Hybrid Cloud and Multi-Cluster Service Connectivity

Sridhar Gaddam
Aswin Suryanarayanan
Red Hat
Agenda

- Kubernetes Networking Model
- Use-cases
- Problem Domain
- Technologies we looked into
- Submariner Architecture
- Future work
Every POD has a unique IP Address and is shared by all the containers in the POD.

A POD can communicate with other PODs/Services in the same cluster, regardless of what node they are on.

Generally the POD IPs are isolated to the outside world and there are explicit mechanisms to allow external traffic into the Cluster.

K8s networking focuses mainly on container networking within the local cluster, but does not talk about cross-cluster network connectivity.
Multiple Clusters

Public Network

OnPrem K8s Cluster

Amazon Web Services

Azure
Use-Cases

- High Availability for your applications.
- Deploying Service Mesh across multiple clusters.
- Stretched Databases
- Enabling access to ClusterIP Services from remote clusters
- App on edge and DB onPrem
- Make use of resources which are only available on public clouds.
Goals:

- The solution should be cloud agnostic
- The solution should be CNI agnostic

Four Aspects:

- Tunnel Management - building secured L3 connectivity between the clusters
- Injecting Routing Rules
- Multi Cluster Network Policy
- Multi Cluster Service Discovery (Global DNS solution)
Problem Domain Cont...

Cluster Connectivity
- Tunnel management
- Injecting Routing Rules

Service Connectivity
- Multi Cluster Service Discovery
- Multi Cluster Network Policy

Orchestrator
Responsible for configuring, monitoring, upgrading the cross-cluster connections.
Technologies we looked into?

- In-house POCs.
  - Designed and prototyped couple of proposals.

- External Open Source Projects
  - Cilium
  - Federation (KubeFed)
  - Istio
  - Submariner
Submariner
Submariner Architecture

Broker Cluster

POD CIDR: 10.0.0.0/16
Service CIDR: 172.11.0.0/16

POD CIDR: 20.0.0.0/16
Service CIDR: 172.22.0.0/16

West Cluster
Gateway Node
Master
Worker1
Worker2

East Cluster
Gateway Node
Master
Worker1
Worker2

Public Network

clusters.submariner.io
endpoints.submariner.io

GW to Broker Connection
IPSEC Traffic
Submariner Components

- **Submariner Engine**
  - is run on the gateway nodes
  - will perform leader election
  - responsible for running/interfacing with Charon to establish IPsec tunnels
  - updates local cluster information into the central broker to share information between clusters

- **Submariner Route Agent (DaemonSet)**
  - is run on every node.
  - is aware of the current leader in the local cluster.
  - on the gateway node, it will simply sit idle awaiting leader loss
  - on other nodes inserts route rules to allow all pods/nodes to communicate through the elected gateway node to the remote cluster networks
The worker nodes in each cluster which are marked as the g/w nodes take part in the election.

- Uses Kubernetes Simple Leader Election.
- One node in each cluster will be elected as the leader.
Upon the failure of a leader in a cluster another submariner pod labelled as gateway pod gains the leadership.

- The new leader in the cluster and the remote leaders in other cluster performs the reconciliation process to re-establish the tunnel.
- The submariner route-agent in the cluster will update routes on each node to point towards the new leader.
Cluster Connectivity

What works today?

- **Tunnel Management** -
  - The connected clusters are now reachable through the ipsec tunnels created
- **Routing Rules** -
  - Each gateway node learns about the pod and service CIDR reachable in each cluster
  - Routing rules are programmed on all the worker nodes

Prerequisites:

- Knowledge of cluster configuration
- Non-overlapping POD and Service CIDRs
Packet Traversal

West Cluster

- **Master**
- **Worker1**
- **Worker2**

POD CIDR: 10.0.0.0/16
Service CIDR: 172.11.0.0/16

IPSEC Traffic

East Cluster

- **Master**
- **Worker1**
- **Worker2**

POD CIDR: 20.0.0.0/16
Service CIDR: 172.22.0.0/16

Public Network

- **Gateway Node**

1. 192.168.101.0/24
2. 192.168.201.0/24
3. 192.168.101.0/24
4. 192.168.201.0/24
5. 192.168.201.0/24
6. 192.168.201.0/24

Un-encrypted data

IPSEC Traffic
Future Work

● Currently, its Pre-Alpha, and the team is working on making it production grade
● Multi-cluster Network Policy
● Multi-cluster Service Discovery
● Support for various topologies
● Support for different types of tunnels (Currently only supports IPsec, looking at OpenVPN, VxLAN and IPinIP)
● Monitoring of the solution
● Support for Overlapping CIDRs
● Leveraging public cloud services
  ○ AWS Transit GW, GCP Network Director, Azure Gateway
Submariner github repo: https://github.com/rancher/submariner

Weekly Agenda: https://docs.google.com/document/d/1qnZ2LpF_rXGfnYYPNTldQ4WbeEUxwnuQD-xTC6GbZdg

SIG-Multicluster: Efforts in progress to move the project to sig-multicluster

Weekly calls: Tuesdays @15:00 UTC
  https://bluejeans.com/3472508766
Thank you