pod
...
Annotations:        k8s.v1.cni.cncf.io/networks-status:
                    [{
```

```
                    "name": "openshift-sdn",
                    "interface": "eth0",
                    "ips": [
                        "10.131.0.13"
                    ],
                    "default": true, 1
                    "dns": {}
                }]
Status:             Running
...
```

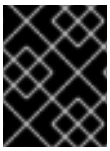
1 Only the default cluster network is attached to the Pod.

7.4. CONFIGURING A BRIDGE NETWORK

As a cluster administrator, you can configure an additional network for your cluster using the bridge Container Network Interface (CNI) plug-in. When configured, all Pods on a node are connected to a virtual switch. Each Pod is assigned an IP address on the additional network.

7.4.1. Creating an additional network attachment with the bridge CNI plug-in

The Cluster Network Operator (CNO) manages additional network definitions. When you specify an additional network to create, the CNO creates the NetworkAttachmentDefinition Custom Resource (CR) automatically.



IMPORTANT

Do not edit the NetworkAttachmentDefinition CRs that the Cluster Network Operator manages. Doing so might disrupt network traffic on your additional network.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- Log in as a user with **cluster-admin** privileges.

Procedure

To create an additional network for your cluster, complete the following steps:

1. Edit the CNO CR by running the following command:

```
$ oc edit networks.operator.openshift.io cluster
```

2. Modify the CR that you are creating by adding the configuration for the additional network you are creating, as in the following example CR.
   The following YAML configures the bridge CNI plug-in:

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
```

```
additionalNetworks: (1)
- name: test-network-1
  type: Raw
  rawCNIConfig: '{
    "cniVersion": "0.3.1",
    "type": "bridge",
    "master": "eth1",
    "ipam": {
      "type": "static",
      "addresses": [
        {
          "address": "191.168.1.1/24"
        }
      ]
    }
  }'
```

(1) Specify the configuration for the additional network attachment definition.

3. Save your changes and quit the text editor to commit your changes.

4. Optional: Confirm that the CNO created the NetworkAttachmentDefinition CR by running the following command. There might be a delay before the CNO creates the CR.

```
$ oc get network-attachment-definitions
NAME              AGE
example-network   14m
example-macvlan   21m
```

7.4.1.1. Configuration for bridge

The configuration for an additional network attachment that uses the bridge Container Network Interface (CNI) plug-in is provided in two parts:

- Cluster Network Operator (CNO) configuration
- CNI plug-in configuration

The CNO configuration specifies the name for the additional network attachment and the namespace to create the attachment in. The plug-in is configured by a JSON object specified by the **rawCNIConfig** parameter in the CNO configuration.

The following YAML describes the configuration parameters for the CNO:

Cluster Network Operator YAML configuration

```
name: <name> (1)
namespace: <namespace> (2)
rawCNIConfig: '{ (3)
  ...
}'
type: Raw
```

1. Specify a name for the additional network attachment that you are creating. The name must be unique within the specified **namespace**.
2. Specify the namespace to create the network attachment in. If you do not specify a value, then the **default** namespace is used.
3. Specify the CNI plug-in configuration in JSON format, which is based on the following template.

The following object describes the configuration parameters for the bridge CNI plug-in:

bridge CNI plug-in JSON configuration object

```
{
  "cniVersion": "0.3.1",
  "name": "<name>", 1
  "type": "bridge",
  "bridge": "<bridge>", 2
  "ipam": { 3
    ...
  },
  "ipMasq": false, 4
  "isGateway": false, 5
  "isDefaultGateway": false, 6
  "forceAddress": false, 7
  "hairpinMode": false, 8
  "promiscMode": false, 9
  "vlan": <vlan>, 10
  "mtu": <mtu> 11
}
```

1. Specify the value for the **name** parameter you provided previously for the CNO configuration.
2. Specify the name of the virtual bridge to use. If the bridge interface does not exist on the host, it is created. The default value is **cni0**.
3. Specify a configuration object for the ipam CNI plug-in. The plug-in manages IP address assignment for the network attachment definition.
4. Set to **true** to enable IP masquerading for traffic that leaves the virtual network. The source IP address for all traffic is rewritten to the bridge's IP address. If the bridge does not have an IP address, this setting has no effect. The default value is **false**.
5. Set to **true** to assign an IP address to the bridge. The default value is **false**.
6. Set to **true** to configure the bridge as the default gateway for the virtual network. The default value is **false**. If **isDefaultGateway** is set to **true**, then **isGateway** is also set to **true** automatically.
7. Set to **true** to allow assignment of a previously assigned IP address to the virtual bridge. When set to **false**, if an IPv4 address or an IPv6 address from overlapping subsets is assigned to the virtual bridge, an error occurs. The default value is **false**.
8. Set to **true** to allow the virtual bridge to send an ethernet frame back through the virtual port it was received on. This mode is also known as *reflective relay*. The default value is **false**.

9 Set to **true** to enable promiscuous mode on the bridge. The default value is **false**.

10 Specify a virtual LAN (VLAN) tag as an integer value. By default, no VLAN tag is assigned.

11 Set the maximum transmission unit (MTU) to the specified value. The default value is automatically set by the kernel.

7.4.1.1.1. bridge configuration example

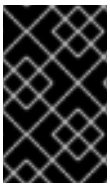
The following example configures an additional network named **bridge-net**:

```
name: bridge-net
type: Raw
rawCNIConfig: '{ (1)
  "cniVersion": "0.3.1",
  "type": "bridge",
  "master": "eth1",
  "isGateway": true,
  "vlan": 2,
  "ipam": {
    "type": "dhcp"
    }
}'
```

1 The CNI configuration object is specified as a YAML string.

7.4.1.2. Configuration for ipam CNI plug-in

The IP address management (IPAM) CNI plug-in manages IP address assignment for other CNI plug-ins. You can configure ipam for either static IP address assignment or dynamic IP address assignment by using DHCP. The DHCP server you specify must be reachable from the additional network.



IMPORTANT

In OpenShift Container Platform 4.2.0, if you attach a Pod to an additional network that uses DHCP for IP address management, the Pod will fail to start. This will be fixed in a future release. For more information, see BZ#1754686.

The following JSON configuration object describes the parameters that you can set.

IMPORTANT

If you set the **type** parameter to the **DHCP** value, you cannot set any other parameters.

ipam CNI plug-in JSON configuration object

```
{
  "ipam": {
    "type": "<type>", (1)
    "addresses": [ (2)
      {
```

```
        "address": "<address>", 3
        "gateway": "<gateway>" 4
      }
    ],
    "routes": [ 5
      {
        "dst": "<dst>" 6
        "gw": "<gw>" 7
      }
    ],
    "dns": { 8
      "nameservers": ["<nameserver>"], 9
      "domain": "<domain>", 10
      "search": ["<search_domain>"] 11
    }
  }
}
```

1. Specify **static** to configure the plug-in to manage IP address assignment. Specify **DHCP** to allow a DHCP server to manage IP address assignment. You cannot specify any additional parameters if you specify a value of **DHCP**.
2. An array of that define IP addresses to assign to the virtual interface. Both IPv4 and IPv6 IP addresses are supported.
3. A block of IP addresses that you specify in CIDR format to assign to Pods on a worker node, such as **10.1.1.0/24**.
4. The default gateway to route egress network traffic to.
5. An array describing routes to configure inside the Pod.
6. The IP address range in CIDR format.
7. The gateway to use to route network traffic to.
8. The DNS configuration. Optional.
9. An of array of one or more IP addresses for to send DNS queries to.
10. The default domain to append to a host name. For example, if the domain is set to **example.com**, a DNS lookup query for **example-host** will be rewritten as **example-host.example.com**.
11. An array of domain names to append to an unqualified host name, such as **example-host**, during a DNS lookup query.

7.4.1.2.1. Static IP address assignment configuration example

You can configure ipam for static IP address assignment:

```
{
  "ipam": {
    "type": "static",
      "addresses": [
```

```
        {
          "address": "191.168.1.1/24"
        }
      ]
  }
}
```

7.4.1.2.2. Dynamic IP address assignment configuration example

You can configure ipam for DHCP:

```
{
  "ipam": {
    "type": "DHCP"
  }
}
```

Next steps

- Attach a Pod to an additional network .

7.5. CONFIGURING A MACVLAN NETWORK

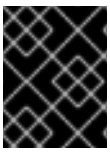
As a cluster administrator, you can configure an additional network for your cluster using the macvlan CNI plug-in. When a Pod is attached to the network, the plug-in creates a sub-interface from the parent interface on the host. A unique hardware mac address is generated for each sub-device.

IMPORTANT

The unique MAC addresses this plug-in generates for sub-interfaces might not be compatible with the security polices of your cloud provider.

7.5.1. Creating an additional network attachment with the macvlan CNI plug-in

The Cluster Network Operator (CNO) manages additional network definitions. When you specify an additional network to create, the CNO creates the NetworkAttachmentDefinition Custom Resource (CR) automatically.



IMPORTANT

Do not edit the NetworkAttachmentDefinition CRs that the Cluster Network Operator manages. Doing so might disrupt network traffic on your additional network.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- Log in as a user with **cluster-admin** privileges.

Procedure

To create an additional network for your cluster, complete the following steps:

1. Edit the CNO CR by running the following command:

```
$ oc edit networks.operator.openshift.io cluster
```

2. Modify the CR that you are creating by adding the configuration for the additional network you are creating, as in the following example CR.
   The following YAML configures the macvlan CNI plug-in:

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
  additionalNetworks: (1)
  - name: example-addn-network
    type: SimpleMacvlan
    simpleMacvlanConfig:
      ipamConfig:
        type: static
        staticIPAMConfig:
          addresses:
          - address: 10.1.1.0/24
```

(1) Specify the configuration for the additional network attachment definition.

3. Save your changes and quit the text editor to commit your changes.

4. Optional: Confirm that the CNO created the NetworkAttachmentDefinition CR by running the following command. There might be a delay before the CNO creates the CR.

```
$ oc get network-attachment-definitions
NAME              AGE
example-network     14m
example-macvlan     21m
```

7.5.1.1. Configuration for macvlan CNI plug-in

The following YAML describes the configuration parameters for the macvlan Container Network Interface (CNI) plug-in:

macvlan YAML configuration

```
name: <name> (1)
namespace: <namespace> (2)
type: SimpleMacvlan
simpleMacvlanConfig:
  master: <master> (3)
  mode: <mode> (4)
  mtu: <mtu> (5)
  ipamConfig: (6)
    ...
```

1. Specify a name for the additional network attachment that you are creating. The name must be unique within the specified **namespace**.
2. Specify the namespace to create the network attachment in. If a value is not specified, the **default** namespace is used.
3. The ethernet interface to associate with the virtual interface. If a value for **master** is not specified, then the host system's primary ethernet interface is used.
4. Configures traffic visibility on the virtual network. Must be either **bridge**, **passthru**, **private**, or **vepa**. If a value for **mode** is not provided, the default value is **bridge**.
5. Set the maximum transmission unit (MTU) to the specified value. The default value is automatically set by the kernel.
6. Specify a configuration object for the ipam CNI plug-in. The plug-in manages IP address assignment for the attachment definition.

7.5.1.1.1. macvlan configuration example

The following example configures an additional network named **macvlan-net**:

```
name: macvlan-net
type: SimpleMacvlan
simpleMacvlanConfig:
  ipamConfig:
    type: DHCP
```

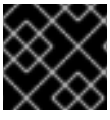
7.5.1.2. Configuration for ipam CNI plug-in

The IP address management (IPAM) CNI plug-in manages IP address assignment for other CNI plug-ins. You can configure ipam for either static IP address assignment or dynamic IP address assignment by using DHCP. The DHCP server you specify must be reachable from the additional network.

IMPORTANT

In OpenShift Container Platform 4.2.0, if you attach a Pod to an additional network that uses DHCP for IP address management, the Pod will fail to start. This will be fixed in a future release. For more information, see BZ#1754686.

The following YAML configuration describes the parameters that you can set.



IMPORTANT

If you set the **type** parameter to the **DHCP** value, you cannot set any other parameters.

ipam CNI plug-in YAML configuration object

```
ipamConfig:
  type: <type> 1
  ... 2
```

1. Specify **static** to configure the plug-in to manage IP address assignment. Specify **DHCP** to allow a DHCP server to manage IP address assignment. You cannot specify any additional parameters if
2. If you set the **type** parameter to **static**, then provide the **staticIPAMConfig** parameter.

7.5.1.2.1. Static ipam configuration YAML

The following YAML describes a configuration for static IP address assignment:

Static ipam configuration YAML

```
ipamConfig:
  type: static
  staticIPAMConfig:
    addresses: 1
    - address: <address> 2
      gateway: <gateway> 3
    routes: 4
    - destination: <destination> 5
      gateway: <gateway> 6
    dns: 7
      nameservers: 8
      - <nameserver>
      domain: <domain> 9
      search: 10
      - <search_domain>
```

1. A collection of mappings that define IP addresses to assign to the virtual interface. Both IPv4 and IPv6 IP addresses are supported.
2. A block of IP addresses that you specify in CIDR format to assign to Pods on a worker node, such as **10.1.1.0/24**.
3. The default gateway to route egress network traffic to.
4. A collection of mappings describing routes to configure inside the Pod.
5. The IP address range in CIDR format.
6. The gateway to use to route network traffic to.
7. The DNS configuration. Optional.
8. A collection of one or more IP addresses for to send DNS queries to.
9. The default domain to append to a host name. For example, if the domain is set to **example.com**, a DNS lookup query for **example-host** will be rewritten as **example-host.example.com**.
10. An array of domain names to append to an unqualified host name, such as **example-host**, during a DNS lookup query.

7.5.1.2.2. Dynamic ipam configuration YAML

The following YAML describes a configuration for static IP address assignment:

Dynamic ipam configuration YAML

```
ipamConfig:
  type: DHCP
```

7.5.1.2.3. Static IP address assignment configuration example

The following example shows an ipam configuration for static IP addresses:

```
ipamConfig:
  type: static
  staticIPAMConfig:
    addresses:
    - address: 198.51.100.11/24
      gateway: 198.51.100.10
    routes:
    - destination: 0.0.0.0/0
      gateway: 198.51.100.1
    dns:
      nameservers:
      - 198.51.100.1
      - 198.51.100.2
      domain: testDNS.example
      search:
      - testdomain1.example
      - testdomain2.example
```

7.5.1.2.4. Dynamic IP address assignment configuration example

The following example shows an ipam configuration for DHCP:

```
ipamConfig:
  type: DHCP
```

Next steps

- Attach a Pod to an additional network .

7.6. CONFIGURING AN IPVLAN NETWORK

As a cluster administrator, you can configure an additional network for your cluster by using the ipvlan Container Network Interface (CNI) plug-in. The virtual network created by this plug-in is associated with a physical interface that you specify.

7.6.1. Creating an additional network attachment with the ipvlan CNI plug-in

The Cluster Network Operator (CNO) manages additional network definitions. When you specify an additional network to create, the CNO creates the NetworkAttachmentDefinition Custom Resource (CR) automatically.

IMPORTANT

Do not edit the NetworkAttachmentDefinition CRs that the Cluster Network Operator manages. Doing so might disrupt network traffic on your additional network.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- Log in as a user with **cluster-admin** privileges.

Procedure

To create an additional network for your cluster, complete the following steps:

1. Edit the CNO CR by running the following command:

```
$ oc edit networks.operator.openshift.io cluster
```

2. Modify the CR that you are creating by adding the configuration for the additional network you are creating, as in the following example CR.
The following YAML configures the ipvlan CNI plug-in:

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
  additionalNetworks: (1)
  - name: test-network-1
    type: Raw
    rawCNIConfig: '{
      "cniVersion": "0.3.1",
      "type": "ipvlan",
      "master": "eth1",
      "mode": "l2",
      "ipam": {
        "type": "static",
        "addresses": [
          {
            "address": "191.168.1.1/24"
          }
        ]
      }
    }'
```

(1) Specify the configuration for the additional network attachment definition.

3. Save your changes and quit the text editor to commit your changes.

4. Optional: Confirm that the CNO created the NetworkAttachmentDefinition CR by running the following command. There might be a delay before the CNO creates the CR.

```
$ oc get network-attachment-definitions
NAME                AGE
```

```
example-network     14m
example-macvlan     21m
```

7.6.1.1. Configuration for ipvlan

The configuration for an additional network attachment that uses the ipvlan Container Network Interface (CNI) plug-in is provided in two parts:

- Cluster Network Operator (CNO) configuration
- CNI plug-in configuration

The CNO configuration specifies the name for the additional network attachment and the namespace to create the attachment in. The plug-in is configured by a JSON object specified by the **rawCNIConfig** parameter in the CNO configuration.

The following YAML describes the configuration parameters for the CNO:

Cluster Network Operator YAML configuration

```
name: <name> 1
namespace: <namespace> 2
rawCNIConfig: '{ 3
  ...
}'
type: Raw
```

1. Specify a name for the additional network attachment that you are creating. The name must be unique within the specified **namespace**.
2. Specify the namespace to create the network attachment in. If you do not specify a value, then the **default** namespace is used.
3. Specify the CNI plug-in configuration in JSON format, which is based on the following template.

The following object describes the configuration parameters for the ipvlan CNI plug-in:

ipvlan CNI plug-in JSON configuration object

```
{
  "cniVersion": "0.3.1",
  "name": "<name>", 1
  "type": "ipvlan",
  "mode": "<mode>", 2
  "master": "<master>", 3
  "mtu": <mtu>, 4
  "ipam": { 5
    ...
  }
}
```

1. Specify the value for the **name** parameter you provided previously for the CNO configuration.

2 Specify the operating mode for the virtual network. The value must be **l2**, **l3**, or **l3s**. The default value is **l2**.

3 Specify the ethernet interface to associate with the network attachment. If a **master** is not specified, the interface for the default network route is used.

4 Set the maximum transmission unit (MTU) to the specified value. The default value is automatically set by the kernel.

5 Specify a configuration object for the ipam CNI plug-in. The plug-in manages IP address assignment for the attachment definition.

7.6.1.1.1. ipvlan configuration example

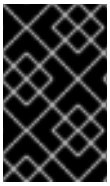
The following example configures an additional network named **ipvlan-net**:

```
name: ipvlan-net
type: Raw
rawCNIConfig: '{ 1
  "cniVersion": "0.3.1",
  "type": "ipvlan",
  "master": "eth1",
  "mode": "l3",
  "ipam": {
    "type": "dhcp"
    }
}'
```

1 The CNI configuration object is specified as a YAML string.

7.6.1.2. Configuration for ipam CNI plug-in

The IP address management (IPAM) CNI plug-in manages IP address assignment for other CNI plug-ins. You can configure ipam for either static IP address assignment or dynamic IP address assignment by using DHCP. The DHCP server you specify must be reachable from the additional network.



IMPORTANT

In OpenShift Container Platform 4.2.0, if you attach a Pod to an additional network that uses DHCP for IP address management, the Pod will fail to start. This will be fixed in a future release. For more information, see BZ#1754686.

The following JSON configuration object describes the parameters that you can set.

IMPORTANT

If you set the **type** parameter to the **DHCP** value, you cannot set any other parameters.

ipam CNI plug-in JSON configuration object

```
{
```

```
"ipam": {
  "type": "<type>", 1
  "addresses": [ 2
    {
      "address": "<address>", 3
      "gateway": "<gateway>" 4
    }
  ],
  "routes": [ 5
    {
      "dst": "<dst>" 6
      "gw": "<gw>" 7
    }
  ],
  "dns": { 8
    "nameservers": ["<nameserver>"], 9
    "domain": "<domain>", 10
    "search": ["<search_domain>"] 11
  }
 }
}
```

1. Specify **static** to configure the plug-in to manage IP address assignment. Specify **DHCP** to allow a DHCP server to manage IP address assignment. You cannot specify any additional parameters if you specify a value of **DHCP**.
2. An array of that define IP addresses to assign to the virtual interface. Both IPv4 and IPv6 IP addresses are supported.
3. A block of IP addresses that you specify in CIDR format to assign to Pods on a worker node, such as **10.1.1.0/24**.
4. The default gateway to route egress network traffic to.
5. An array describing routes to configure inside the Pod.
6. The IP address range in CIDR format.
7. The gateway to use to route network traffic to.
8. The DNS configuration. Optional.
9. An of array of one or more IP addresses for to send DNS queries to.
10. The default domain to append to a host name. For example, if the domain is set to **example.com**, a DNS lookup query for **example-host** will be rewritten as **example-host.example.com**.
11. An array of domain names to append to an unqualified host name, such as **example-host**, during a DNS lookup query.

7.6.1.2.1. Static IP address assignment configuration example

You can configure ipam for static IP address assignment:

```
{
  "ipam": {
    "type": "static",
     "addresses": [
       {
         "address": "191.168.1.1/24"
       }
     ]
  }
}
```

7.6.1.2.2. Dynamic IP address assignment configuration example

You can configure ipam for DHCP:

```
{
  "ipam": {
    "type": "DHCP"
  }
}
```

Next steps

- Attach a Pod to an additional network .

7.7. CONFIGURING A HOST-DEVICE NETWORK

As a cluster administrator, you can configure an additional network for your cluster by using the host-device Container Network Interface (CNI) plug-in. The plug-in allows you to move the specified network device from the host's network namespace into the Pod's network namespace.

7.7.1. Creating an additional network attachment with the host-device CNI plug-in

The Cluster Network Operator (CNO) manages additional network definitions. When you specify an additional network to create, the CNO creates the NetworkAttachmentDefinition Custom Resource (CR) automatically.

IMPORTANT

Do not edit the NetworkAttachmentDefinition CRs that the Cluster Network Operator manages. Doing so might disrupt network traffic on your additional network.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- Log in as a user with **cluster-admin** privileges.

Procedure

To create an additional network for your cluster, complete the following steps:

1. Edit the CNO CR by running the following command:
   -

```
$ oc edit networks.operator.openshift.io cluster
```

2. Modify the CR that you are creating by adding the configuration for the additional network you are creating, as in the following example CR.
The following YAML configures the host-device CNI plug-in:

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
  additionalNetworks: 1
  - name: test-network-1
    type: Raw
    rawCNIConfig: '{
      "cniVersion": "0.3.1",
      "type": "host-device",
      "device": "eth1"
    }'
```

1 Specify the configuration for the additional network attachment definition.

3. Save your changes and quit the text editor to commit your changes.

4. Optional: Confirm that the CNO created the NetworkAttachmentDefinition CR by running the following command. There might be a delay before the CNO creates the CR.

```
$ oc get network-attachment-definitions
NAME              AGE
example-network   14m
example-macvlan   21m
```

7.7.1.1. Configuration for host-device

The configuration for an additional network attachment that uses the host-device Container Network Interface (CNI) plug-in is provided in two parts:

- Cluster Network Operator (CNO) configuration
- CNI plug-in configuration

The CNO configuration specifies the name for the additional network attachment and the namespace to create the attachment in. The plug-in is configured by a JSON object specified by the **rawCNIConfig** parameter in the CNO configuration.

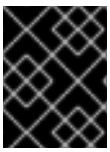
The following YAML describes the configuration parameters for the CNO:

Cluster Network Operator YAML configuration

```
name: <name> 1
namespace: <namespace> 2
rawCNIConfig: '{ 3
```

```
  ...
}'
type: Raw
```

1. Specify a name for the additional network attachment that you are creating. The name must be unique within the specified **namespace**.
2. Specify the namespace to create the network attachment in. If you do not specify a value, then the **default** namespace is used.
3. Specify the CNI plug-in configuration in JSON format, which is based on the following template.



IMPORTANT

Specify your network device by setting only one of the following parameters: **device**, **hwaddr**, **kernelpath**, or **pciBusID**.

The following object describes the configuration parameters for the host-device CNI plug-in:

host-device CNI plug-in JSON configuration object

```
{
  "cniVersion": "0.3.1",
  "name": "<name>", 1
  "type": "host-device",
  "device": "<device>", 2
  "hwaddr": "<hwaddr>", 3
  "kernelpath": "<kernelpath>", 4
  "pciBusID": "<pciBusID>" 5
}
```

1. Specify the value for the **name** parameter you provided previously for the CNO configuration.
2. Specify the name of the device, such as **eth0**.
3. Specify the device hardware MAC address.
4. Specify the Linux kernel device path, such as **/sys/devices/pci0000:00/0000:00:1f.6**.
5. Specify the PCI address of the network device, such as **0000:00:1f.6**.

7.7.1.1.1. host-device configuration example

The following example configures an additional network named **hostdev-net**:

```
name: hostdev-net
type: Raw
rawCNIConfig: '{ 1
  "cniVersion": "0.3.1",
  "type": "host-device",
  "device": "eth1"
}'
```

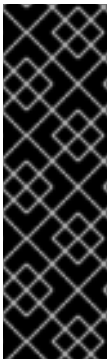
–

1 The CNI configuration object is specified as a YAML string.

Next steps

- Attach a Pod to an additional network .

7.8. CONFIGURING AN ADDITIONAL NETWORK FOR SR-IOV



IMPORTANT

SR-IOV multinetwork support is a Technology Preview feature only. Technology Preview features are not supported with Red Hat production service level agreements (SLAs) and might not be functionally complete. Red Hat does not recommend using them in production. These features provide early access to upcoming product features, enabling customers to test functionality and provide feedback during the development process.

For more information about the support scope of Red Hat Technology Preview features, see https://access.redhat.com/support/offerings/techpreview/.

7.8.1. Configuring SR-IOV

OpenShift Container Platform includes the capability to use SR-IOV hardware on OpenShift Container Platform nodes, which enables you to attach SR-IOV virtual function (VF) interfaces to Pods in addition to other network interfaces.

Two components are required to provide this capability: the SR-IOV network device plug-in and the SR-IOV CNI plug-in.

- The SR-IOV network device plug-in is a Kubernetes device plug-in for discovering, advertising, and allocating SR-IOV network virtual function (VF) resources. Device plug-ins are used in Kubernetes to enable the use of limited resources, typically in physical devices. Device plug-ins give the Kubernetes scheduler awareness of which resources are exhausted, allowing Pods to be scheduled to worker nodes that have sufficient resources available.
- The SR-IOV CNI plug-in plumbs VF interfaces allocated from the SR-IOV device plug-in directly into a Pod.

7.8.1.1. Supported Devices

The following Network Interface Card (NIC) models are supported in OpenShift Container Platform:

- Intel XXV710-DA2 25G card with vendor ID 0x8086 and device ID 0x158b
- Mellanox MT27710 Family [ConnectX-4 Lx] 25G card with vendor ID 0x15b3 and device ID 0x1015
- Mellanox MT27800 Family [ConnectX-5] 100G card with vendor ID 0x15b3 and device ID 0x1017

NOTE

For Mellanox cards, ensure that SR-IOV is enabled in the firmware before provisioning VFs on the host.

7.8.1.2. Creating SR-IOV plug-ins and daemonsets

NOTE

The creation of SR-IOV VFs is not handled by the SR-IOV device plug-in and SR-IOV CNI. To provision SR-IOV VF on hosts, you must configure it manually.

To use the SR-IOV network device plug-in and SR-IOV CNI plug-in, run both plug-ins in daemon mode on each node in your cluster.

1. Create a YAML file for the **openshift-sriov** namespace with the following contents:

```
apiVersion: v1
kind: Namespace
metadata:
  name: openshift-sriov
  labels:
    name: openshift-sriov
    openshift.io/run-level: "0"
  annotations:
    openshift.io/node-selector: ""
    openshift.io/description: "Openshift SR-IOV network components"
```

2. Run the following command to create the **openshift-sriov** namespace:

```
$ oc create -f openshift-sriov.yaml
```

3. Create a YAML file for the **sriov-device-plugin** service account with the following contents:

```
apiVersion: v1
kind: ServiceAccount
metadata:
  name: sriov-device-plugin
  namespace: openshift-sriov
```

4. Run the following command to create the **sriov-device-plugin** service account:

```
$ oc create -f sriov-device-plugin.yaml
```

5. Create a YAML file for the **sriov-cni** service account with the following contents:

```
apiVersion: v1
kind: ServiceAccount
metadata:
  name: sriov-cni
  namespace: openshift-sriov
```

6. Run the following command to create the **sriov-cni** service account:

```
$ oc create -f sriov-cni.yaml
```

7. Create a YAML file for the **sriov-device-plugin** DaemonSet with the following contents:

NOTE

The SR-IOV network device plug-in daemon, when launched, will discover all the configured SR-IOV VFs (of supported NIC models) on each node and advertise discovered resources. The number of available SR-IOV VF resources that are capable of being allocated can be reviewed by describing a node with the **oc describe node <node-name>** command. The resource name for the SR-IOV VF resources is **openshift.io/sriov**. When no SR-IOV VFs are available on the node, a value of zero is displayed.

```
kind: DaemonSet
apiVersion: apps/v1
metadata:
  name: sriov-device-plugin
  namespace: openshift-sriov
  annotations:
    kubernetes.io/description: |
      This daemon set launches the SR-IOV network device plugin on each node.
spec:
  selector:
    matchLabels:
      app: sriov-device-plugin
  updateStrategy:
    type: RollingUpdate
  template:
    metadata:
      labels:
        app: sriov-device-plugin
        component: network
        type: infra
        openshift.io/component: network
    spec:
      hostNetwork: true
      nodeSelector:
        beta.kubernetes.io/os: linux
      tolerations:
      - operator: Exists
      serviceAccountName: sriov-device-plugin
      containers:
      - name: sriov-device-plugin
        image: quay.io/openshift/ose-sriov-network-device-plugin:v4.0.0
        args:
        - --log-level=10
        securityContext:
          privileged: true
        volumeMounts:
        - name: devicesock
          mountPath: /var/lib/kubelet/
          readOnly: false
        - name: net
          mountPath: /sys/class/net
          readOnly: true
      volumes:
        - name: devicesock
          hostPath:
```

```
          path: /var/lib/kubelet/
        - name: net
          hostPath:
            path: /sys/class/net
```

8. Run the following command to create the **sriov-device-plugin** DaemonSet:

```
oc create -f sriov-device-plugin.yaml
```

9. Create a YAML file for the **sriov-cni** DaemonSet with the following contents:

```
kind: DaemonSet
apiVersion: apps/v1
metadata:
  name: sriov-cni
  namespace: openshift-sriov
  annotations:
    kubernetes.io/description: |
      This daemon set launches the SR-IOV CNI plugin on SR-IOV capable worker nodes.
spec:
  selector:
    matchLabels:
      app: sriov-cni
  updateStrategy:
    type: RollingUpdate
  template:
    metadata:
      labels:
        app: sriov-cni
        component: network
        type: infra
        openshift.io/component: network
    spec:
      nodeSelector:
        beta.kubernetes.io/os: linux
      tolerations:
      - operator: Exists
      serviceAccountName: sriov-cni
      containers:
      - name: sriov-cni
        image: quay.io/openshift/ose-sriov-cni:v4.0.0
        securityContext:
          privileged: true
        volumeMounts:
        - name: cnibin
          mountPath: /host/opt/cni/bin
      volumes:
        - name: cnibin
          hostPath:
            path: /var/lib/cni/bin
```

10. Run the following command to create the **sriov-cni** DaemonSet:

```
$ oc create -f sriov-cni.yaml
```

7.8.1.3. Configuring additional interfaces using SR-IOV

1. Create a YAML file for the Custom Resource (CR) with SR-IOV configuration. The **name** field in the following CR has the value **sriov-conf**.

```
apiVersion: "k8s.cni.cncf.io/v1"
kind: NetworkAttachmentDefinition
metadata:
  name: sriov-conf
  annotations:
    k8s.v1.cni.cncf.io/resourceName: openshift.io/sriov (1)
spec:
  config: '{
      "type": "sriov", (2)
      "name": "sriov-conf",
      "ipam": {
        "type": "host-local",
        "subnet": "10.56.217.0/24",
        "routes": [{
          "dst": "0.0.0.0/0"
        }],
        "gateway": "10.56.217.1"
      }
    }'
```

(1) **k8s.v1.cni.cncf.io/resourceName** annotation is set to **openshift.io/sriov**.

(2) **type** is set to **sriov**.

2. Run the following command to create the **sriov-conf** CR:

```
$ oc create -f sriov-conf.yaml
```

3. Create a YAML file for a Pod which references the name of the **NetworkAttachmentDefinition** and requests one **openshift.io/sriov** resource:

```
apiVersion: v1
kind: Pod
metadata:
  name: sriovsamplepod
  annotations:
    k8s.v1.cni.cncf.io/networks: sriov-conf
spec:
  containers:
  - name: sriovsamplepod
    command: ["/bin/bash", "-c", "sleep 2000000000000"]
    image: centos/tools
    resources:
      requests:
        openshift.io/sriov: '1'
      limits:
        openshift.io/sriov: '1'
```

4. Run the following command to create the **sriovsamplepod** Pod:

```
$ oc create -f sriovsamplepod.yaml
```

5. View the additional interface by executing the **ip** command:

```
$ oc exec sriovsamplepod -- ip a
```

7.9. EDITING AN ADDITIONAL NETWORK

As a cluster administrator you can modify the configuration for an existing additional network.

7.9.1. Modifying an additional network attachment definition

As a cluster administrator, you can make changes to an existing additional network. Any existing Pods attached to the additional network will not be updated.

Prerequisites

- You have configured an additional network for your cluster.
- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- Log in as a user with **cluster-admin** privileges.

Procedure

To edit an additional network for your cluster, complete the following steps:

1. Run the following command to edit the Cluster Network Operator (CNO) CR in your default text editor:

```
$ oc edit networks.operator.openshift.io cluster
```

2. In the **additionalNetworks** collection, update the additional network with your changes.
3. Save your changes and quit the text editor to commit your changes.
4. Optional: Confirm that the CNO updated the NetworkAttachmentDefinition CR by running the following command. Replace **<network-name>** with the name of the additional network to display. There might be a delay before the CNO updates the NetworkAttachmentDefinition CR to reflect your changes.

```
$ oc get network-attachment-definitions <network-name> -o yaml
```

For example, the following console output displays a NetworkAttachmentDefinition that is named **net1**:

```
$ oc get network-attachment-definitions net1 -o go-template='{{printf "%s\n" .spec.config}}'
{ "cniVersion": "0.3.1", "type": "macvlan",
"master": "ens5",
"mode": "bridge",
"ipam":       {"type":"static","routes":[{"dst":"0.0.0.0/0","gw":"10.128.2.1"}],"addresses":
[{"address":"10.128.2.100/23","gateway":"10.128.2.1"}],"dns":{"nameservers":
["172.30.0.10"],"domain":"us-west-2.compute.internal","search":["us-west-
2.compute.internal"]}} }
```

-

7.10. REMOVING AN ADDITIONAL NETWORK

As a cluster administrator you can remove an additional network attachment.

7.10.1. Removing an additional network attachment definition

As a cluster administrator, you can remove an additional network from your OpenShift Container Platform cluster. The additional network is not removed from any Pods it is attached to.

IMPORTANT

In OpenShift Container Platform 4.2.0 you must manually delete the additional network CR after removing it from the Cluster Network Operator configuration. This will be fixed in a future release. For more information, see BZ#1755908.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- Log in as a user with **cluster-admin** privileges.

Procedure

To remove an additional network from your cluster, complete the following steps:

1. Edit the Cluster Network Operator (CNO) in your default text editor by running the following command:

```
$ oc edit networks.operator.openshift.io cluster
```

2. Modify the CR by removing the configuration from the **additionalNetworks** collection for the network attachment definition you are removing.

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
  additionalNetworks: [] 1
```

1 If you are removing the configuration mapping for the only additional network attachment definition in the **additionalNetworks** collection, you must specify an empty collection.

3. Save your changes and quit the text editor to commit your changes.
4. Delete the NetworkAttachmentDefinition CR for the additional network by running the following command. Replace **<name>** with the name of the additional network to remove.

```
$ oc delete network-attachment-definition <name>
```

5. Optional: Confirm that the additional network CR was deleted by running the following command:

```
$ oc get network-attachment-definition --all-namespaces
```

CHAPTER 8. OPENSHIFT SDN

8.1. ABOUT OPENSHIFT SDN

OpenShift Container Platform uses a software-defined networking (SDN) approach to provide a unified cluster network that enables communication between Pods across the OpenShift Container Platform cluster. This Pod network is established and maintained by the OpenShift SDN, which configures an overlay network using Open vSwitch (OVS).

OpenShift SDN provides three SDN modes for configuring the Pod network:

- The *network policy* mode allows project administrators to configure their own isolation policies using NetworkPolicy objects. NetworkPolicy is the default mode in OpenShift Container Platform 4.2.
- The *multitenant* mode provides project-level isolation for Pods and Services. Pods from different projects cannot send packets to or receive packets from Pods and Services of a different project. You can disable isolation for a project, allowing it to send network traffic to all Pods and Services in the entire cluster and receive network traffic from those Pods and Services.
- The *subnet* mode provides a flat Pod network where every Pod can communicate with every other Pod and Service. The network policy mode provides the same functionality as the subnet mode.

8.2. ASSIGNING EGRESS IPS TO A PROJECT

As a cluster administrator, you can configure OpenShift Software Defined Network (SDN) to assign one or more egress IP addresses to a project. All outgoing external connections from the specified project will share the same, fixed source IP, allowing external resources to recognize the traffic based on the egress IP. An egress IP address assigned to a project is different from the egress router, which is used to send traffic to specific destinations.

Egress IPs are implemented as additional IP addresses on the primary network interface of the node and must be in the same subnet as the node's primary IP.

IMPORTANT

Egress IPs must not be configured in any Linux network configuration files, such as **ifcfg-eth0**.

Allowing additional IP addresses on the primary network interface might require extra configuration when using some cloud or VM solutions.

You can assign egress IP addresses to namespaces by setting the **egressIPs** parameter of the **NetNamespace** resource. After an egress IP is associated with a project, OpenShift SDN allows you to assign egress IPs to hosts in two ways:

- In the *automatically assigned* approach, an egress IP address range is assigned to a node. You set the **egressCIDRs** parameter of each node's **HostSubnet** resource to indicate the range of egress IP addresses that can be hosted by a node. This is the preferred approach.

- In the *manually assigned* approach, a list of one or more egress IP address is assigned to a node. You set the **egressIPs** parameter of each node's **HostSubnet** resource to indicate the IP addresses that can be hosted by a node.

Namespaces that request an egress IP addresses are matched with nodes that are able to host those egress IP addresses, and then the egress IP addresses are assigned to those nodes. If **egressIPs** is set on a **NetNamespace** resource, but no node hosts that egress IP address, then egress traffic from the namespace will be dropped.

High availability of nodes is automatic. If a node that hosts egress IP addresses is unreachable and there are nodes that are able to host those egress IP addresses, then the egress IP addresses will move to a new node. When the original egress IP address node comes back online, the egress IP addresses automatically move to balance egress IP addresses across nodes.

IMPORTANT

You cannot use manually assigned and automatically assigned egress IP addresses on the same nodes. If you manually assign egress IP addresses from an IP address range, you must not make that range available for automatic IP assignment.

8.2.1. Enabling automatically assigned egress IPs for a namespace

In OpenShift Container Platform you can enable automatic assignment of an egress IP address for a specific namespace across one or more nodes.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must be logged in to the cluster with the **cluster-admin** role.

Procedure

1. Update the **NetNamespace** resource with the egress IP address using the following JSON:

```
$ oc patch netnamespace <project_name> --type=merge -p \ 1
 '{
   "egressIPs": [
     "<ip_address>" 2
   ]
 }'
```

1 Specify the name of the project.

2 Specify a single egress IP address. Using multiple IP addresses is not supported.

For example, to assign **project1** to an IP address of 192.168.1.100 and **project2** to an IP address of 192.168.1.101:

```
$ oc patch netnamespace project1 --type=merge -p \
  '{"egressIPs": ["192.168.1.100"]}'
$ oc patch netnamespace project2 --type=merge -p \
  '{"egressIPs": ["192.168.1.101"]}'
```

2. Indicate which nodes can host egress IP addresses by setting the **egressCIDRs** parameter for each host using the following JSON:

```
$ oc patch hostsubnet <node_name> --type=merge -p \ 1
  '{
    "egressCIDRs": [
      "<ip_address_range_1>", "<ip_address_range_2>" 2
    ]
  }'
```

1 Specify a node name.

2 Specify one or more IP address ranges in CIDR format.

For example, to set **node1** and **node2** to host egress IP addresses in the range 192.168.1.0 to 192.168.1.255:

```
$ oc patch hostsubnet node1 --type=merge -p \
  '{"egressCIDRs": ["192.168.1.0/24"]}'
$ oc patch hostsubnet node2 --type=merge -p \
  '{"egressCIDRs": ["192.168.1.0/24"]}'
```

3. OpenShift Container Platform automatically assigns specific egress IP addresses to available nodes in a balanced way. In this case, it assigns the egress IP address 192.168.1.100 to **node1** and the egress IP address 192.168.1.101 to **node2** or vice versa.

8.2.2. Configuring manually assigned egress IPs

In OpenShift Container Platform you can associate one or more egress IPs with a project.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster with a user that has the **cluster-admin** role.

Procedure

1. Update the **NetNamespace** resource by specifying the following JSON object with the desired IP addresses:

```
$ oc patch netnamespace <project> --type=merge -p \ 1
  '{
    "egressIPs": [ 2
      "<ip_address>"
    ]
  }'
```

1 Specify the name of the project.

2 Specify one or more egress IP addresses. The **egressIPs** parameter is an array.

For example, to assign the **project1** project to an IP address of **192.168.1.100**:

```
$ oc patch netnamespace project1 --type=merge \
  -p '{"egressIPs": ["192.168.1.100"]}'
```

You can set **egressIPs** to two or more IP addresses on different nodes to provide high availability. If multiple egress IP addresses are set, pods use the first IP in the list for egress, but if the node hosting that IP address fails, pods switch to using the next IP in the list after a short delay.

2. Manually assign the egress IP to the node hosts. Set the **egressIPs** parameter on the **HostSubnet** object on the node host. Using the following JSON, include as many IPs as you want to assign to that node host:

```
$ oc patch hostsubnet <node_name> --type=merge -p \ 1
  '{
    "egressIPs": [ 2
      "<ip_address_1>",
      "<ip_address_N>"
      ]
  }'
```

1 Specify the name of the project.

2 Specify one or more egress IP addresses. The **egressIPs** field is an array.

For example, to specify that **node1** should have the egress IPs **192.168.1.100**, **192.168.1.101**, and **192.168.1.102**:

```
$ oc patch hostsubnet node1 --type=merge -p \
  '{"egressIPs": ["192.168.1.100", "192.168.1.101", "192.168.1.102"]}'
```

In the previous example, all egress traffic for **project1** will be routed to the node hosting the specified egress IP, and then connected (using NAT) to that IP address.

8.3. USING MULTICAST

8.3.1. About multicast

With IP multicast, data is broadcast to many IP addresses simultaneously.

IMPORTANT

At this time, multicast is best used for low-bandwidth coordination or service discovery and not a high-bandwidth solution.

Multicast traffic between OpenShift Container Platform Pods is disabled by default. If you are using the OpenShift SDN network plug-in, you can enable multicast on a per-project basis.

When using the OpenShift SDN network plug-in in **networkpolicy** isolation mode:

- Multicast packets sent by a Pod will be delivered to all other Pods in the project, regardless of NetworkPolicy objects. Pods might be able to communicate over multicast even when they cannot communicate over unicast.
- Multicast packets sent by a Pod in one project will never be delivered to Pods in any other project, even if there are NetworkPolicy objects that allow communication between the projects.

When using the OpenShift SDN network plug-in in **multitenant** isolation mode:

- Multicast packets sent by a Pod will be delivered to all other Pods in the project.
- Multicast packets sent by a Pod in one project will be delivered to Pods in other projects only if each project is joined together and multicast is enabled in each joined project.

8.3.2. Enabling multicast between Pods

You can enable multicast between Pods for your project.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster with a user that has the **cluster-admin** role.

Procedure

- Run the following command to enable multicast for a project:

```
$ oc annotate netnamespace <namespace> \ 1
    netnamespace.network.openshift.io/multicast-enabled=true
```

1 The **namespace** for the project you want to enable multicast for.

8.3.3. Disabling multicast between Pods

You can disable multicast between Pods for your project.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster with a user that has the **cluster-admin** role.

Procedure

- Disable multicast by running the following command:

```
$ oc annotate netnamespace <namespace> \ 1
    netnamespace.network.openshift.io/multicast-enabled-
```

1 The **namespace** for the project you want to disable multicast for.

8.4. CONFIGURING NETWORK ISOLATION USING OPENSHIFT SDN

When your cluster is configured to use the multitenant isolation mode for the OpenShift SDN CNI plug-in, each project is isolated by default. Network traffic is not allowed between Pods or services in different projects in multitenant isolation mode.

You can change the behavior of multitenant isolation for a project in two ways:

- You can join one or more projects, allowing network traffic between Pods and services in different projects.
- You can disable network isolation for a project. It will be globally accessible, accepting network traffic from Pods and services in all other projects. A globally accessible project can access Pods and services in all other projects.

Prerequisites

- You must have a cluster configured to use the OpenShift SDN Container Network Interface (CNI) plug-in in multitenant isolation mode.

8.4.1. Joining projects

You can join two or more projects to allow network traffic between Pods and services in different projects.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster with a user that has the **cluster-admin** role.

Procedure

1. Use the following command to join projects to an existing project network:

```
$ oc adm pod-network join-projects --to=<project1> <project2> <project3>
```

   Alternatively, instead of specifying specific project names, you can use the **--selector=<project_selector>** option to specify projects based upon an associated label.

2. Optional: Run the following command to view the pod networks that you have joined together:

```
$ oc get netnamespaces
```

   Projects in the same pod-network have the same network ID in the **NETID** column.

8.4.2. Isolating a project

You can isolate a project so that Pods and services in other projects cannot access its Pods and services.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.

- You must log in to the cluster with a user that has the **cluster-admin** role.

Procedure

- To isolate the projects in the cluster, run the following command:

```
$ oc adm pod-network isolate-projects <project1> <project2>
```

  Alternatively, instead of specifying specific project names, you can use the **--selector=<project_selector>** option to specify projects based upon an associated label.

8.4.3. Disabling network isolation for a project

You can disable network isolation for a project.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- You must log in to the cluster with a user that has the **cluster-admin** role.

Procedure

- Run the following command for the project:

```
$ oc adm pod-network make-projects-global <project1> <project2>
```

  Alternatively, instead of specifying specific project names, you can use the **--selector=<project_selector>** option to specify projects based upon an associated label.

8.5. CONFIGURING KUBE-PROXY

The Kubernetes network proxy (kube-proxy) runs on each node and is managed by the Cluster Network Operator (CNO). kube-proxy maintains network rules for forwarding connections for endpoints associated with services.

8.5.1. About iptables rules synchronization

The synchronization period determines how frequently the Kubernetes network proxy (kube-proxy) syncs the iptables rules on a node.

A sync begins when either of the following events occurs:

- An event occurs, such as service or endpoint is added to or removed from the cluster.
- The time since the last sync exceeds the sync period defined for kube-proxy.

8.5.2. Modifying the kube-proxy configuration

You can modify the Kubernetes network proxy configuration for your cluster.

Prerequisites

- Install the OpenShift Command-line Interface (CLI), commonly known as **oc**.
- Log in to a running cluster with the **cluster-admin** role.

Procedure

1. Edit the **Network.operator.openshift.io** Custom Resource (CR) by running the following command:

```
$ oc edit network.operator.openshift.io cluster
```

2. Modify the **kubeProxyConfig** parameter in the CR with your changes to the kube-proxy configuration, such as in the following example CR:

```
apiVersion: operator.openshift.io/v1
kind: Network
metadata:
  name: cluster
spec:
  kubeProxyConfig:
    iptablesSyncPeriod: 30s
    proxyArguments:
      iptables-min-sync-period: ["30s"]
```

3. Save the file and exit the text editor.
   The syntax is validated by the **oc** command when you save the file and exit the editor. If your modifications contain a syntax error, the editor opens the file and displays an error message.

4. Run the following command to confirm the configuration update:

```
$ oc get networks.operator.openshift.io -o yaml
```

   The command returns output similar to the following example:

```
apiVersion: v1
items:
- apiVersion: operator.openshift.io/v1
  kind: Network
  metadata:
    name: cluster
  spec:
    clusterNetwork:
    - cidr: 10.128.0.0/14
      hostPrefix: 23
    defaultNetwork:
      type: OpenShiftSDN
    kubeProxyConfig:
      iptablesSyncPeriod: 30s
      proxyArguments:
        iptables-min-sync-period:
        - 30s
    serviceNetwork:
    - 172.30.0.0/16
  status: {}
kind: List
```

5. Optional: Run the following command to confirm that the Cluster Network Operator accepted the configuration change:

```
$ oc get clusteroperator network
NAME      VERSION     AVAILABLE   PROGRESSING   DEGRADED   SINCE
network   4.1.0-0.9   True        False         False      1m
```

The **AVAILABLE** field is **True** when the configuration update is applied successfully.

8.5.3. kube-proxy configuration parameters

You can modify the following **kubeProxyConfig** parameters:

Table 8.1. Parameters

Parameter	Description	Values	Default
iptablesSyncPeriod	The refresh period for **iptables** rules.	A time interval, such as **30s** or **2m**. Valid suffixes include **s**, **m**, and **h** and are described in the Go time package documentation.	**30s**
proxyArguments.iptables-min-sync-period	The minimum duration before refreshing **iptables** rules. This parameter ensures that the refresh does not happen too frequently.	A time interval, such as **30s** or **2m**. Valid suffixes include **s**, **m**, and **h** and are described in the Go time package	**30s**

CHAPTER 9. CONFIGURING ROUTES

9.1. ROUTE CONFIGURATION

9.1.1. Configuring route timeouts

You can configure the default timeouts for an existing route when you have services in need of a low timeout, which is required for Service Level Availability (SLA) purposes, or a high timeout, for cases with a slow back end.

Prerequisites

- You need a deployed Ingress Controller on a running cluster.

Procedure

1. Using the **oc annotate** command, add the timeout to the route:

```
$ oc annotate route <route_name> \
    --overwrite haproxy.router.openshift.io/timeout=<timeout><time_unit> 1
```

1 Supported time units are microseconds (us), milliseconds (ms), seconds (s), minutes (m), hours (h), or days (d).

The following example sets a timeout of two seconds on a route named **myroute**:

```
$ oc annotate route myroute --overwrite haproxy.router.openshift.io/timeout=2s
```

9.1.2. Enabling HTTP strict transport security

HTTP Strict Transport Security (HSTS) policy is a security enhancement, which ensures that only HTTPS traffic is allowed on the host. Any HTTP requests are dropped by default. This is useful for ensuring secure interactions with websites, or to offer a secure application for the user's benefit.

When HSTS is enabled, HSTS adds a Strict Transport Security header to HTTPS responses from the site. You can use the **insecureEdgeTerminationPolicy** value in a route to redirect to send HTTP to HTTPS. However, when HSTS is enabled, the client changes all requests from the HTTP URL to HTTPS before the request is sent, eliminating the need for a redirect. This is not required to be supported by the client, and can be disabled by setting **max-age=0**.

IMPORTANT

HSTS works only with secure routes (either edge terminated or re-encrypt). The configuration is ineffective on HTTP or passthrough routes.

Procedure

- To enable HSTS on a route, add the **haproxy.router.openshift.io/hsts_header** value to the edge terminated or re-encrypt route:

```
apiVersion: v1
```

```
kind: Route
metadata:
  annotations:
    haproxy.router.openshift.io/hsts_header: max-age=31536000;includeSubDomains;preload
```
(1) (2) (3)

(1) **max-age** is the only required parameter. It measures the length of time, in seconds, that the HSTS policy is in effect. The client updates **max-age** whenever a response with a HSTS header is received from the host. When **max-age** times out, the client discards the policy.

(2) **includeSubDomains** is optional. When included, it tells the client that all subdomains of the host are to be treated the same as the host.

(3) **preload** is optional. When **max-age** is greater than 0, then including **preload** in **haproxy.router.openshift.io/hsts_header** allows external services to include this site in their HSTS preload lists. For example, sites such as Google can construct a list of sites that have **preload** set. Browsers can then use these lists to determine which sites they can communicate with over HTTPS, before they have interacted with the site. Without **preload** set, browsers must have interacted with the site over HTTPS to get the header.

9.1.3. Troubleshooting throughput issues

Sometimes applications deployed through OpenShift Container Platform can cause network throughput issues such as unusually high latency between specific services.

Use the following methods to analyze performance issues if Pod logs do not reveal any cause of the problem:

- Use a packet analyzer, such as ping or tcpdump to analyze traffic between a Pod and its node. For example, run the tcpdump tool on each Pod while reproducing the behavior that led to the issue. Review the captures on both sides to compare send and receive timestamps to analyze the latency of traffic to and from a Pod. Latency can occur in OpenShift Container Platform if a node interface is overloaded with traffic from other Pods, storage devices, or the data plane.

  ```
  $ tcpdump -s 0 -i any -w /tmp/dump.pcap host <podip 1> && host <podip 2>
  ```
  (1)

  (1) **podip** is the IP address for the Pod. Run the **oc get pod <pod_name> -o wide** command to get the IP address of a Pod.

  tcpdump generates a file at **/tmp/dump.pcap** containing all traffic between these two Pods. Ideally, run the analyzer shortly before the issue is reproduced and stop the analyzer shortly after the issue is finished reproducing to minimize the size of the file. You can also run a packet analyzer between the nodes (eliminating the SDN from the equation) with:

  ```
  $ tcpdump -s 0 -i any -w /tmp/dump.pcap port 4789
  ```

- Use a bandwidth measuring tool, such as iperf, to measure streaming throughput and UDP throughput. Run the tool from the Pods first, then from the nodes, to locate any bottlenecks.
  - For information on installing and using iperf, see this Red Hat Solution.

9.1.4. Using cookies to keep route statefulness

OpenShift Container Platform provides sticky sessions, which enables stateful application traffic by ensuring all traffic hits the same endpoint. However, if the endpoint Pod terminates, whether through restart, scaling, or a change in configuration, this statefulness can disappear.

OpenShift Container Platform can use cookies to configure session persistence. The Ingress controller selects an endpoint to handle any user requests, and creates a cookie for the session. The cookie is passed back in the response to the request and the user sends the cookie back with the next request in the session. The cookie tells the Ingress Controller which endpoint is handling the session, ensuring that client requests use the cookie so that they are routed to the same Pod.

9.1.4.1. Annotating a route with a cookie

You can set a cookie name to overwrite the default, auto-generated one for the route. This allows the application receiving route traffic to know the cookie name. By deleting the cookie it can force the next request to re-choose an endpoint. So, if a server was overloaded it tries to remove the requests from the client and redistribute them.

Procedure

1. Annotate the route with the desired cookie name:

```
$ oc annotate route <route_name> router.openshift.io/<cookie_name>="-
<cookie_annotation>"
```

   For example, to annotate the cookie name of **my_cookie** to the **my_route** with the annotation of **my_cookie_annotation**:

```
$ oc annotate route my_route router.openshift.io/my_cookie="-my_cookie_annotation"
```

2. Save the cookie, and access the route:

```
$ curl $my_route -k -c /tmp/my_cookie
```

9.2. SECURED ROUTES

The following sections describe how to create re-encrypt and edge routes with custom certificates.

IMPORTANT

If you create routes in Microsoft Azure through public endpoints, the resource names are subject to restriction. You cannot create resources that use certain terms. For a list of terms that Azure restricts, see Resolve reserved resource name errors in the Azure documentation.

9.2.1. Creating a re-encrypt route with a custom certificate

You can configure a secure route using reencrypt TLS termination with a custom certificate by using the **oc create route** command.

Prerequisites

- You must have a certificate/key pair in PEM-encoded files, where the certificate is valid for the route host.

- You may have a separate CA certificate in a PEM-encoded file that completes the certificate chain.
- You must have a separate destination CA certificate in a PEM-encoded file.
- You must have a **Service** resource that you want to expose.

NOTE

Password protected key files are not supported. To remove a passphrase from a key file, use the following command:

```
$ openssl rsa -in password_protected_tls.key -out tls.key
```

Procedure

This procedure creates a **Route** resource with a custom certificate and reencrypt TLS termination. The following assumes that the certificate/key pair are in the **tls.crt** and **tls.key** files in the current working directory. You must also specify a destination CA certificate to enable the Ingress Controller to trust the service's certificate. You may also specify a CA certificate if needed to complete the certificate chain. Substitute the actual path names for **tls.crt**, **tls.key**, **cacert.crt**, and (optionally) **ca.crt**. Substitute the name of the **Service** resource that you want to expose for **frontend**. Substitute the appropriate host name for **www.example.com**.

- Create a secure **Route** resource using reencrypt TLS termination and a custom certificate:

```
$ oc create route reencrypt --service=frontend --cert=tls.crt --key=tls.key --dest-ca-
cert=destca.crt --ca-cert=ca.crt --hostname=www.example.com
```

If you examine the resulting **Route** resource, it should look similar to the following:

YAML Definition of the Secure Route

```
apiVersion: v1
kind: Route
metadata:
  name: frontend
spec:
  host: www.example.com
  to:
    kind: Service
    name: frontend
  tls:
    termination: reencrypt
    key: |-
      -----BEGIN PRIVATE KEY-----
      [...]
      -----END PRIVATE KEY-----
    certificate: |-
      -----BEGIN CERTIFICATE-----
      [...]
      -----END CERTIFICATE-----
    caCertificate: |-
      -----BEGIN CERTIFICATE-----
      [...]
```

```
  -----END CERTIFICATE-----
destinationCACertificate: |-
  -----BEGIN CERTIFICATE-----
  [...]
  -----END CERTIFICATE-----
```

See **oc create route reencrypt --help** for more options.

9.2.2. Creating an edge route with a custom certificate

You can configure a secure route using edge TLS termination with a custom certificate by using the **oc create route** command. With an edge route, the Ingress Controller terminates TLS encryption before forwarding traffic to the destination Pod. The route specifies the TLS certificate and key that the Ingress Controller uses for the route.

Prerequisites

- You must have a certificate/key pair in PEM-encoded files, where the certificate is valid for the route host.
- You may have a separate CA certificate in a PEM-encoded file that completes the certificate chain.
- You must have a **Service** resource that you want to expose.

NOTE

Password protected key files are not supported. To remove a passphrase from a key file, use the following command:

```
$ openssl rsa -in password_protected_tls.key -out tls.key
```

Procedure

This procedure creates a **Route** resource with a custom certificate and edge TLS termination. The following assumes that the certificate/key pair are in the **tls.crt** and **tls.key** files in the current working directory. You may also specify a CA certificate if needed to complete the certificate chain. Substitute the actual path names for **tls.crt**, **tls.key**, and (optionally) **ca.crt**. Substitute the name of the **Service** resource that you want to expose for **frontend**. Substitute the appropriate host name for **www.example.com**.

- Create a secure **Route** resource using edge TLS termination and a custom certificate.

```
$ oc create route edge --service=frontend --cert=tls.crt --key=tls.key --ca-cert=ca.crt --
hostname=www.example.com
```

If you examine the resulting **Route** resource, it should look similar to the following:

YAML Definition of the Secure Route

```
apiVersion: v1
kind: Route
metadata:
  name: frontend
```

```
spec:
  host: www.example.com
  to:
    kind: Service
    name: frontend
  tls:
    termination: edge
    key: |-
      -----BEGIN PRIVATE KEY-----
      [...]
      -----END PRIVATE KEY-----
    certificate: |-
      -----BEGIN CERTIFICATE-----
      [...]
      -----END CERTIFICATE-----
    caCertificate: |-
      -----BEGIN CERTIFICATE-----
      [...]
      -----END CERTIFICATE-----
```

See **oc create route edge --help** for more options.

CHAPTER 10. CONFIGURING INGRESS CLUSTER TRAFFIC

10.1. CONFIGURING INGRESS CLUSTER TRAFFIC OVERVIEW

OpenShift Container Platform provides the following methods for communicating from outside the cluster with services running in the cluster.

The methods are recommended, in order or preference:

- If you have HTTP/HTTPS, use an Ingress Controller.
- If you have a TLS-encrypted protocol other than HTTPS. For example, for TLS with the SNI header, use an Ingress Controller.
- Otherwise, use a Load Balancer, an External IP, or a **NodePort**.

Method	Purpose
Use an Ingress Controller	Allows access to HTTP/HTTPS traffic and TLS-encrypted protocols other than HTTPS (for example, TLS with the SNI header).
Automatically assign an external IP using a load balancer service	Allows traffic to non-standard ports through an IP address assigned from a pool.
Manually assign an external IP to a service	Allows traffic to non-standard ports through a specific IP address.
Configure a **NodePort**	Expose a service on all nodes in the cluster.

10.2. CONFIGURING INGRESS CLUSTER TRAFFIC USING AN INGRESS CONTROLLER

OpenShift Container Platform provides methods for communicating from outside the cluster with services running in the cluster. This method uses an Ingress Controller.

10.2.1. Using Ingress Controllers and routes

The Ingress Operator manages Ingress Controllers and wildcard DNS.

Using an Ingress Controller is the most common way to allow external access to an OpenShift Container Platform cluster.

An Ingress Controller is configured to accept external requests and proxy them based on the configured routes. This is limited to HTTP, HTTPS using SNI, and TLS using SNI, which is sufficient for web applications and services that work over TLS with SNI.

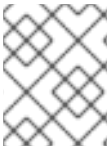
Work with your administrator to configure an Ingress Controller to accept external requests and proxy them based on the configured routes.

The administrator can create a wildcard DNS entry and then set up an Ingress Controller. Then, you can work with the edge Ingress Controller without having to contact the administrators.

When a set of routes is created in various projects, the overall set of routes is available to the set of Ingress Controllers. Each Ingress Controller admits routes from the set of routes. By default, all Ingress Controllers admit all routes.

The Ingress Controller:

- Has two replicas by default, which means it should be running on two worker nodes.
- Can be scaled up to have more replicas on more nodes.



NOTE

The procedures in this section require prerequisites performed by the cluster administrator.

Prerequisites

Before starting the following procedures, the administrator must:

- Set up the external port to the cluster networking environment so that requests can reach the cluster.
- Make sure there is at least one user with cluster admin role. To add this role to a user, run the following command:

```
oc adm policy add-cluster-role-to-user cluster-admin username
```

- Have an OpenShift Container Platform cluster with at least one master and at least one node and a system outside the cluster that has network access to the cluster. This procedure assumes that the external system is on the same subnet as the cluster. The additional networking required for external systems on a different subnet is out-of-scope for this topic.

10.2.2. Creating a project and service

If the project and service that you want to expose do not exist, first create the project, then the service.

If the project and service already exist, go to the next step: **Exposing the service to create a route**

1. Log in to OpenShift Container Platform.
2. Create a new project for your service:

```
$ oc new-project <project_name>
```

   For example:

```
$ oc new-project <myproject>
```

3. Use the **oc new-app** command to create a service:
   For example:

```
$ oc new-app \
    -e MYSQL_USER=admin \
    -e MYSQL_PASSWORD=redhat \
    -e MYSQL_DATABASE=mysqldb \
    registry.redhat.io/openshift3/mysql-55-rhel7
```

4. Run the following command to see that the new service is created:

```
$ oc get svc -n openshift-ingress
NAME                       TYPE          CLUSTER-IP      EXTERNAL-IP      PORT(S)
AGE
router-default             LoadBalancer   172.30.16.119   52.230.228.163
80:30745/TCP,443:32561/TCP   2d6h
router-internal-default    ClusterIP      172.30.101.15   <none>
80/TCP,443/TCP,1936/TCP      2d6h
```

By default, the new service does not have an external IP address.

10.2.3. Exposing the service by creating a route

You can expose the service as a route using the **oc expose** command.

To expose the service:

1. Log in to OpenShift Container Platform.
2. Log in to the project where the service you want to expose is located.

```
$ oc project project1
```

3. Run the following command to expose the route:

```
oc expose service <service-name>
```

For example:

```
oc expose service mysql-55-rhel7
route "mysql-55-rhel7" exposed
```

4. Use a tool, such as cURL, to make sure you can reach the service using the cluster IP address for the service:

```
curl <pod-ip>:<port>
```

For example:

```
curl 172.30.131.89:3306
```

The examples in this section use a MySQL service, which requires a client application. If you get a string of characters with the **Got packets out of order** message,you are connected to the service.

If you have a MySQL client, log in with the standard CLI command:

```
$ mysql -h 172.30.131.89 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.

MySQL [(none)]>
```

10.2.4. Configuring ingress controller sharding by using route labels

Ingress Controller sharding by using route labels means that the Ingress Controller serves any route in any namespace that is selected by the route selector.

Ingress Controller sharding is useful when balancing incoming traffic load among a set of Ingress Controllers and when isolating traffic to a specific Ingress Controller. For example, company A goes to one Ingress Controller and company B to another.

Procedure

1. Edit the **router-internal.yaml** file:

```
# cat router-internal.yaml
apiVersion: v1
items:
- apiVersion: operator.openshift.io/v1
  kind: IngressController
  metadata:
    name: sharded
    namespace: openshift-ingress-operator
  spec:
    domain: <apps-sharded.basedomain.example.net>
    nodePlacement:
      nodeSelector:
        matchLabels:
          node-role.kubernetes.io/worker: ""
    routeSelector:
      matchLabels:
        type: sharded
  status: {}
kind: List
metadata:
  resourceVersion: ""
  selfLink: ""
```

2. Apply the Ingress Controller **router-internal.yaml** file:

```
# oc apply -f router-internal.yaml
```

The Ingress Controller selects routes in any namespace that have the label **type: sharded**.

10.2.5. Configuring ingress controller sharding by using namespace labels

Ingress Controller sharding by using namespace labels means that the Ingress Controller serves any route in any namespace that is selected by the namespace selector.

Ingress Controller sharding is useful when balancing incoming traffic load among a set of Ingress Controllers and when isolating traffic to a specific Ingress Controller. For example, company A goes to one Ingress Controller and company B to another.

Procedure

1. Edit the **router-internal.yaml** file:

```
# cat router-internal.yaml
apiVersion: v1
items:
- apiVersion: operator.openshift.io/v1
  kind: IngressController
  metadata:
    name: sharded
    namespace: openshift-ingress-operator
  spec:
    domain: <apps-sharded.basedomain.example.net>
    nodePlacement:
      nodeSelector:
        matchLabels:
          node-role.kubernetes.io/worker: ""
    routeSelector:
      matchLabels:
        type: sharded
  status: {}
kind: List
metadata:
  resourceVersion: ""
  selfLink: ""
```

2. Apply the Ingress Controller **router-internal.yaml** file:

```
# oc apply -f router-internal.yaml
```

The Ingress Controller selects routes in any namespace that is selected by the namespace selector that have the label **type: sharded**.

10.2.6. Additional resources

- The Ingress Operator manages wildcard DNS. For more information, see Ingress Operator in OpenShift Container Platform, Installing a cluster on bare metal , and Installing a cluster on vSphere.

10.3. CONFIGURING INGRESS CLUSTER TRAFFIC USING A LOAD BALANCER

OpenShift Container Platform provides methods for communicating from outside the cluster with services running in the cluster. This method uses a load balancer.

10.3.1. Using a load balancer to get traffic into the cluster

If you do not need a specific external IP address, you can configure a load balancer service to allow external access to an OpenShift Container Platform cluster.

A load balancer service allocates a unique IP. The load balancer has a single edge router IP, which can be a virtual IP (VIP), but is still a single machine for initial load balancing.

NOTE

If a pool is configured, it is done at the infrastructure level, not by a cluster administrator.

NOTE

The procedures in this section require prerequisites performed by the cluster administrator.

Prerequisites

Before starting the following procedures, the administrator must:

- Set up the external port to the cluster networking environment so that requests can reach the cluster.
- Make sure there is at least one user with cluster admin role. To add this role to a user, run the following command:

```
oc adm policy add-cluster-role-to-user cluster-admin username
```

- Have an OpenShift Container Platform cluster with at least one master and at least one node and a system outside the cluster that has network access to the cluster. This procedure assumes that the external system is on the same subnet as the cluster. The additional networking required for external systems on a different subnet is out-of-scope for this topic.

10.3.2. Creating a project and service

If the project and service that you want to expose do not exist, first create the project, then the service.

If the project and service already exist, go to the next step: **Exposing the service to create a route**

1. Log in to OpenShift Container Platform.
2. Create a new project for your service:

```
$ oc new-project <project_name>
```

   For example:

```
$ oc new-project <myproject>
```

3. Use the **oc new-app** command to create a service:
   For example:

```
$ oc new-app \
    -e MYSQL_USER=admin \
    -e MYSQL_PASSWORD=redhat \
    -e MYSQL_DATABASE=mysqldb \
    registry.redhat.io/openshift3/mysql-55-rhel7
```

4. Run the following command to see that the new service is created:

```
$ oc get svc -n openshift-ingress
NAME                      TYPE           CLUSTER-IP      EXTERNAL-IP      PORT(S)
AGE
router-default            LoadBalancer   172.30.16.119   52.230.228.163
80:30745/TCP,443:32561/TCP   2d6h
router-internal-default   ClusterIP      172.30.101.15   <none>
80/TCP,443/TCP,1936/TCP      2d6h
```

By default, the new service does not have an external IP address.

10.3.3. Exposing the service by creating a route

You can expose the service as a route using the **oc expose** command.

To expose the service:

1. Log in to OpenShift Container Platform.
2. Log in to the project where the service you want to expose is located.

```
$ oc project project1
```

3. Run the following command to expose the route:

```
oc expose service <service-name>
```

For example:

```
oc expose service mysql-55-rhel7
route "mysql-55-rhel7" exposed
```

4. Use a tool, such as cURL, to make sure you can reach the service using the cluster IP address for the service:

```
curl <pod-ip>:<port>
```

For example:

```
curl 172.30.131.89:3306
```

The examples in this section use a MySQL service, which requires a client application. If you get a string of characters with the **Got packets out of order** message,you are connected to the service.

If you have a MySQL client, log in with the standard CLI command:

```
$ mysql -h 172.30.131.89 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.

MySQL [(none)]>
```

10.3.4. Creating a load balancer service

Use the following procedure to create a load balancer service.

Prerequisites

- Make sure that the project and service you want to expose exist.

Procedure

To create a load balancer service:

1. Log in to OpenShift Container Platform.
2. Load the project where the service you want to expose is located.

```
$ oc project project1
```

3. Open a text file on the master node and paste the following text, editing the file as needed:

Example 10.1. Sample load balancer configuration file

```
apiVersion: v1
kind: Service
metadata:
  name: egress-2 1
spec:
  ports:
  - name: db
    port: 3306 2
  loadBalancerIP:
  type: LoadBalancer 3
  selector:
    name: mysql 4
```

1. Enter a descriptive name for the load balancer service.
2. Enter the same port that the service you want to expose is listening on.
3. Enter **loadbalancer** as the type.
4. Enter the name of the service.

4. Save and exit the file.
5. Run the following command to create the service:

```
oc create -f <file-name>
```

For example:

```
oc create -f mysql-lb.yaml
```

6. Execute the following command to view the new service:

```
$ oc get svc -n openshift-ingress
NAME                      TYPE          CLUSTER-IP      EXTERNAL-IP      PORT(S)
AGE
router-default            LoadBalancer   172.30.16.119   52.230.228.163
80:30745/TCP,443:32561/TCP   2d6h
router-internal-default   ClusterIP      172.30.101.15   <none>
80/TCP,443/TCP,1936/TCP      2d6h
```

The service has an external IP address automatically assigned if there is a cloud provider enabled.

7. On the master, use a tool, such as cURL, to make sure you can reach the service using the public IP address:

```
$ curl <public-ip>:<port>
```

++ For example:

```
$ curl 172.29.121.74:3306
```

The examples in this section use a MySQL service, which requires a client application. If you get a string of characters with the **Got packets out of order** message, you are connecting with the service:

If you have a MySQL client, log in with the standard CLI command:

```
$ mysql -h 172.30.131.89 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.

MySQL [(none)]>
```

10.4. CONFIGURING INGRESS CLUSTER TRAFFIC USING A SERVICE EXTERNAL IP

OpenShift Container Platform provides methods for communicating from outside the cluster with services running in the cluster. This method uses a service external IP.

10.4.1. Using a service external IP to get traffic into the cluster

One method to expose a service is to assign an external IP address directly to the service you want to make accessible from outside the cluster.

The external IP address that you use must be provisioned on your infrastructure platform and attached to a cluster node.

With an external IP on the service, OpenShift Container Platform sets up sets up NAT rules to allow traffic arriving at any cluster node attached to that IP address to be sent to one of the internal pods. This is similar to the internal service IP addresses, but the external IP tells OpenShift Container Platform

that this service should also be exposed externally at the given IP. The administrator must assign the IP address to a host (node) interface on one of the nodes in the cluster. Alternatively, the address can be used as a virtual IP (VIP).

These IPs are not managed by OpenShift Container Platform and administrators are responsible for ensuring that traffic arrives at a node with this IP.

NOTE

The procedures in this section require prerequisites performed by the cluster administrator.

Prerequisites

Before starting the following procedures, the administrator must:

- Set up the external port to the cluster networking environment so that requests can reach the cluster.
- Make sure there is at least one user with cluster admin role. To add this role to a user, run the following command:

```
oc adm policy add-cluster-role-to-user cluster-admin username
```

- Have an OpenShift Container Platform cluster with at least one master and at least one node and a system outside the cluster that has network access to the cluster. This procedure assumes that the external system is on the same subnet as the cluster. The additional networking required for external systems on a different subnet is out-of-scope for this topic.

10.4.2. Creating a project and service

If the project and service that you want to expose do not exist, first create the project, then the service.

If the project and service already exist, go to the next step: **Exposing the service to create a route**

1. Log in to OpenShift Container Platform.
2. Create a new project for your service:

```
$ oc new-project <project_name>
```

   For example:

```
$ oc new-project <myproject>
```

3. Use the **oc new-app** command to create a service:
   For example:

```
$ oc new-app \
    -e MYSQL_USER=admin \
    -e MYSQL_PASSWORD=redhat \
    -e MYSQL_DATABASE=mysqldb \
    registry.redhat.io/openshift3/mysql-55-rhel7
```

4. Run the following command to see that the new service is created:

```
$ oc get svc -n openshift-ingress
NAME                      TYPE          CLUSTER-IP      EXTERNAL-IP      PORT(S)
AGE
router-default            LoadBalancer   172.30.16.119   52.230.228.163
80:30745/TCP,443:32561/TCP   2d6h
router-internal-default   ClusterIP      172.30.101.15   <none>
80/TCP,443/TCP,1936/TCP      2d6h
```

By default, the new service does not have an external IP address.

10.4.3. Exposing the service by creating a route

You can expose the service as a route using the **oc expose** command.

To expose the service:

1. Log in to OpenShift Container Platform.
2. Log in to the project where the service you want to expose is located.

```
$ oc project project1
```

3. Run the following command to expose the route:

```
oc expose service <service-name>
```

For example:

```
oc expose service mysql-55-rhel7
route "mysql-55-rhel7" exposed
```

4. Use a tool, such as cURL, to make sure you can reach the service using the cluster IP address for the service:

```
curl <pod-ip>:<port>
```

For example:

```
curl 172.30.131.89:3306
```

The examples in this section use a MySQL service, which requires a client application. If you get a string of characters with the **Got packets out of order** message,you are connected to the service.

If you have a MySQL client, log in with the standard CLI command:

```
$ mysql -h 172.30.131.89 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.

MySQL [(none)]>
```

10.5. CONFIGURING INGRESS CLUSTER TRAFFIC USING A NODEPORT

OpenShift Container Platform provides methods for communicating from outside the cluster with services running in the cluster. This method uses a **NodePort**.

10.5.1. Using a NodePort to get traffic into the cluster

Use a **NodePort**-type **Service** resource to expose a service on a specific port on all nodes in the cluster. The port is specified in the **Service** resource's **.spec.ports[*].nodePort** field.

NOTE

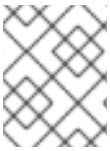
Using `NodePort`s requires additional port resources.

A node port exposes the service on a static port on the node IP address.

NodePort`s are in the `30000-32767 range by default, which means a **NodePort** is unlikely to match a service's intended port. For example, **8080** may be exposed as **31020**.

The administrator must ensure the external IPs are routed to the nodes.

`NodePort`s and external IPs are independent and both can be used concurrently.



NOTE

The procedures in this section require prerequisites performed by the cluster administrator.

Prerequisites

Before starting the following procedures, the administrator must:

- Set up the external port to the cluster networking environment so that requests can reach the cluster.
- Make sure there is at least one user with cluster admin role. To add this role to a user, run the following command:

```
oc adm policy add-cluster-role-to-user cluster-admin username
```

- Have an OpenShift Container Platform cluster with at least one master and at least one node and a system outside the cluster that has network access to the cluster. This procedure assumes that the external system is on the same subnet as the cluster. The additional networking required for external systems on a different subnet is out-of-scope for this topic.

10.5.2. Creating a project and service

If the project and service that you want to expose do not exist, first create the project, then the service.

If the project and service already exist, go to the next step: **Exposing the service to create a route**

1. Log in to OpenShift Container Platform.
2. Create a new project for your service:

```
$ oc new-project <project_name>
```

For example:

```
$ oc new-project <myproject>
```

3. Use the **oc new-app** command to create a service:
For example:

```
$ oc new-app \
    -e MYSQL_USER=admin \
    -e MYSQL_PASSWORD=redhat \
    -e MYSQL_DATABASE=mysqldb \
    registry.redhat.io/openshift3/mysql-55-rhel7
```

4. Run the following command to see that the new service is created:

```
$ oc get svc -n openshift-ingress
NAME                      TYPE          CLUSTER-IP      EXTERNAL-IP      PORT(S)
AGE
router-default            LoadBalancer   172.30.16.119   52.230.228.163
80:30745/TCP,443:32561/TCP   2d6h
router-internal-default   ClusterIP      172.30.101.15   <none>
80/TCP,443/TCP,1936/TCP      2d6h
```

By default, the new service does not have an external IP address.

10.5.3. Exposing the service by creating a route

You can expose the service as a route using the **oc expose** command.

To expose the service:

1. Log in to OpenShift Container Platform.
2. Log in to the project where the service you want to expose is located.

```
$ oc project project1
```

3. Run the following command to expose the route:

```
oc expose service <service-name>
```

For example:

```
oc expose service mysql-55-rhel7
route "mysql-55-rhel7" exposed
```

4. Use a tool, such as cURL, to make sure you can reach the service using the cluster IP address for the service:

```
curl <pod-ip>:<port>
```

For example:

```
curl 172.30.131.89:3306
```

The examples in this section use a MySQL service, which requires a client application. If you get a string of characters with the **Got packets out of order** message,you are connected to the service.

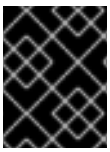
If you have a MySQL client, log in with the standard CLI command:

```
$ mysql -h 172.30.131.89 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.

MySQL [(none)]>
```

CHAPTER 11. CONFIGURING THE CLUSTER-WIDE PROXY

Production environments can deny direct access to the Internet and instead have an HTTP or HTTPS proxy available. You can configure OpenShift Container Platform to use a proxy by modifying the Proxy object for existing clusters or by configuring the proxy settings in the **install-config.yaml** file for new clusters.



IMPORTANT

The cluster-wide proxy is only supported if you used a user-provisioned infrastructure installation for a supported provider.

Prerequisites

- Review the sites that your cluster requires access to and determine whether any of them must bypass the proxy. By default, all cluster egress traffic is proxied, including calls to the cloud provider API for the cloud that hosts your cluster. Add sites to the Proxy object's **spec.noProxy** field to bypass the proxy if necessary.

NOTE

The Proxy object's **status.noProxy** field is populated by default with the instance metadata endpoint (**169.254.169.254**) and with the values of the **networking.machineCIDR**, **networking.clusterNetwork.cidr**, and **networking.serviceNetwork** fields from your installation configuration.

11.1. ENABLING THE CLUSTER-WIDE PROXY

The Proxy object is used to manage the cluster-wide egress proxy. When a cluster is installed or upgraded without the proxy configured, a Proxy object is still generated but it will have a nil **spec**. For example:

```
apiVersion: config.openshift.io/v1
kind: Proxy
metadata:
  name: cluster
spec:
  trustedCA:
    name: ""
status:
```

A cluster administrator can configure the proxy for OpenShift Container Platform by modifying this **cluster** Proxy object.

NOTE

Only the Proxy object named **cluster** is supported, and no additional proxies can be created.

Prerequisites

- Cluster administrator permissions

- OpenShift Container Platform **oc** CLI tool installed

Procedure

1. Create a ConfigMap that contains any additional CA certificates required for proxying HTTPS connections.

NOTE

You can skip this step if the proxy's identity certificate is signed by an authority from the RHCOS trust bundle.

   a. Create a file called **user-ca-bundle.yaml** with the following contents, and provide the values of your PEM-encoded certificates:

```
apiVersion: v1
data:
  ca-bundle.crt: | 1
    <MY_PEM_ENCODED_CERTS> 2
kind: ConfigMap
metadata:
  name: user-ca-bundle 3
  namespace: openshift-config 4
```

   1. This data key must be named **ca-bundle.crt**.
   2. One or more PEM-encoded X.509 certificates used to sign the proxy's identity certificate.
   3. The ConfigMap name that will be referenced from the Proxy object.
   4. The ConfigMap must be in the **openshift-config** namespace.

   b. Create the ConfigMap from this file:

```
$ oc create -f user-ca-bundle.yaml
```

2. Use the **oc edit** command to modify the Proxy object:

```
$ oc edit proxy/cluster
```

3. Configure the necessary fields for the proxy:

```
apiVersion: config.openshift.io/v1
kind: Proxy
metadata:
  name: cluster
spec:
  httpProxy: http://<username>:<pswd>@<ip>:<port> 1
  httpsProxy: http://<username>:<pswd>@<ip>:<port> 2
  noProxy: example.com 3
  readinessEndpoints:
```

```
- http://www.google.com 4
- https://www.google.com
trustedCA:
  name: user-ca-bundle 5
```

1. A proxy URL to use for creating HTTP connections outside the cluster. The URL scheme must be **http**.
2. A proxy URL to use for creating HTTPS connections outside the cluster. If this is not specified, then **httpProxy** is used for both HTTP and HTTPS connections. The URL scheme must be **http**; **https** is currently not supported.
3. A comma-separated list of destination domain names, domains, IP addresses or other network CIDRs to exclude proxying. Preface a domain with **.** to include all subdomains of that domain. Use ***** to bypass proxy for all destinations. Note that if you scale up workers not included in **networking.machineCIDR** from the installation configuration, you must add them to this list to prevent connection issues.
4. One or more URLs external to the cluster to use to perform a readiness check before writing the **httpProxy** and **httpsProxy** values to status.
5. A reference to the ConfigMap in the **openshift-config** namespace that contains additional CA certificates required for proxying HTTPS connections. Note that the ConfigMap must already exist before referencing it here. This field is required unless the proxy's identity certificate is signed by an authority from the RHCOS trust bundle.

4. Save the file to apply the changes.

11.2. REMOVING THE CLUSTER-WIDE PROXY

The **cluster** Proxy object cannot be deleted. To remove the proxy from a cluster, remove all **spec** fields from the Proxy object.

Prerequisites

- Cluster administrator permissions
- OpenShift Container Platform **oc** CLI tool installed

Procedure

1. Use the **oc edit** command to modify the proxy:

```
$ oc edit proxy/cluster
```

2. Remove all **spec** fields from the Proxy object. For example:

```
apiVersion: config.openshift.io/v1
kind: Proxy
metadata:
  name: cluster
spec: {}
status: {}
```

3. Save the file to apply the changes.