CNI Overlay

Last updated by | Leslie Cardoza | Mar 27, 2024 at 12:59 PM GMT+1

Step 1: know your enemy

Read and understand the CNI Overlay concepts.

Step 2: (re-)calculate the network size you require

With the pods taking their IP address from the Overlay network, the nodes are the only ones consuming IP addresses from the SSNS Subnet.
This calls for new math. Check how to (re-)calculate your Subnet requirements if you did not do it already.

Step 3: ensure the cluster's SSNS Subnet is using the required NSG rules

The rules you need are included in the NSG rules of the AKS SSNS pattern.
Specifically, you need those 5 rules at the time of writing:

Type	Name	Direction	Priority	Access	From address	To address	F p
Default VNet rules	Allow Inbound ALL IntraSubnet traffic between nodes in the `aksXX-subnet`	Inbound	100	Allow	<Address prefix>	<Address prefix>	*
AKS specific rules	Allow Inbound ALL from NodeCIDR to PodCIDR|Inbound	Inbound	111	Allow	<Address prefix>	`172.16.0.0/12`	*
AKS specific rule	Allow Inbound ALL from NodeCIDR to PodCIDR|Inbound	Inbound	112	Allow	`172.16.0.0/12`	`172.16.0.0/12`	*
Default VNet rules	Allow Outbound ALL IntraSubnet traffic between nodes in the `aksXX-subnet`	Outbound	100	Allow	<Address prefix>	<Address prefix>	*
AKS specific rule	Allow Outbound ALL from PodCIDR to PodCIDR	Outbound	106	Allow	`172.16.0.0/12`	`172.16.0.0/12`	*

FAQs:

- The inbound rule 112's name is not exact and should be *Allow Inbound ALL from PodCIDR to PodCIDR* instead.
  Currently being taken care of.

- Overlay networks require to be of size `/16`.
  The `/16` version of the above CIDR is smaller and **included** in the `/12` version used in the NSG rules:

  ```
  $ cidr explain 172.16.0.0/16
  Base Address:           172.16.0.0
  Usable Address Range:   172.16.0.1 to 172.16.255.254
  Broadcast Address:      172.16.255.255
  Address Count:          65,534
  Netmask:                255.255.0.0 (/16 bits)

  $ cidr explain 172.16.0.0/12
  Base Address:           172.16.0.0
  Usable Address Range:   172.16.0.1 to 172.31.255.254
  Broadcast Address:      172.31.255.255
  Address Count:          1,048,574
  Netmask:                255.240.0.0 (/12 bits)
  ```

  Go back to step 1 if you are confused here on why.

- **New** Subnets using the pattern should™ have the above rules already.

- An **existing** Subnet created **before** the SSNS pattern change will **not** have those rules.
  You will need to create ta new AKS subnet.

- There is **no** equivalent for the ADR's rule `05 | Outbound | node-to-pod | Any | Any | 10.0.0.0/24 | 172.16.0.0/16 | Allow` in the SSNS pattern.
  This is because communication should **not** be initiated by nodes to pods - only forwarded.
  Forwarding is taken care of by rule `02` in the ADR, `111` in the SSNS pattern.

Step 4: create a cluster from scratch with the Overlay feature enabled

This feature was made GA with Azure CLI 2.48.0.

One cannot reuse an **existing** cluster because changes to those features will be ignored by the ARM, see concepts.
The cluster needs to be created **anew**, with the correct `networkPluginMode` and `podCidr` attributes.
Just set the following values in the definition block for the cluster:

- The network plugin mode to `overlay`, and

- The pod CIDR to a *suitable* CIDR.

  Suitable CIDRs are any `/16` CIDR in the allowed `172.16.0.0/12` range.
  Suggested: `172.16.0.0/16`.

  Go back to step 1 if you are confused here on why.

Examples:

▶ Bicep
▶ Terraform
▶ Azure CLI

Step 5: profit

That's it!
Just continue as usual.

The only change you will notice is that pods (but **not** other resources like DaemonSets) will be assigned IP addresses from the pod CIDR you specified during the cluster's creation:

```
$ kubectl get pods -A -o custom-columns='NAME:.metadata.name,IP:.status.podIP'
NAME                                                   IP
calico-kube-controllers-77665876cb-bvxs4               172.16.2.52
calico-node-95t7v                                      10.146.88.5
calico-node-k9prc                                      10.146.88.7
calico-node-qnrh8                                      10.146.88.6
calico-typha-5b7b5d6d-gn4rn                            10.146.88.7
calico-typha-5b7b5d6d-xpfrc                            10.146.88.6
fluxconfig-agent-669549d8c8-kl28x                      172.16.0.21
fluxconfig-controller-8df65b957-dj46v                  172.16.1.124
helm-controller-79d94b9988-sj5kq                       172.16.2.155
kustomize-controller-5fdfbc5c6-sl257                   172.16.2.13
notification-controller-66bdd585c4-clpxz               172.16.0.4
source-controller-5cf948d8df-qmfm7                     172.16.1.82
gatekeeper-audit-5df84994-c5fdh                        172.16.0.25
gatekeeper-controller-556ff5899d-b2cr9                 172.16.0.2
gatekeeper-controller-556ff5899d-vjtnv                 172.16.0.227
ama-logs-6dr47                                         172.16.1.172
ama-logs-hd4rj                                         172.16.0.163
ama-logs-l9qfd                                         172.16.2.148
ama-logs-rs-847468bd57-vh99d                           172.16.1.156
azure-cns-k42zl                                        10.146.88.5
azure-cns-sw8fd                                        10.146.88.7
azure-cns-zjxrv                                        10.146.88.6
azure-ip-masq-agent-bqvz4                              10.146.88.6
azure-ip-masq-agent-cs7zg                              10.146.88.7
azure-ip-masq-agent-zwvmk                              10.146.88.5
azure-policy-86776486b4-sm52k                          172.16.0.50
azure-policy-webhook-6df78c55d9-d8zmb                  172.16.0.15
azure-wi-webhook-controller-manager-6bfd465bf8-frsdm   172.16.2.25
azure-wi-webhook-controller-manager-6bfd465bf8-gq286   172.16.1.104
cloud-node-manager-p7ptj                               10.146.88.6
cloud-node-manager-ps8nm                               10.146.88.7
cloud-node-manager-xkmsr                               10.146.88.5
coredns-76b9877f49-m6gd4                               172.16.0.58
coredns-76b9877f49-slwpt                               172.16.1.206
coredns-autoscaler-85f7d6b75d-nc4qs                    172.16.0.91
csi-azuredisk-node-2zn6t                               10.146.88.5
csi-azuredisk-node-4qs5l                               10.146.88.7
csi-azuredisk-node-p4c5n                               10.146.88.6
csi-azurefile-node-gqhm4                               10.146.88.5
csi-azurefile-node-sr7c9                               10.146.88.7
csi-azurefile-node-zvnjq                               10.146.88.6
extension-agent-5d85679847-lwdsv                       172.16.2.157
extension-operator-74d6488fd7-78j5s                    172.16.2.251
konnectivity-agent-686847b876-f2cj5                    172.16.1.229
konnectivity-agent-686847b876-r7vpf                    172.16.2.80
kube-proxy-cjr4x                                       10.146.88.7
kube-proxy-gd285                                       10.146.88.6
kube-proxy-tqhjq                                       10.146.88.5
metrics-server-5654598dc8-2qdz5                        172.16.1.30
metrics-server-5654598dc8-kgsbq                        172.16.2.152
prometheus-alertmanager-0                              172.16.2.101
prometheus-prometheus-node-exporter-h52v2              10.146.88.6
prometheus-prometheus-node-exporter-rcjr2              10.146.88.5
prometheus-prometheus-node-exporter-zz5kb              10.146.88.7
prometheus-server-57bc7cb46d-k9szb                     172.16.0.131
grafana-76bbbd7c94-78gz4                               172.16.1.28
nginx-ingress-controller-c6888d76-g268z                172.16.1.161
tigera-operator-66f9f59fb9-s5sbz                       10.146.88.7
```