KubeCon | CloudNativeCon

Europe 2018

The “Silk” Road: Building a CNI Plugin From Scratch

Angela Chin, Senior Software Engineer, Pivotal
Usha Ramachandran, Staff Product Manager, Pivotal

Agenda

- What is CNI
- What is Cloud Foundry
- Cloud Foundry + CNI
- Main Motivations to Build Silk CNI
- Building Silk
- What's Next?
- Key Takeaways

Agenda

- **What is CNI**
- What is Cloud Foundry
- Cloud Foundry + CNI
- Main Motivations to Build Silk CNI
- Building Silk
- What's Next?
- Key Takeaways

What is CNI?

Container Networking Interface (CNI) is an industry standard API for container runtimes to call networking plugins

CNI Plugins

- Creates network interface for containers
- Removes resources when containers are deleted
  - Deletion is idempotent
- Network connectivity

CNI Plugins

- Binaries
- Invoked during a container creation/deletion
- Arguments passed in through combination of env variables and stdin

List of CNI Plugins

- Project Calico - a layer 3 virtual network
- Weave - a multi-host Docker network
- Contiv Networking - policy networking for various use cases
- SR-IOV
- Cilium - BPF & XDP for containers
- Infoblox - enterprise IP address management for containers
- Multus - a Multi plugin
- Romana - Layer 3 CNI plugin supporting network policy for Kubernetes
- CNI-Genie - generic CNI network plugin
- Nuage CNI - Nuage Networks SDN plugin for network policy kubernetes support
- Silk - a CNI plugin designed for Cloud Foundry
- Linen - a CNI plugin designed for overlay networks with Open vSwitch and fit in SDN/OpenFlow network environment
- Vhostuser - a Dataplane network plugin - Supports OVS-DPDK & VPP
- Amazon ECS CNI Plugins - a collection of CNI Plugins to configure containers with Amazon EC2 elastic network interfaces (ENIs)
- Bonding CNI - a Link aggregating plugin to address failover and high availability network

Source: https://github.com/containernetworking/cni#3rd-party-plugins

List of CNI Plugins

- Cilium - BPF & XDP for containers
- Infoblox - enterprise IP address management for containers
- Multus - a Multi plugin
- Romana - Layer 3 CNI plugin supporting network policy for Kubernetes
- CNI-Genie - generic CNI network plugin
- Nuage CNI - Nuage Networks SDN plugin for network policy kubernetes support
- **Silk - a CNI plugin designed for Cloud Foundry**
- Linen - a CNI plugin designed for overlay networks with Open vSwitch and fit in SDN/OpenFlow network environment
- Vhostuser - a Dataplane network plugin - Supports OVS-DPDK & VPP
- Amazon ECS CNI Plugins - a collection of CNI Plugins to configure containers with Amazon EC2 elastic network interfaces (ENIs)
- Bonding CNI - a Link aggregating plugin to address failover and high availability network

List of CNI Plugins

- Cilium - BPF & XDP for containers
- Infoblox - enterprise IP address management for containers
- Multus - a Multi plugin
- Romana - Layer 3 CNI plugin supporting network policy for Kubernetes
- CNI-Genie - generic CNI network plugin
- Nuage CNI - Nuage Networks SDN plugin for network policy kubernetes support
- **Silk - a CNI plugin designed for Cloud Foundry**
- Linen - a CNI plugin designed for overlay networks with Open vSwitch and fit in SDN/OpenFlow network environment
- Vhostuser - a Dataplane network plugin - Supports OVS-DPDK & VPP
- Amazon ECS CNI Plugins - a collection of CNI Plugins to configure containers with Amazon EC2 elastic network interfaces (ENIs)
- Bonding CNI - a Link aggregating plugin to address failover and high availability network

Agenda

- What is CNI
- **What is Cloud Foundry**
- Cloud Foundry + CNI
- Main Motivations to Build Silk CNI
- Building Silk
- What's Next?
- Key Takeaways

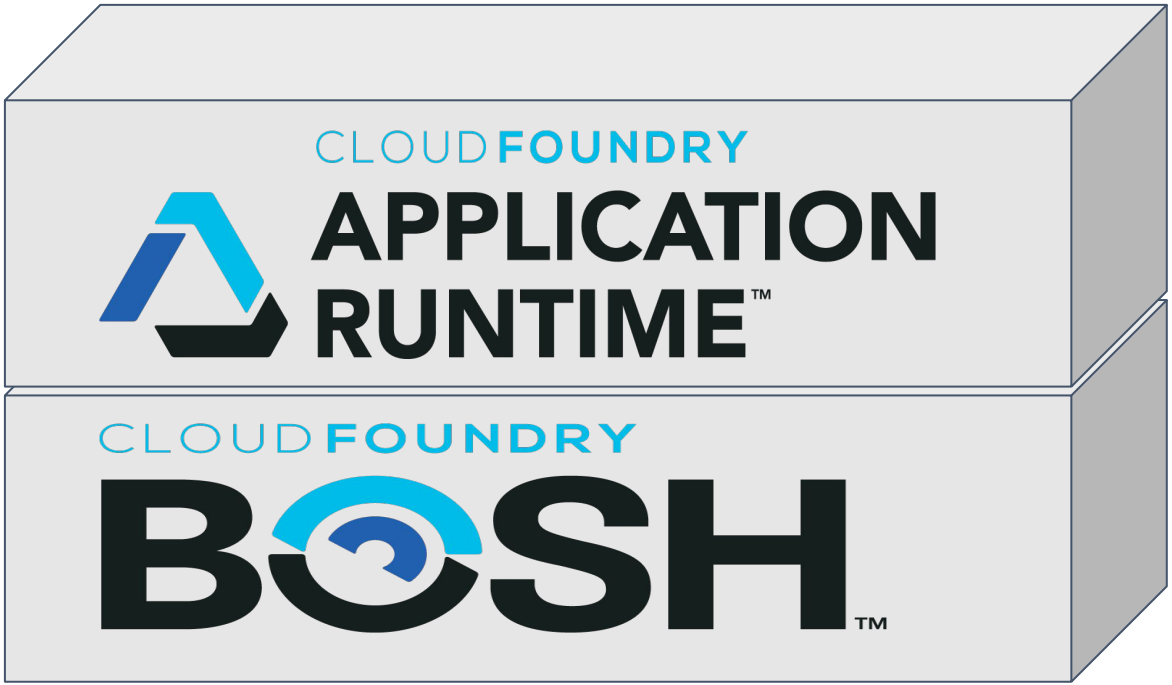
A Haiku

“Here is my source code
Run it on the cloud for me
I do not care how”
- Onsi Fakhouri

Cloud Foundry Application Runtime

Deploy and manage applications

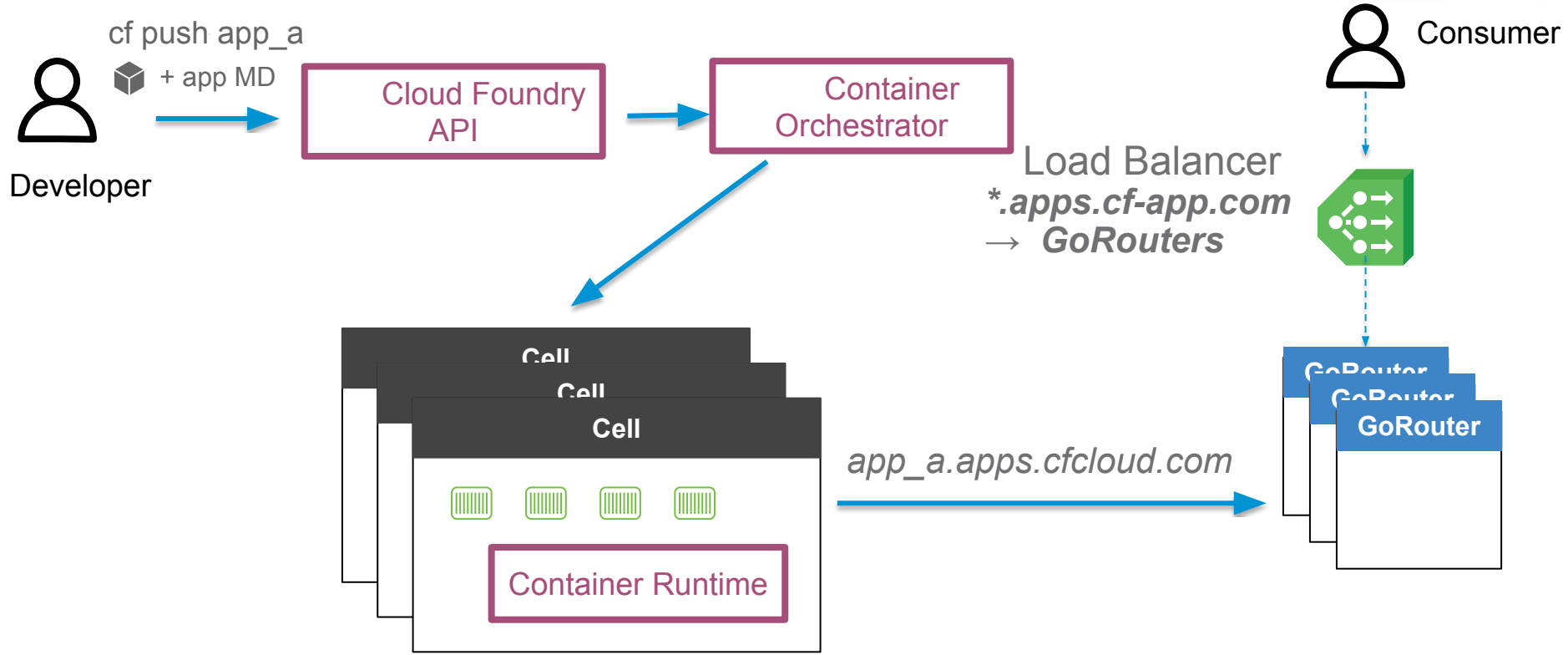
Cloud Foundry Application Runtime



Deploy and manage applications

Deploy and manage infrastructure

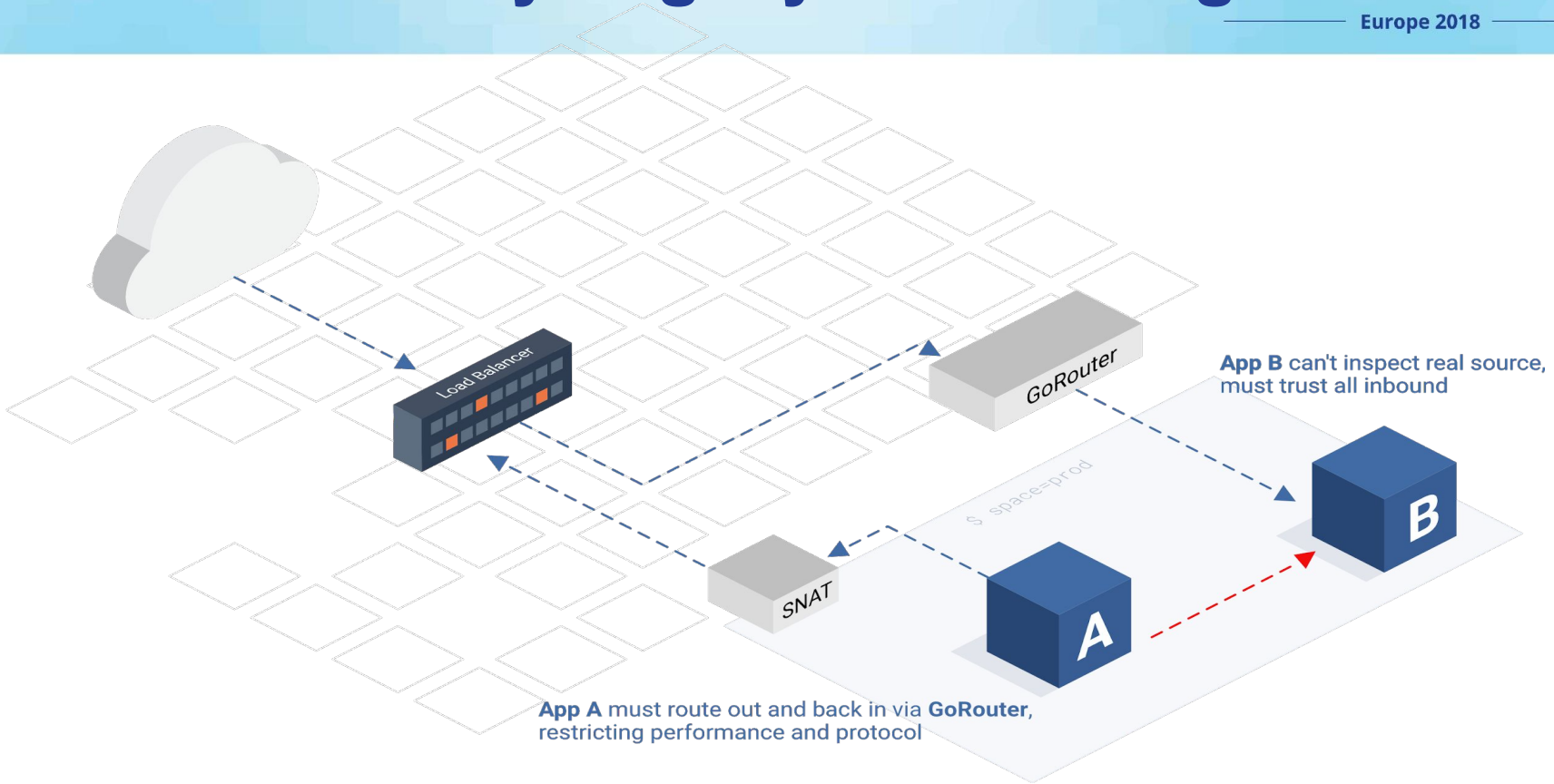
Cloud Foundry Workflow



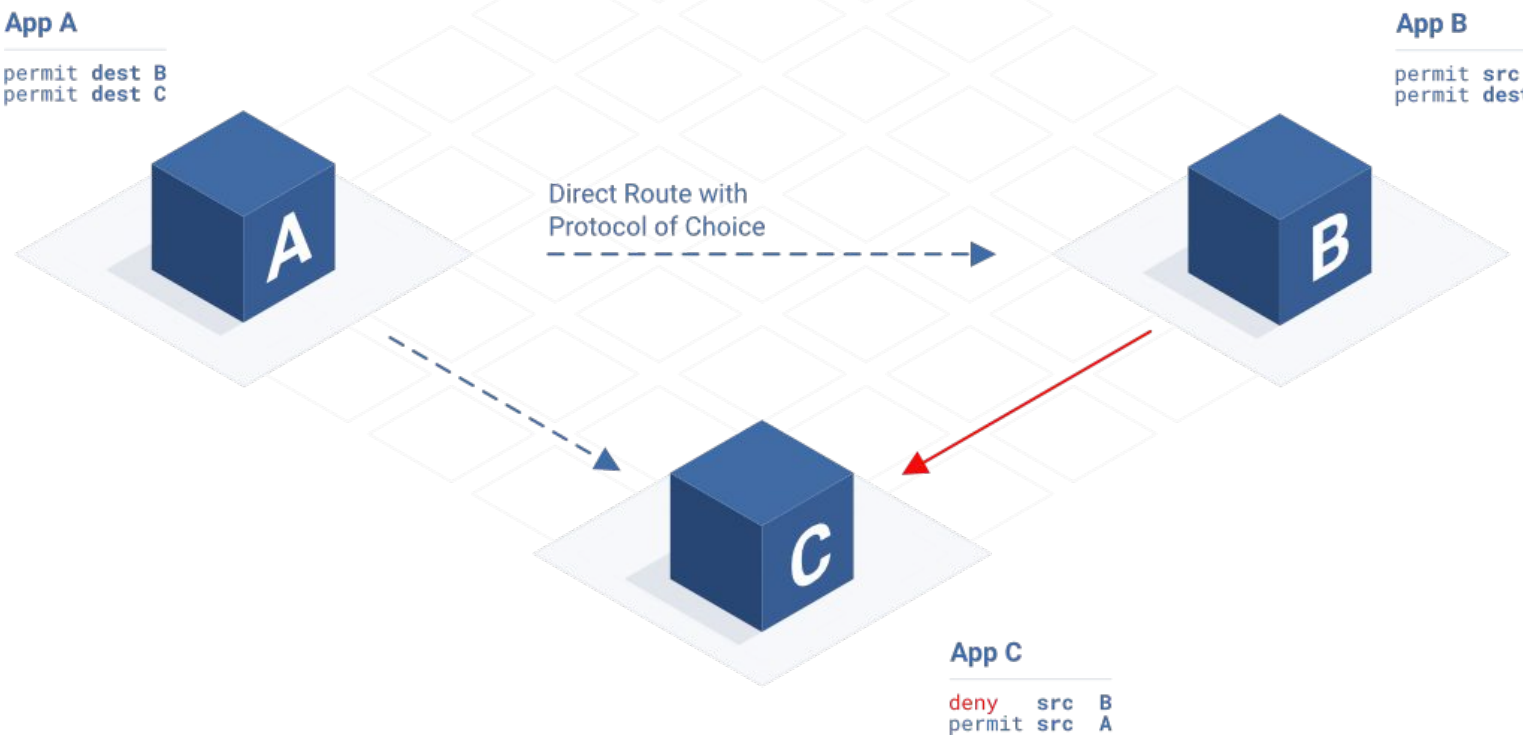
Agenda

- What is CNI
- What is Cloud Foundry
- **Cloud Foundry + CNI**
- Main Motivations to Build Silk CNI
- Building Silk
- What's Next?
- Key Takeaways

Cloud Foundry Legacy Networking



Desired State



Networking Requirements

Application Security Groups

Egress Cell IP:SNAT

Ingress Cell IP:DNAT

Networking Requirements

Application Security Groups

Egress Cell IP:SNAT

Ingress Cell IP:DNAT

IaaS agnostic

Unique IP per container

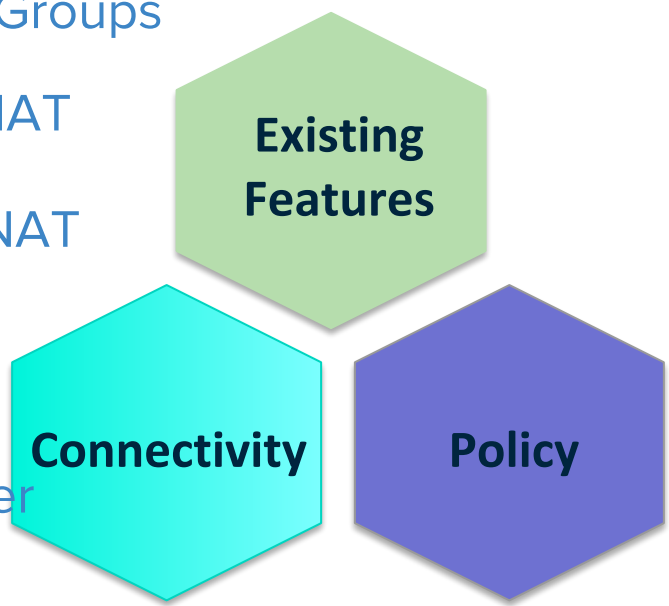
Extensible

Networking Requirements

Application Security Groups

Egress Cell IP:SNAT

Ingress Cell IP:DNAT



IaaS agnostic

Unique IP per container

Extensible

App to App

Dynamic

Self Service

Connectivity - Requirements

IaaS agnostic

Unique IP per container

Extensible

Connectivity - Solution

VXLAN Overlay

Subnet per Cell

CNI

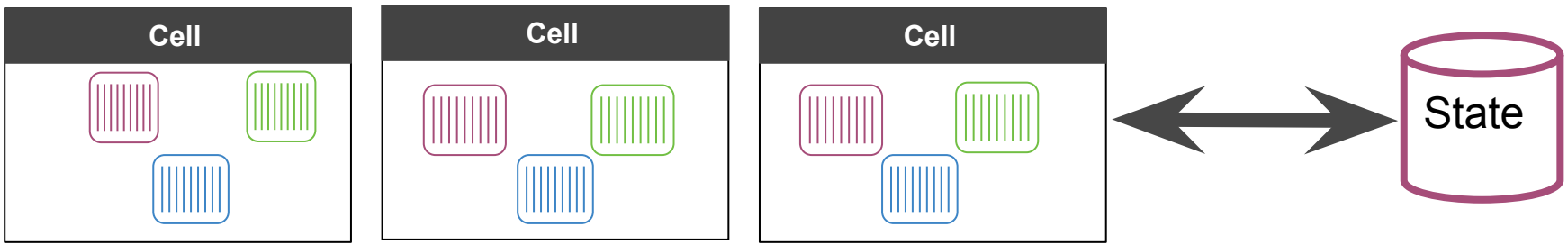
flannel

Agenda

- What is CNI
- What is Cloud Foundry
- Cloud Foundry + CNI
- **Main Motivations to Build Silk CNI**
- Building Silk
- What's Next?
- Key Takeaways

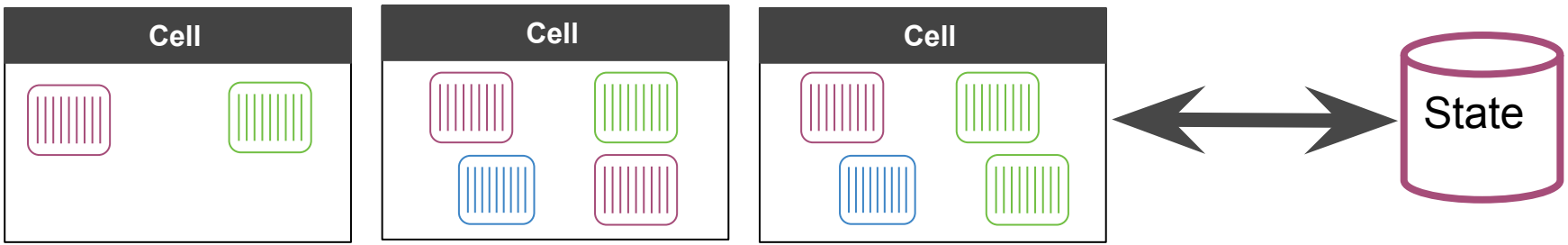
Reliability

Application instances can be added and deleted



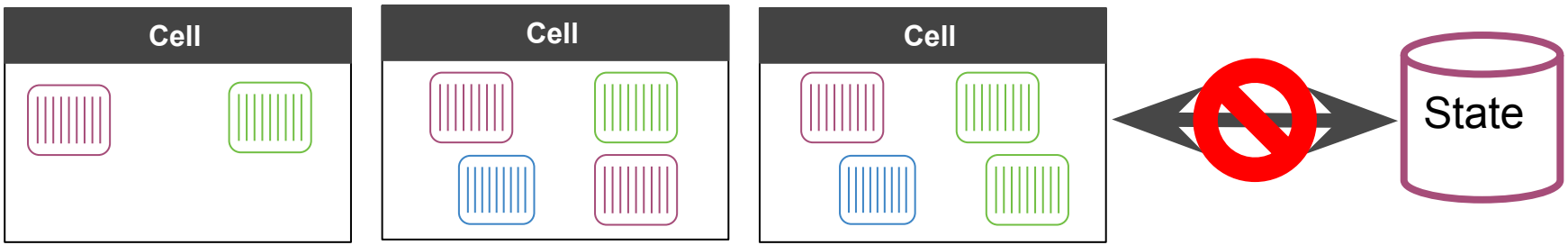
Reliability

Diego Cells can be added/deleted at any time



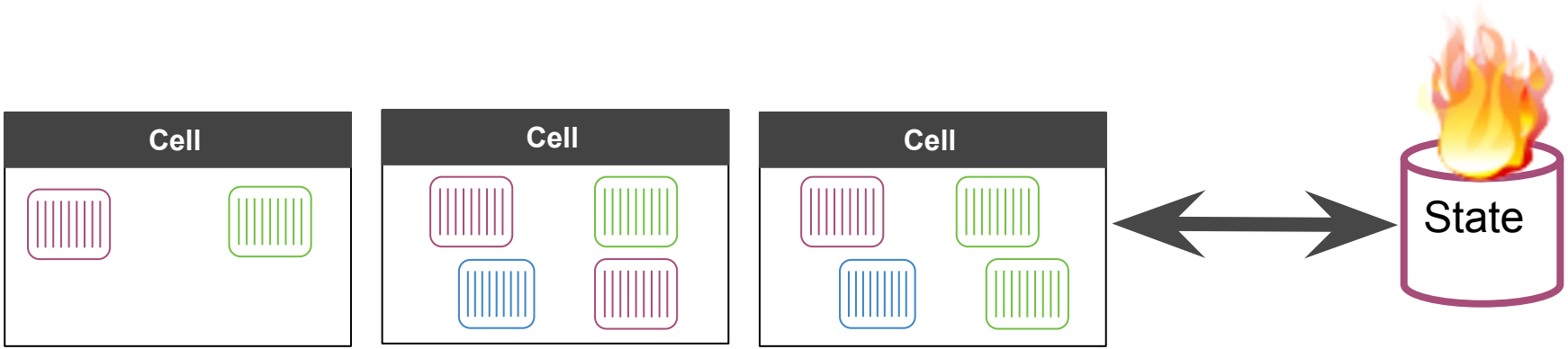
Reliability

Network Partitions



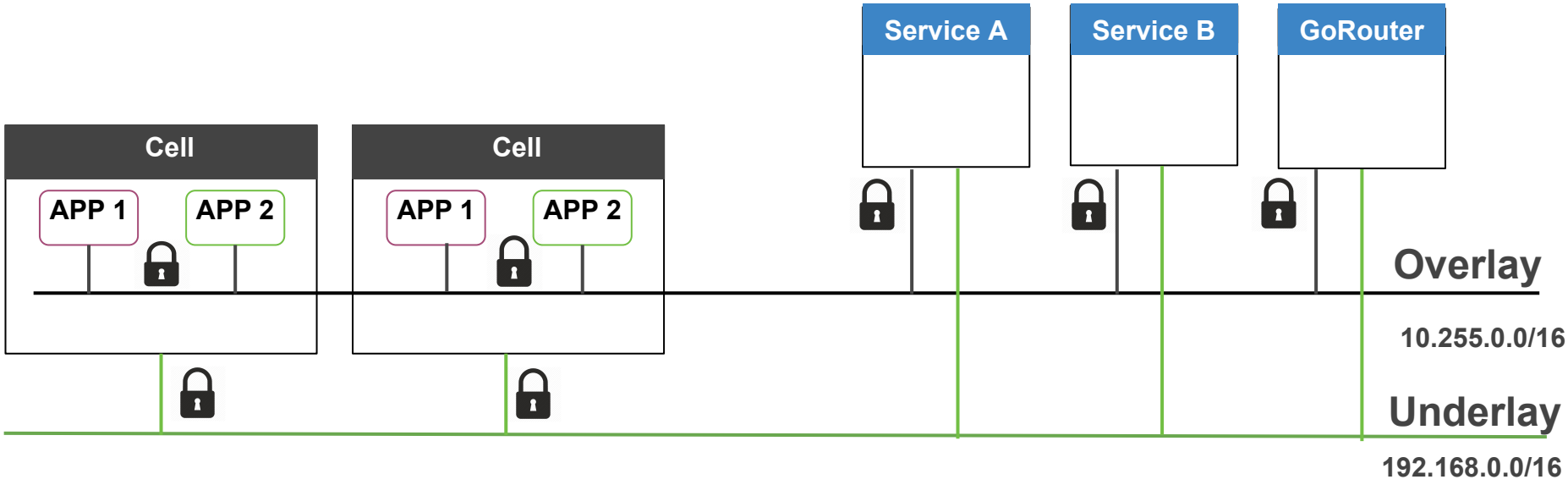
Reliability

Data Loss



Extensibility

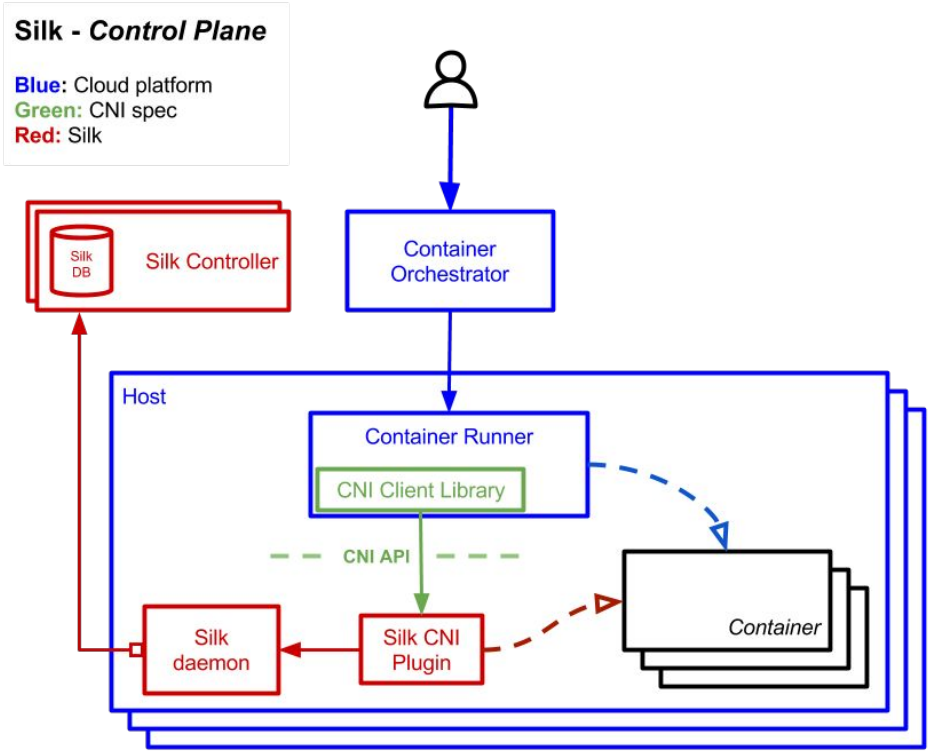
Add support for new features



Agenda

- What is CNI
- What is Cloud Foundry
- Cloud Foundry + CNI
- Main Motivations to Build Silk CNI
- **Building Silk**
- What's Next?
- Key Takeaways

Building Silk



Reliability & Stability

A lease is a subnet given to a host vm.

- Lease is given only if no containers on cell
- Network add will fail if daemon does not have a lease
- Lease is kept for as long as possible

Silk - *Control Plane*

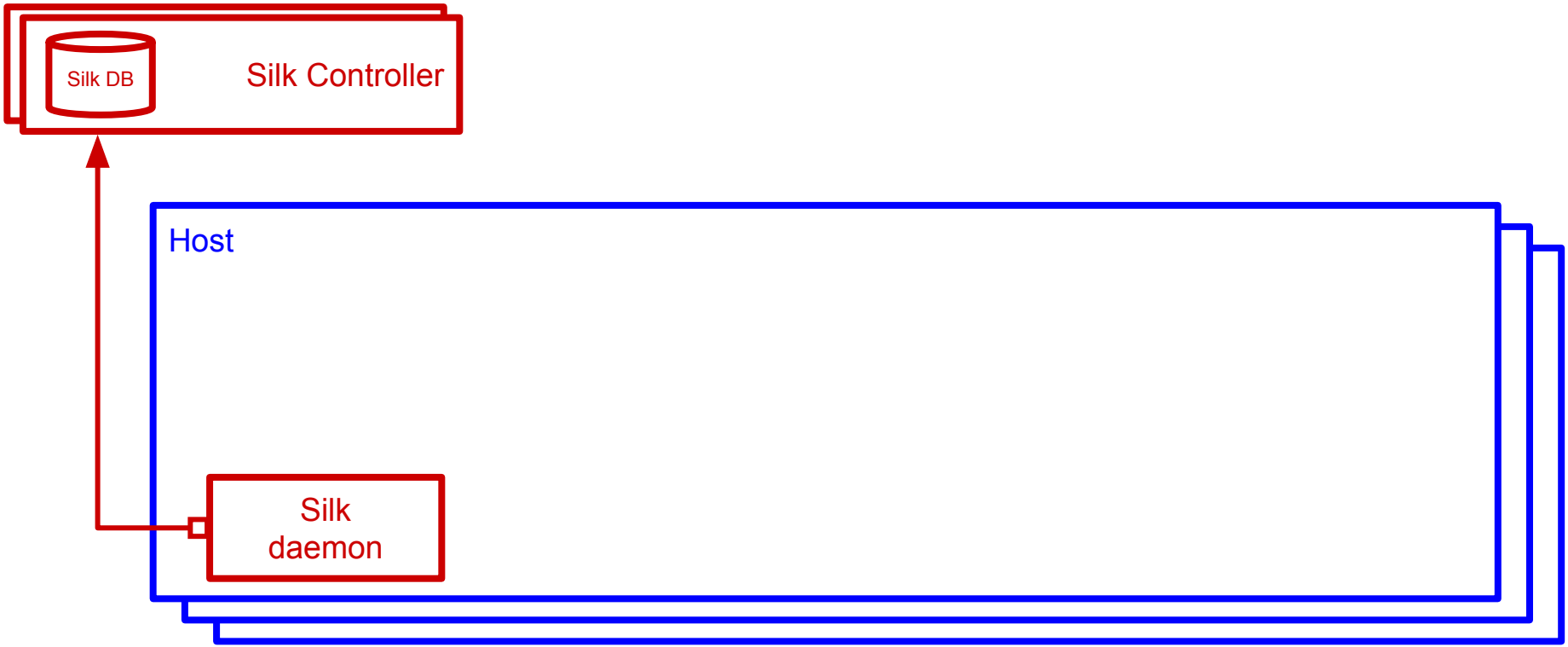
Blue: Cloud platform
Green: CNI spec
Red: Silk

How Silk Works: Subnet Assignment

Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

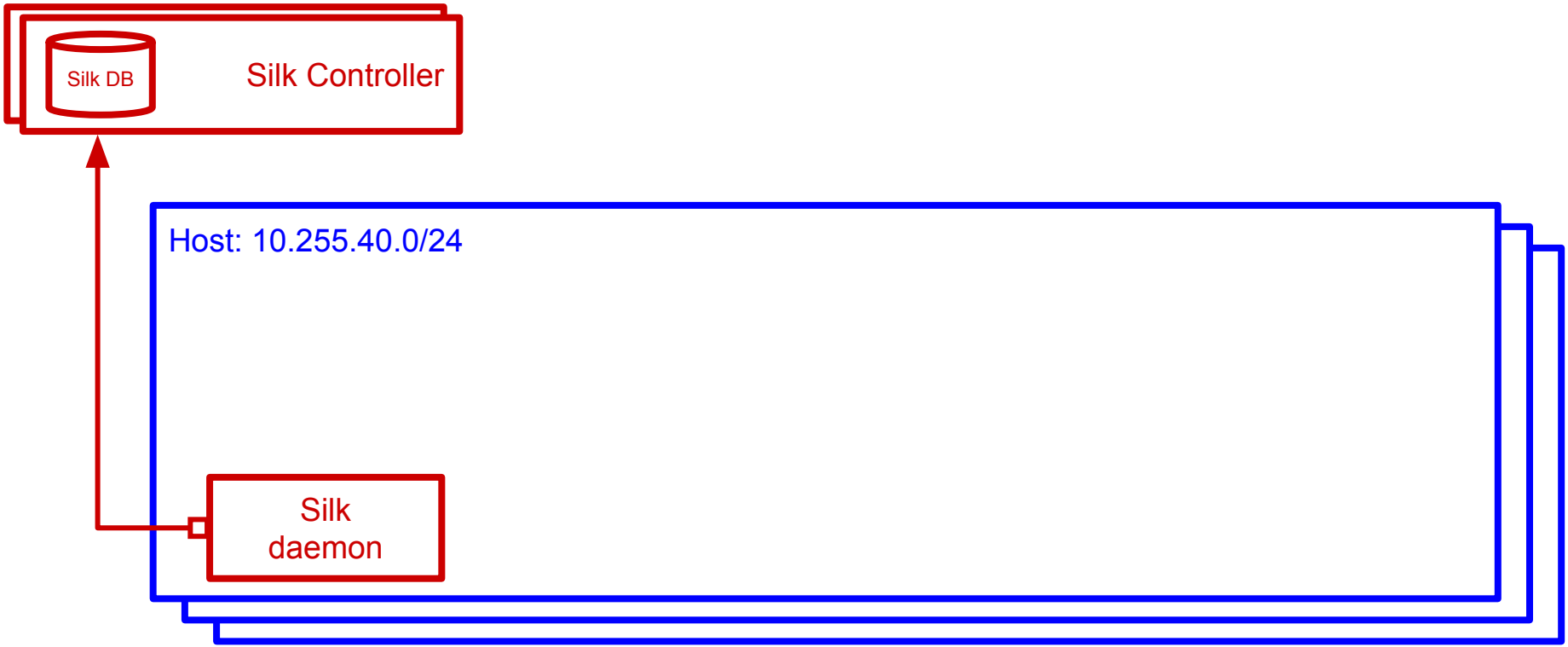
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

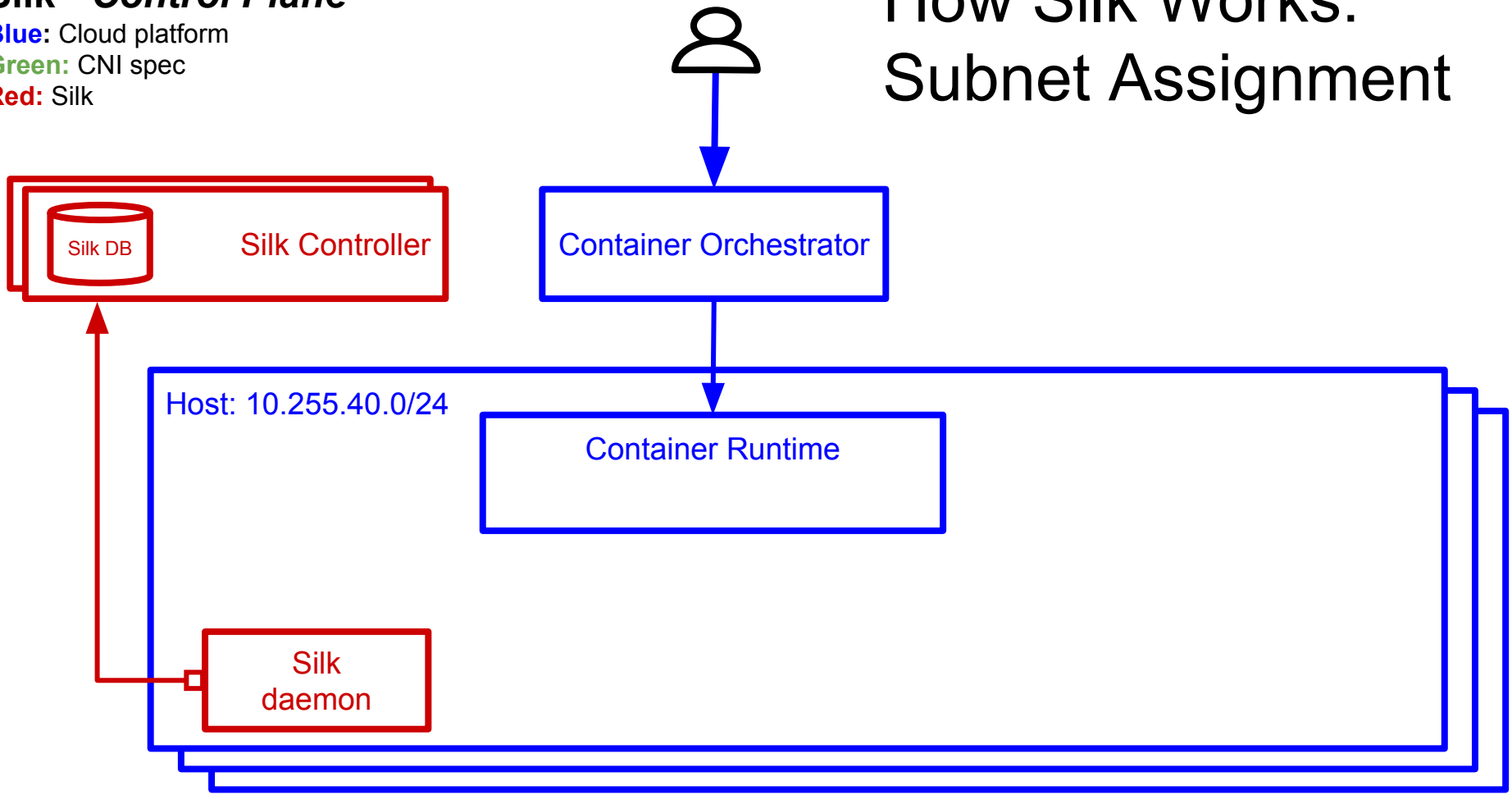
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

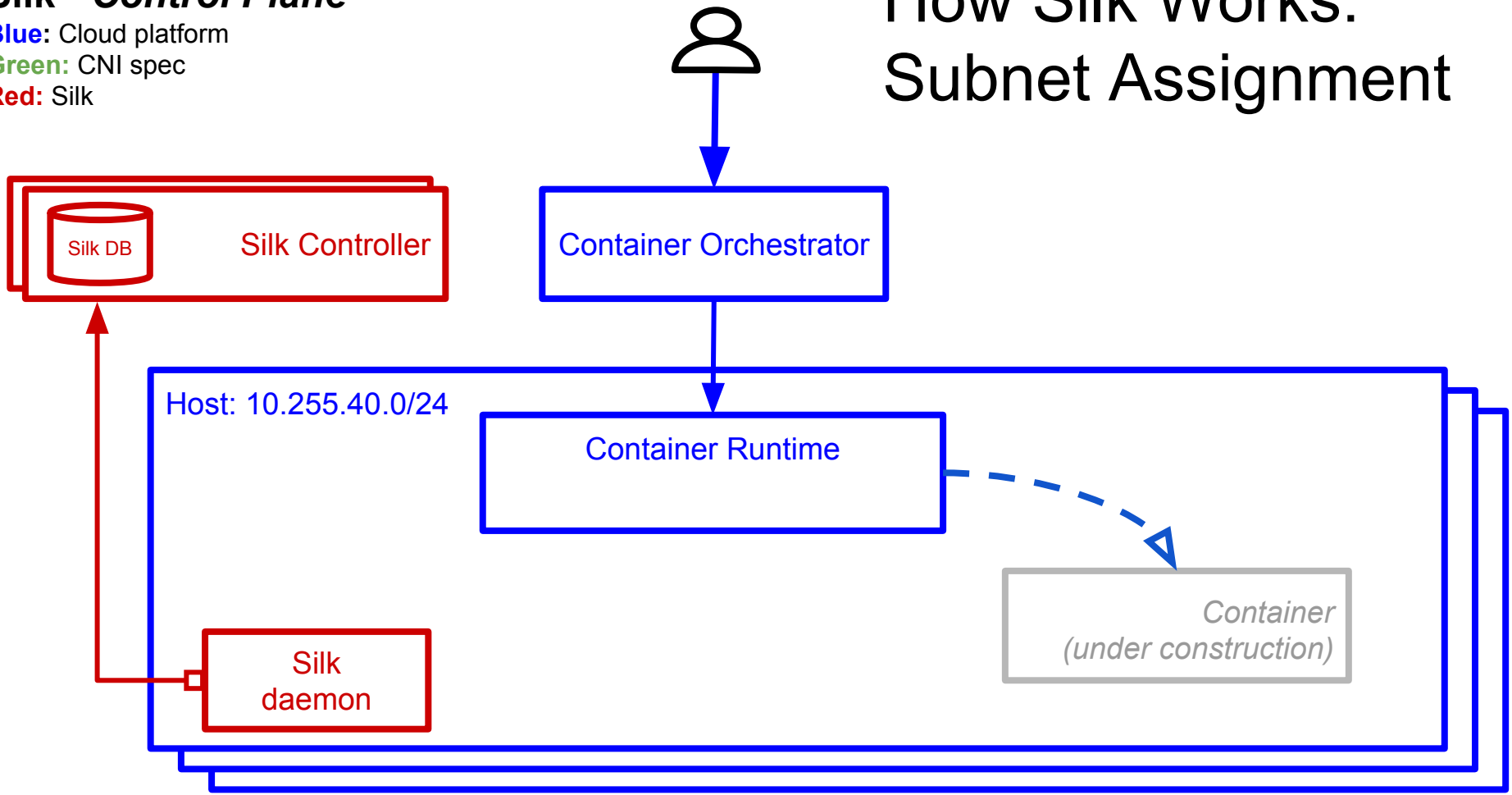
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

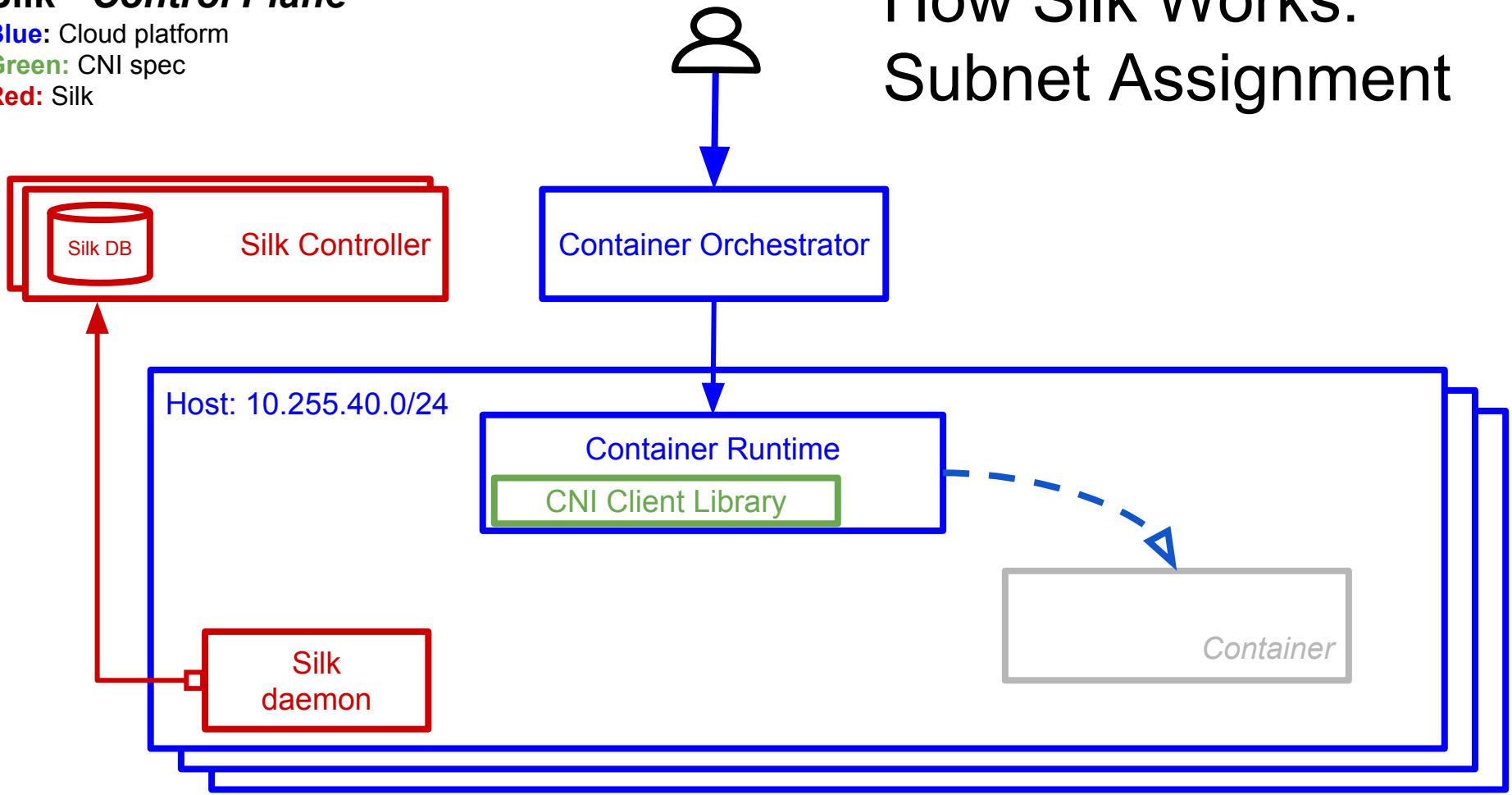
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

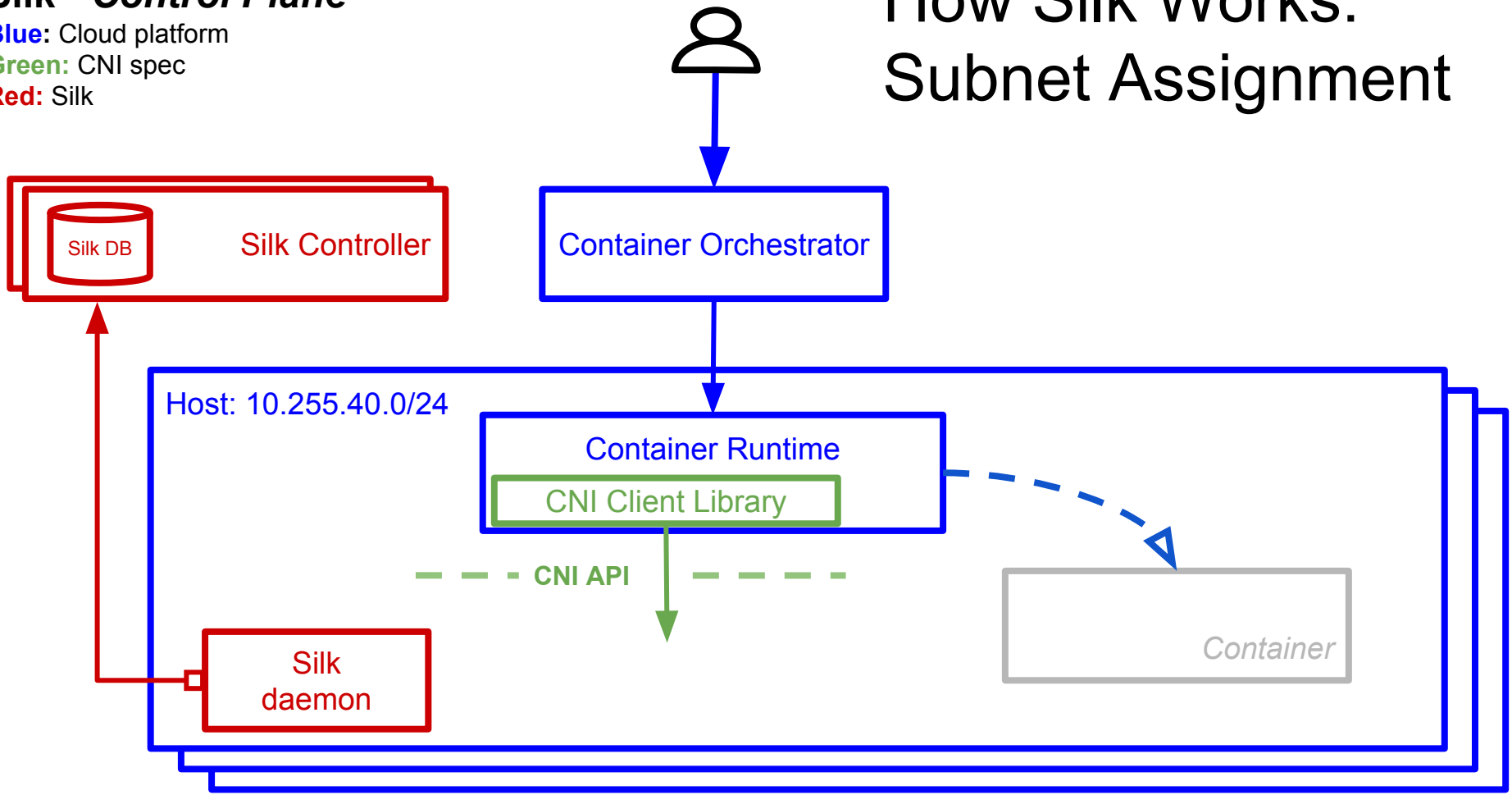
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

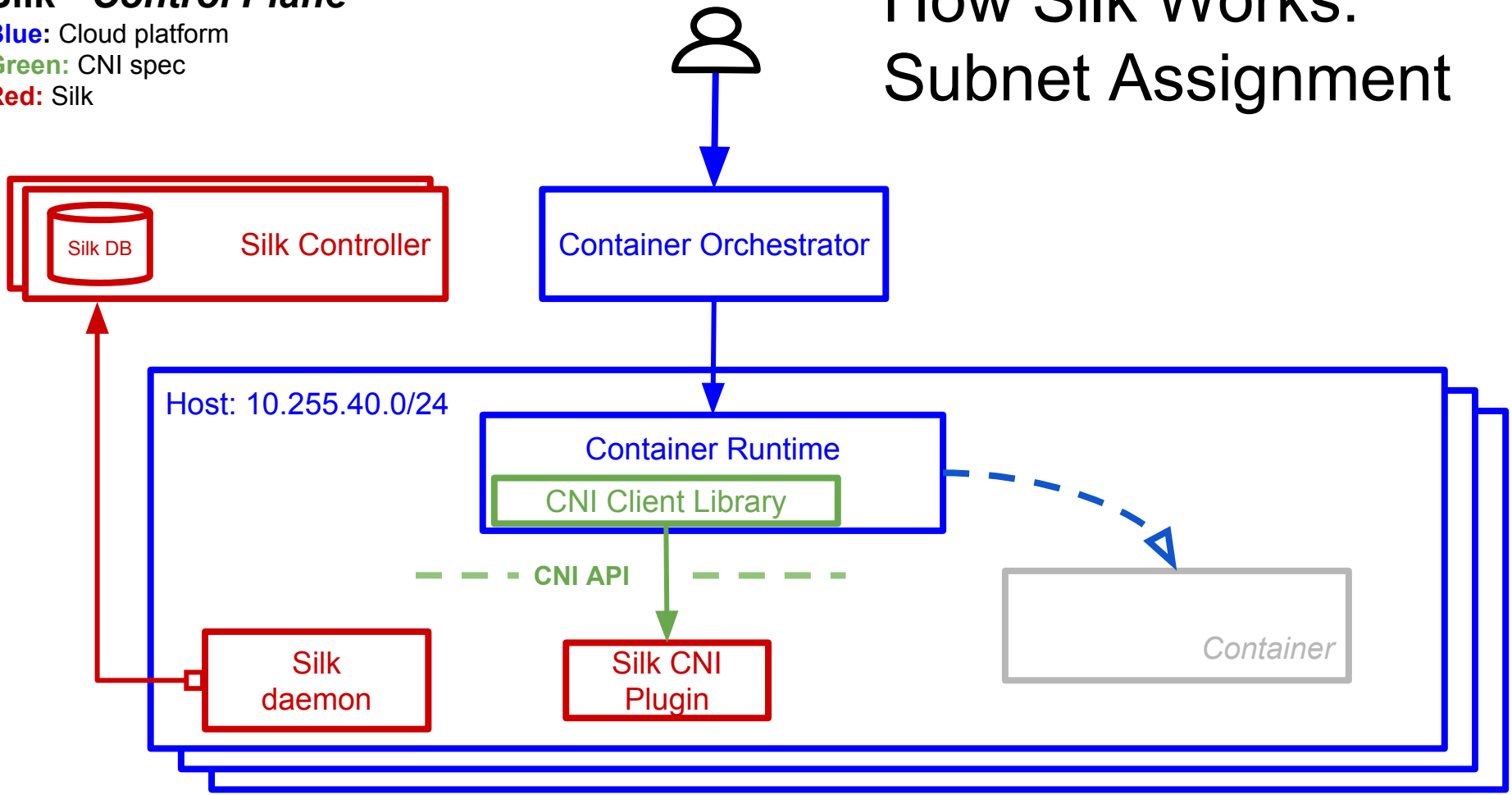
How Silk Works: Subnet Assignment



How Silk Works: Subnet Assignment

Silk - *Control Plane*

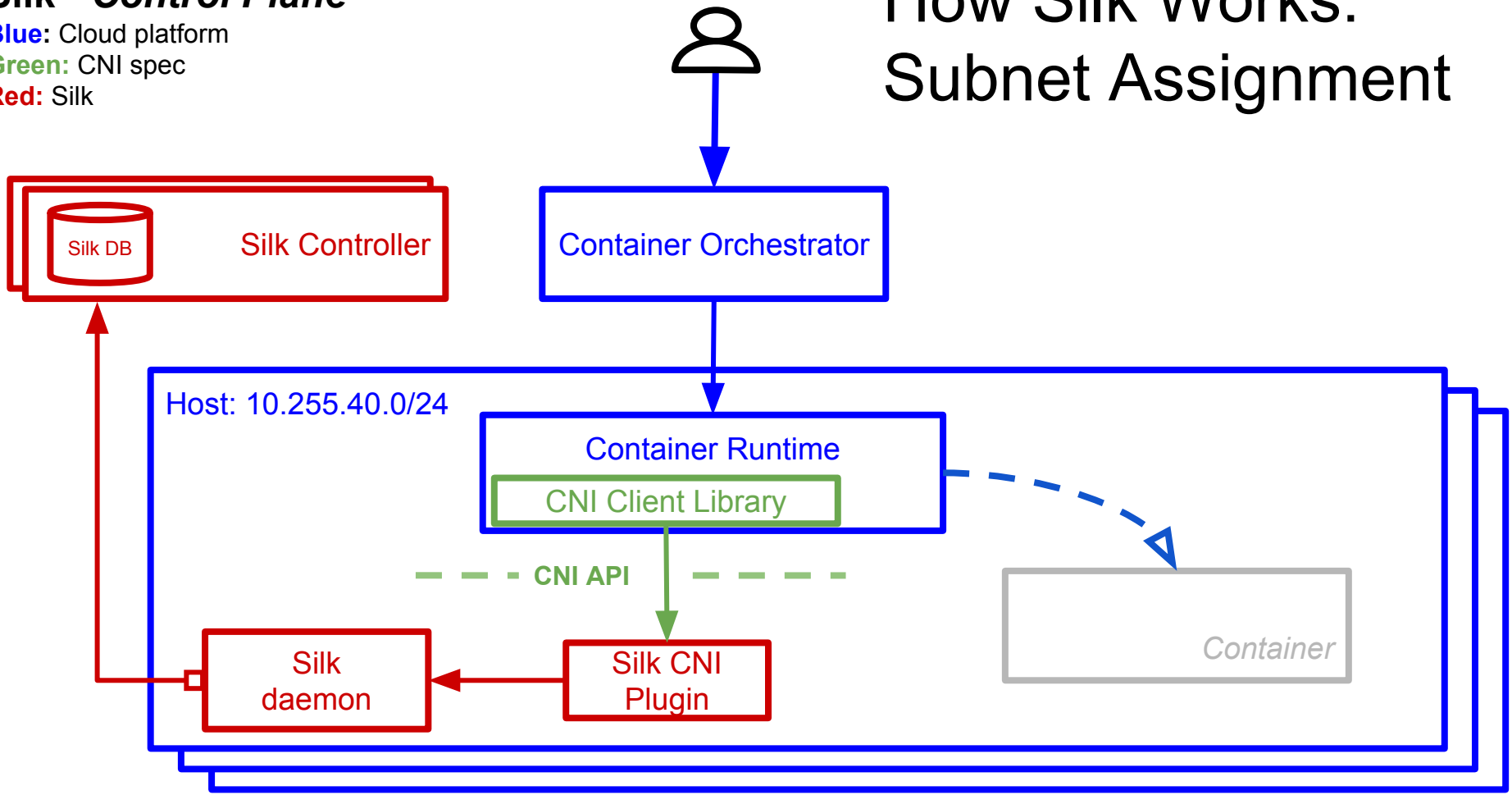
Blue: Cloud platform
Green: CNI spec
Red: Silk



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

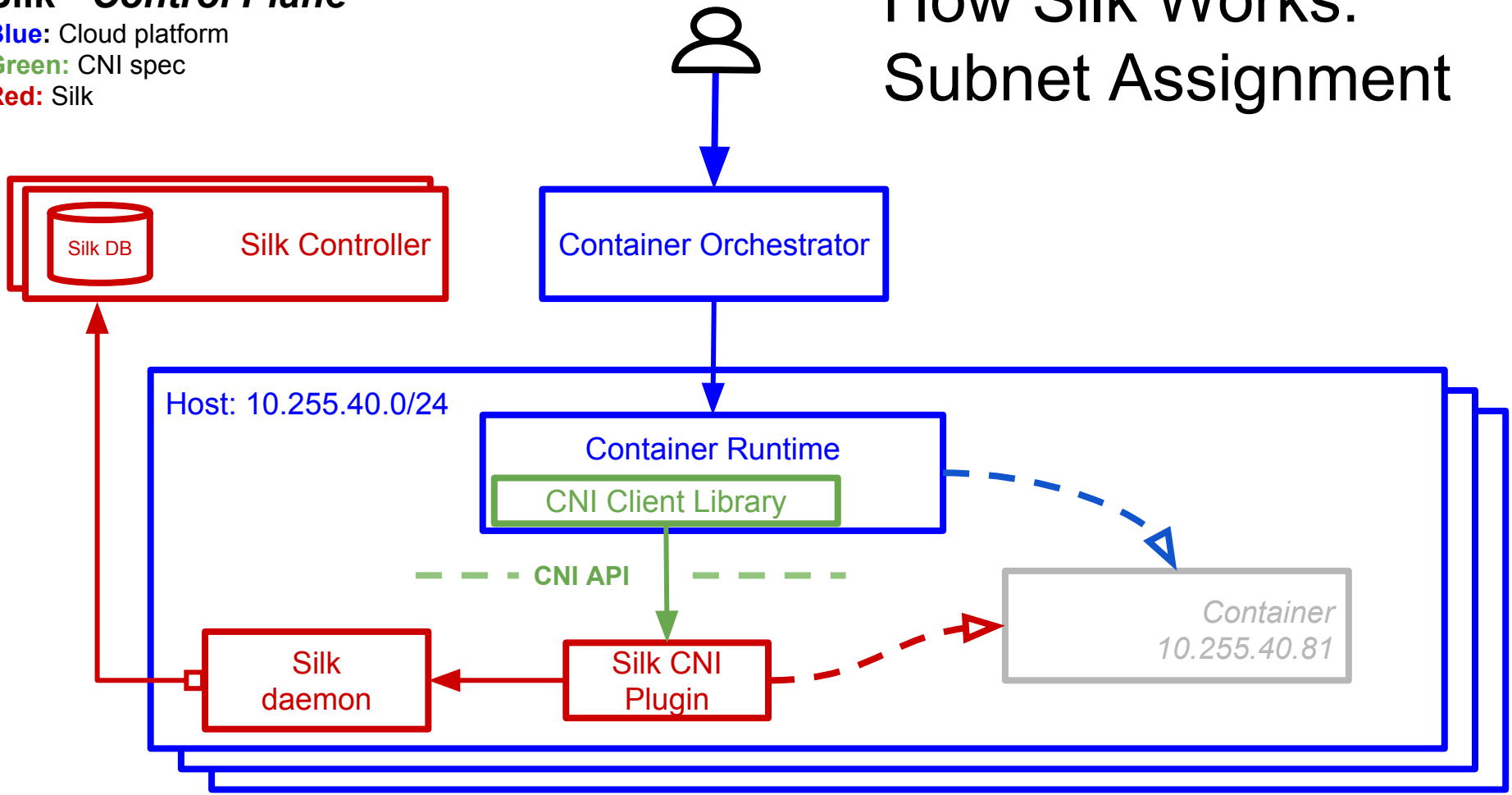
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

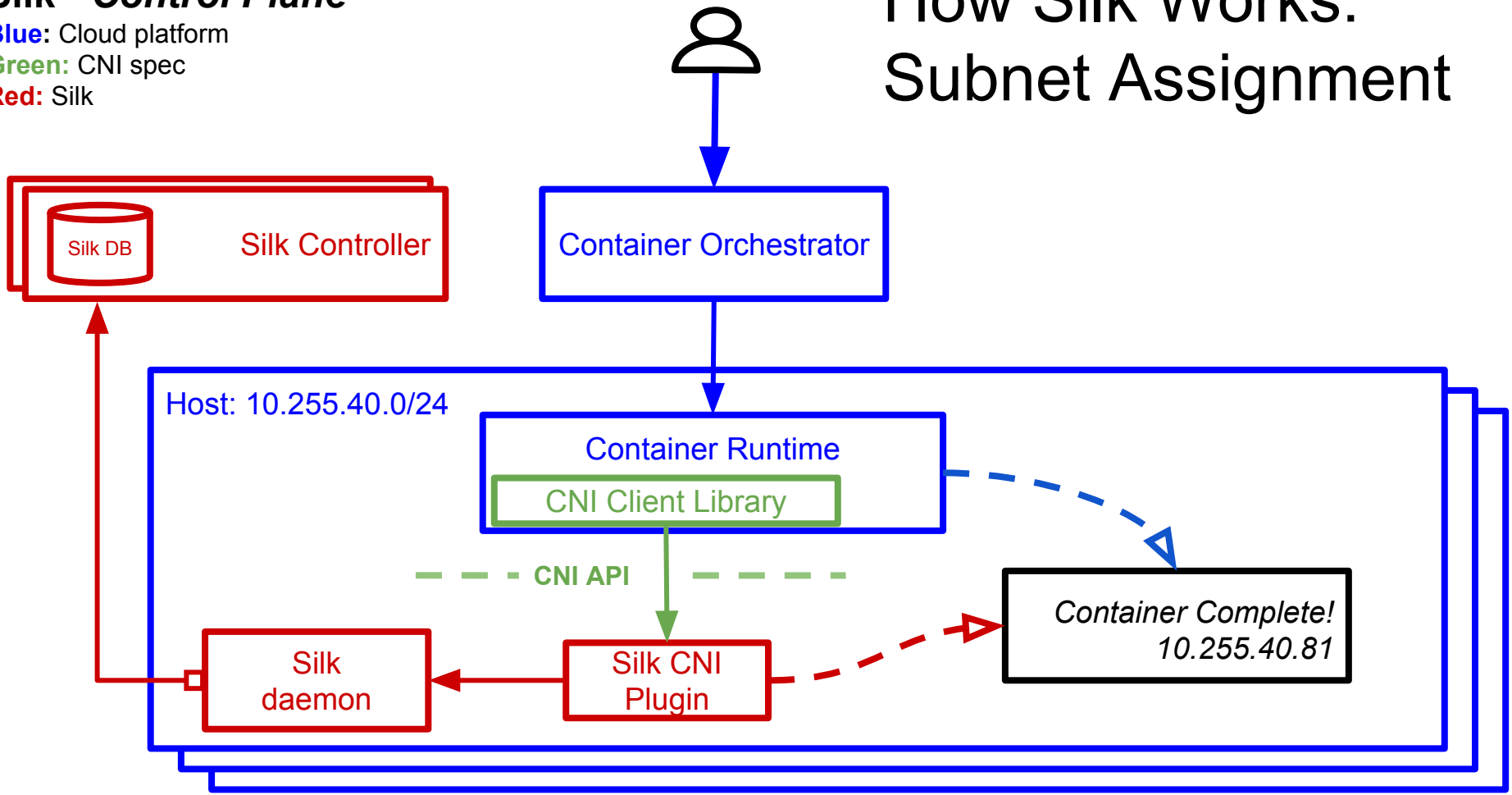
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

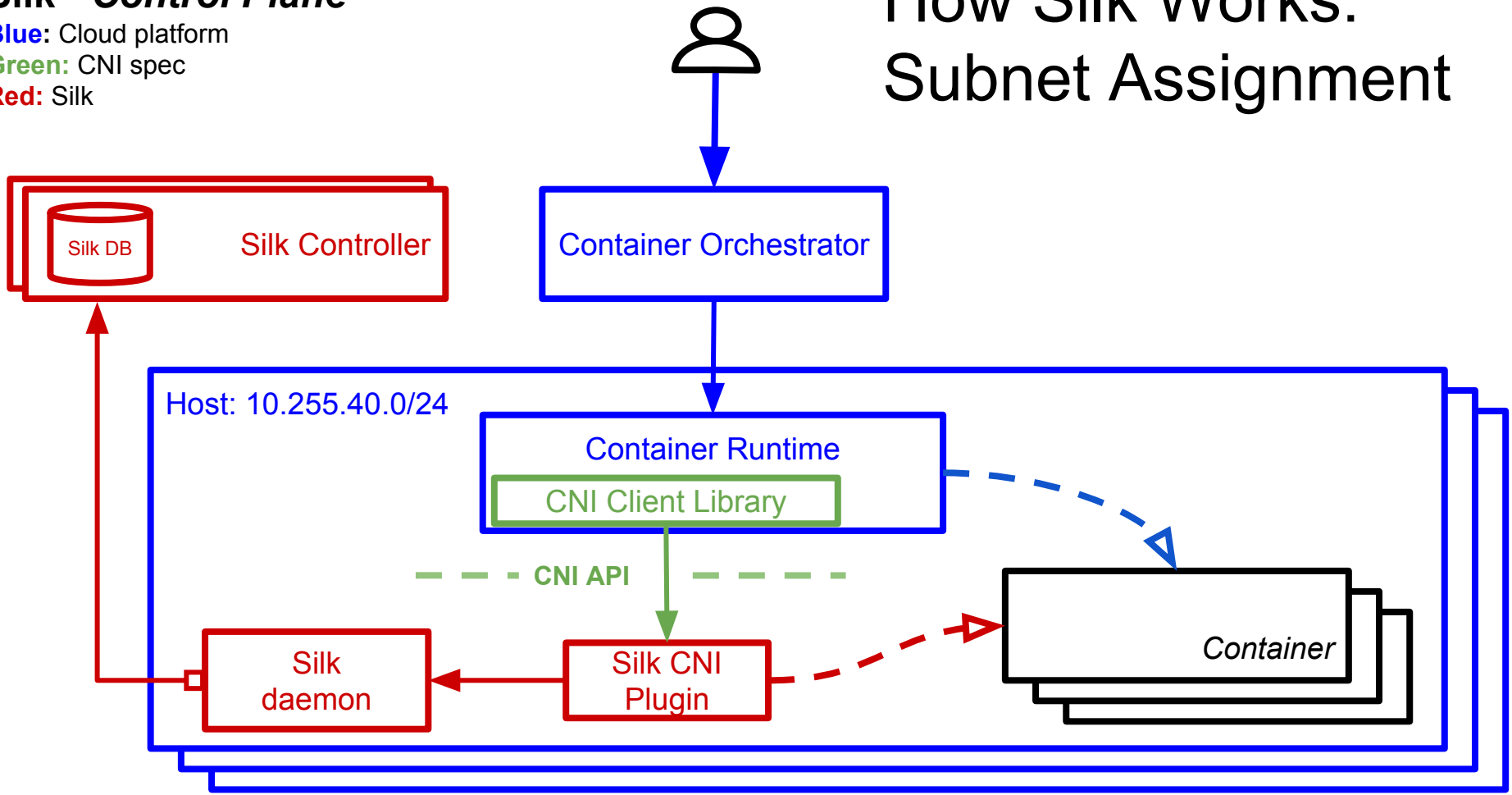
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

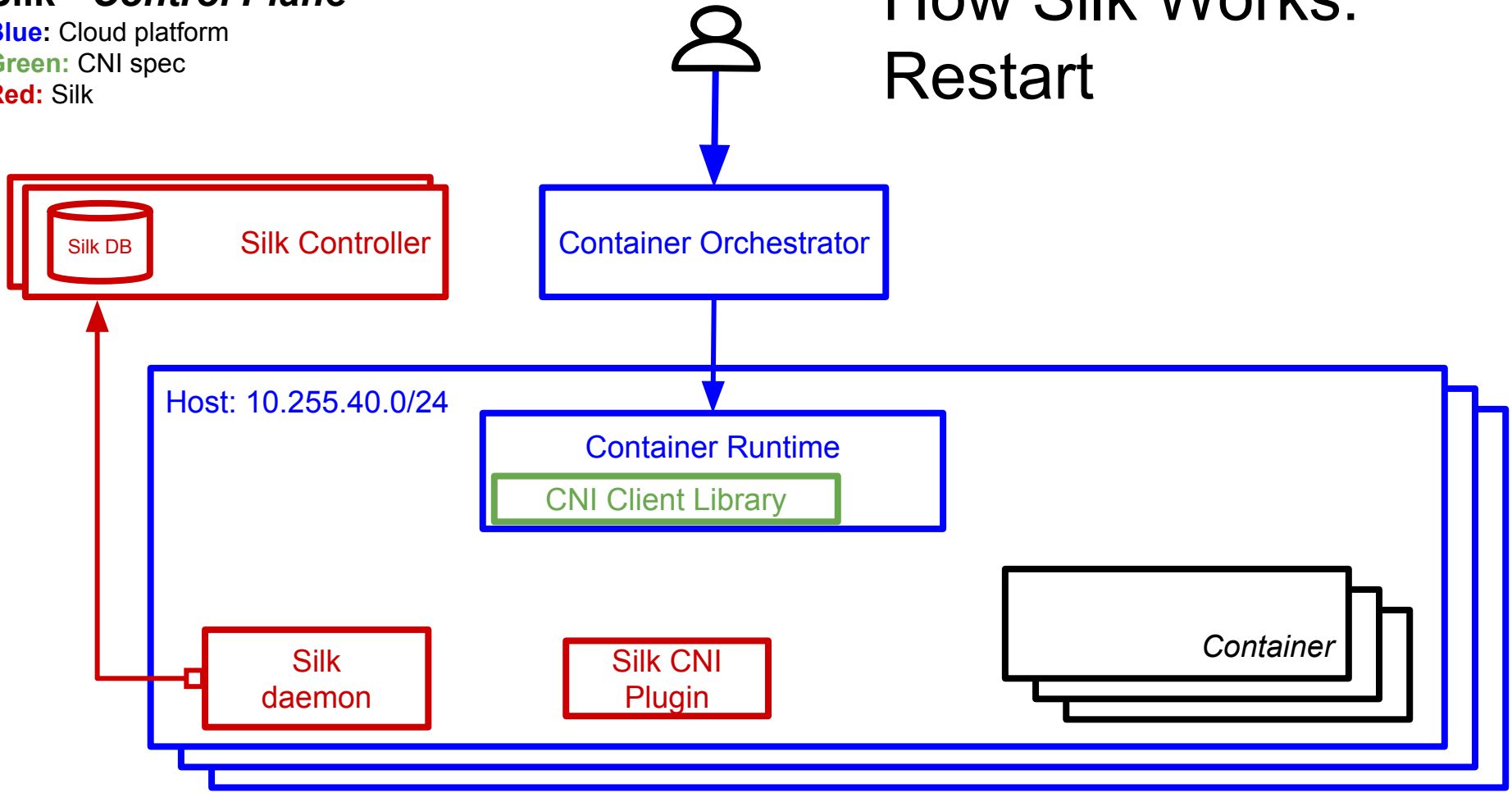
How Silk Works: Subnet Assignment



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

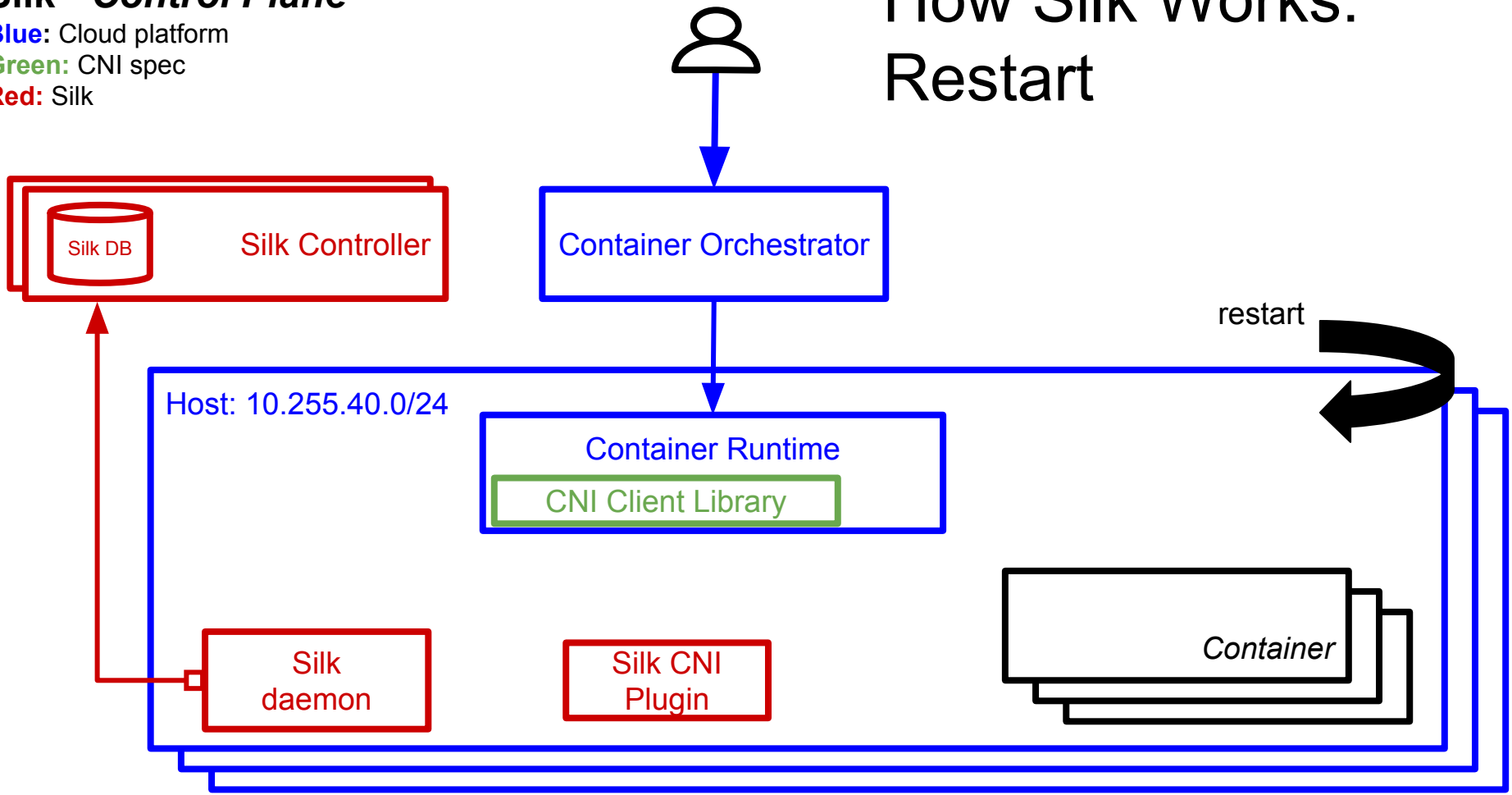
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

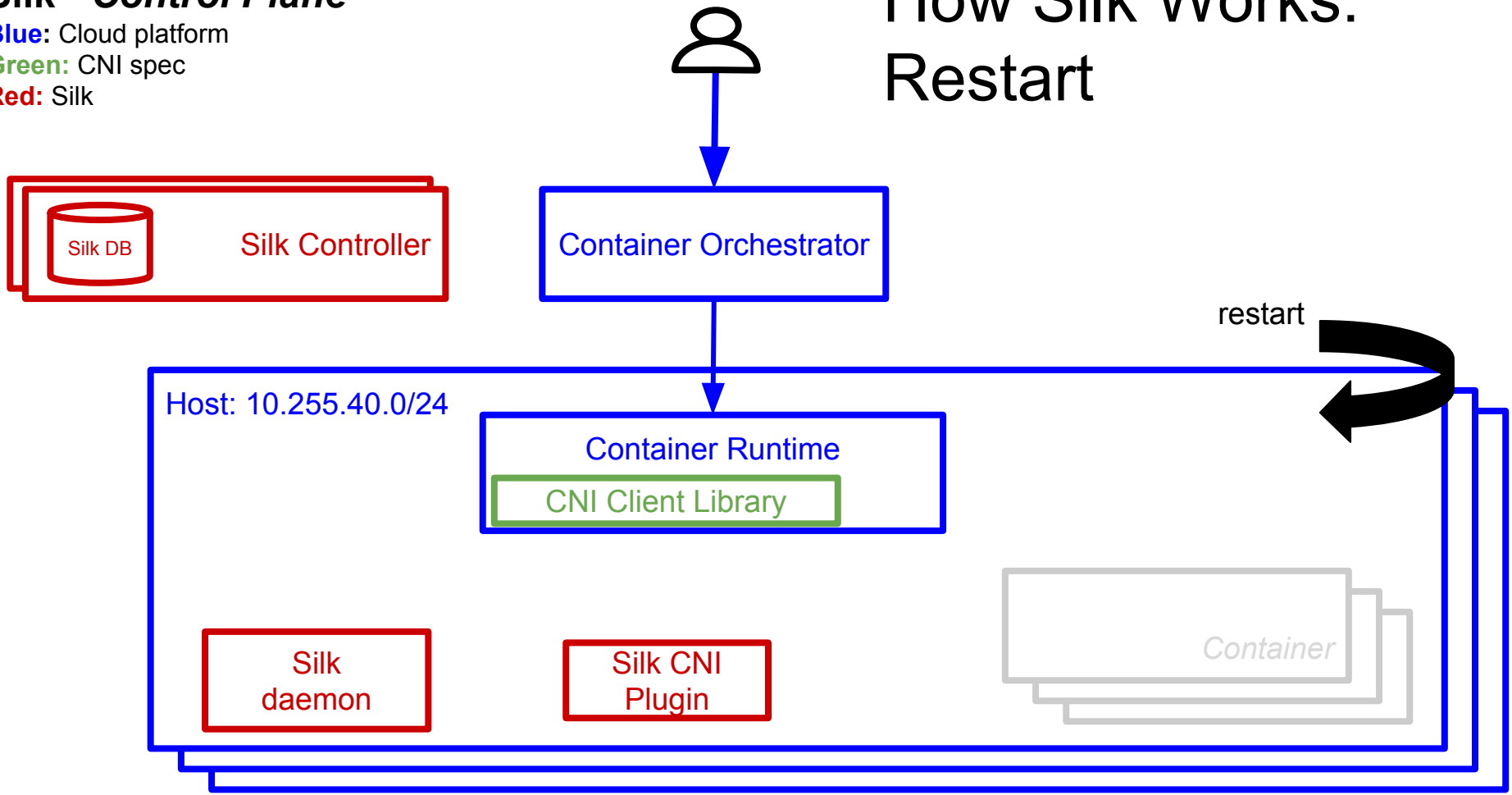
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

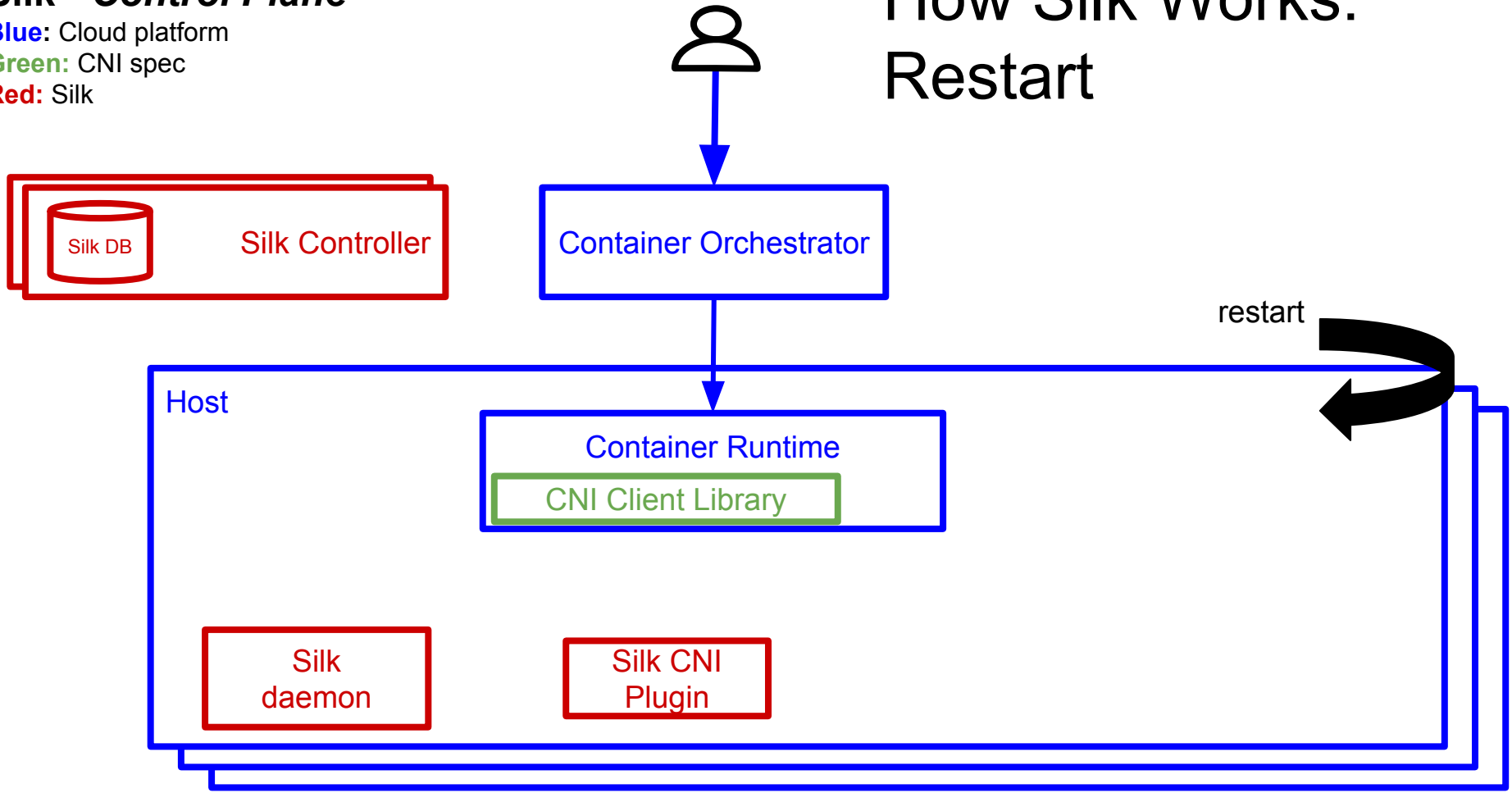
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

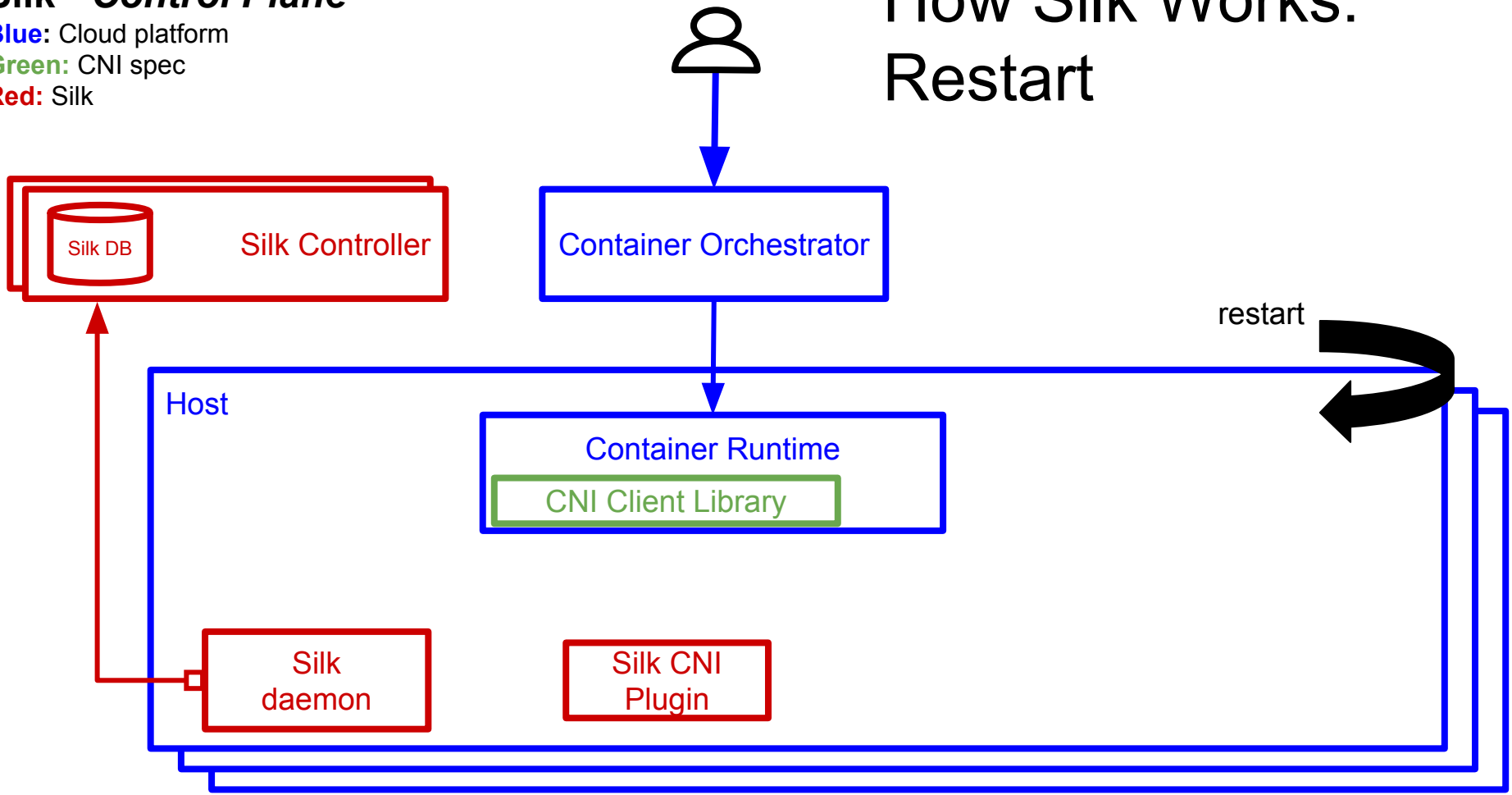
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

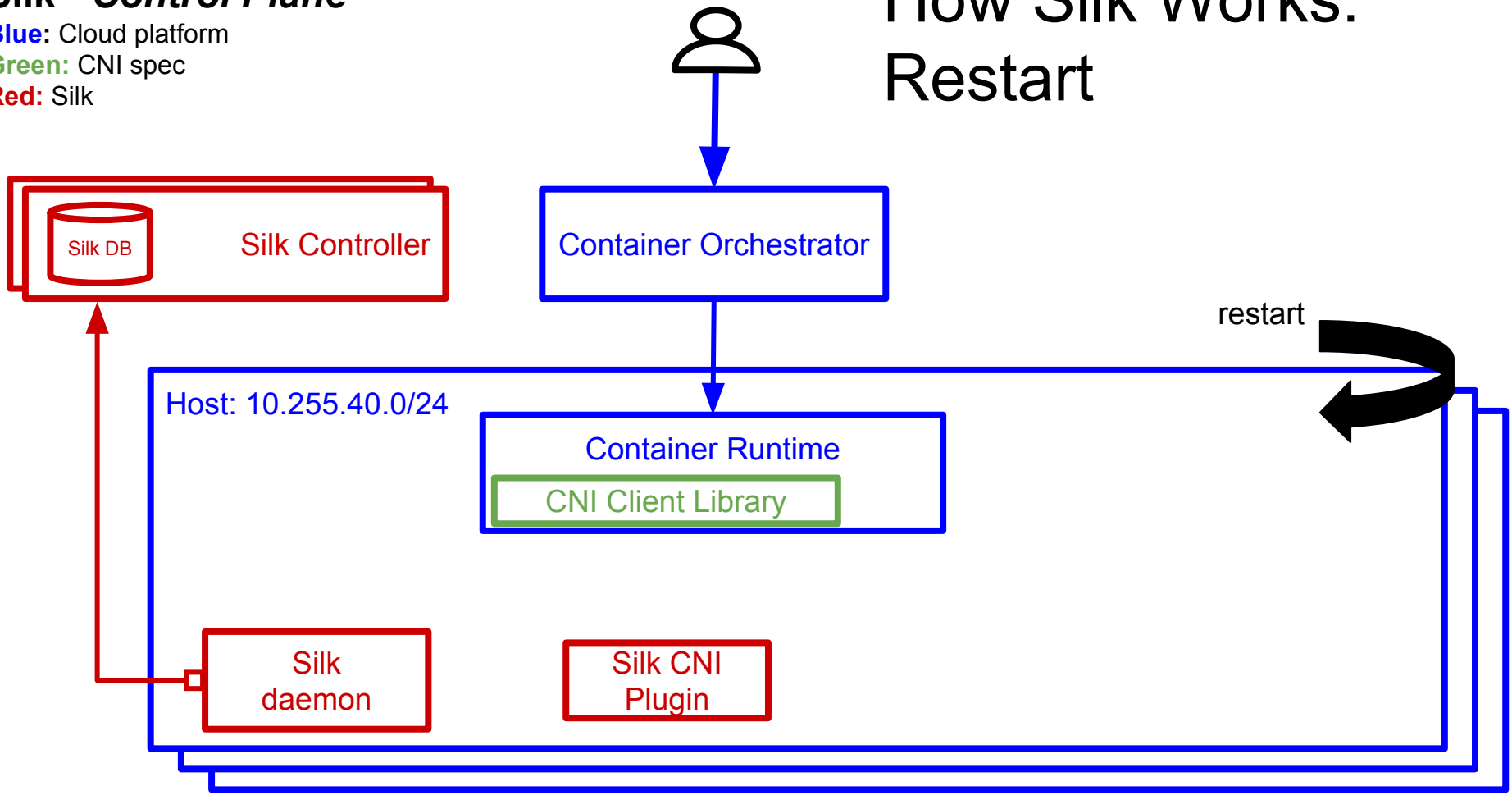
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

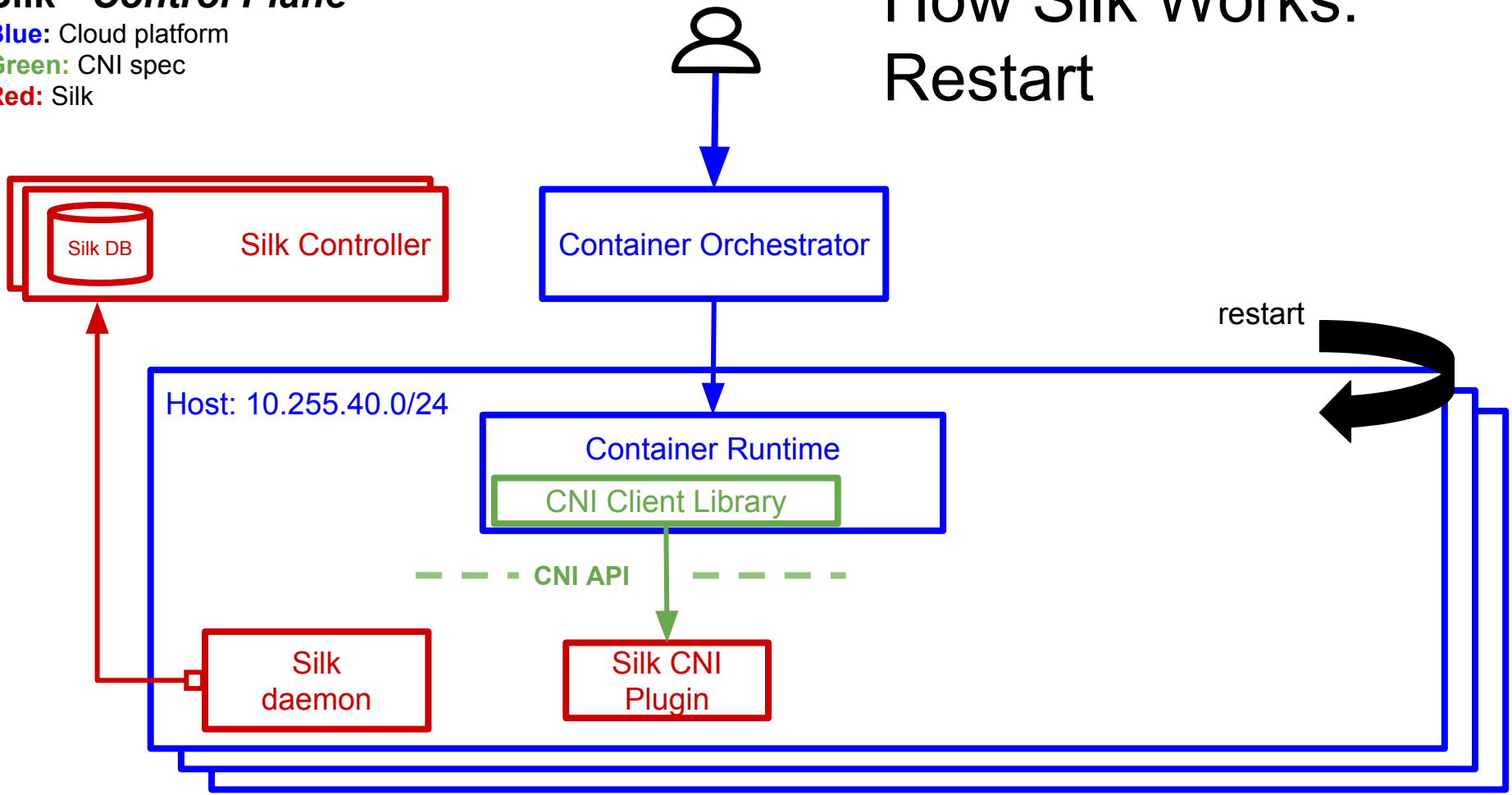
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

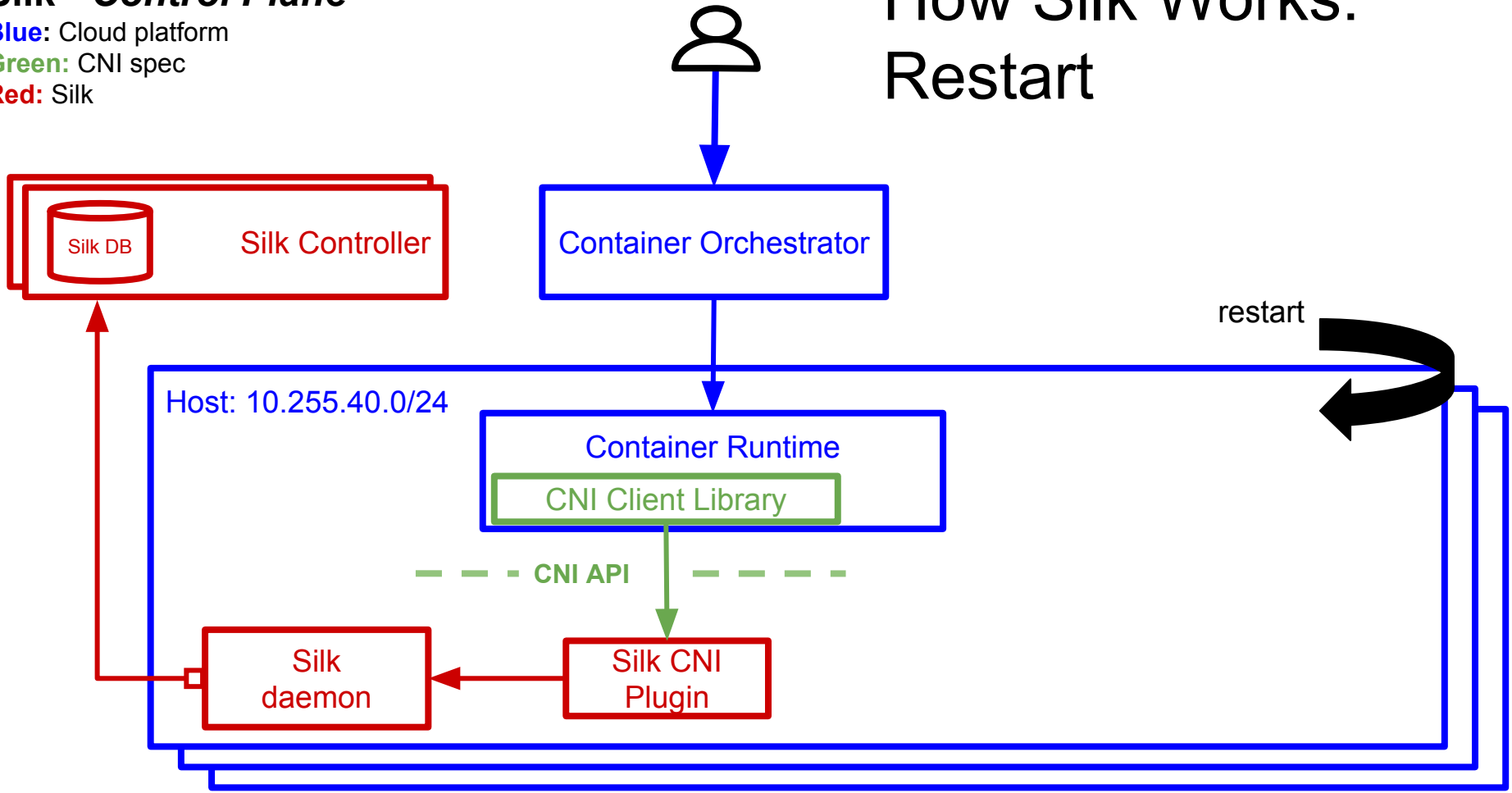
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

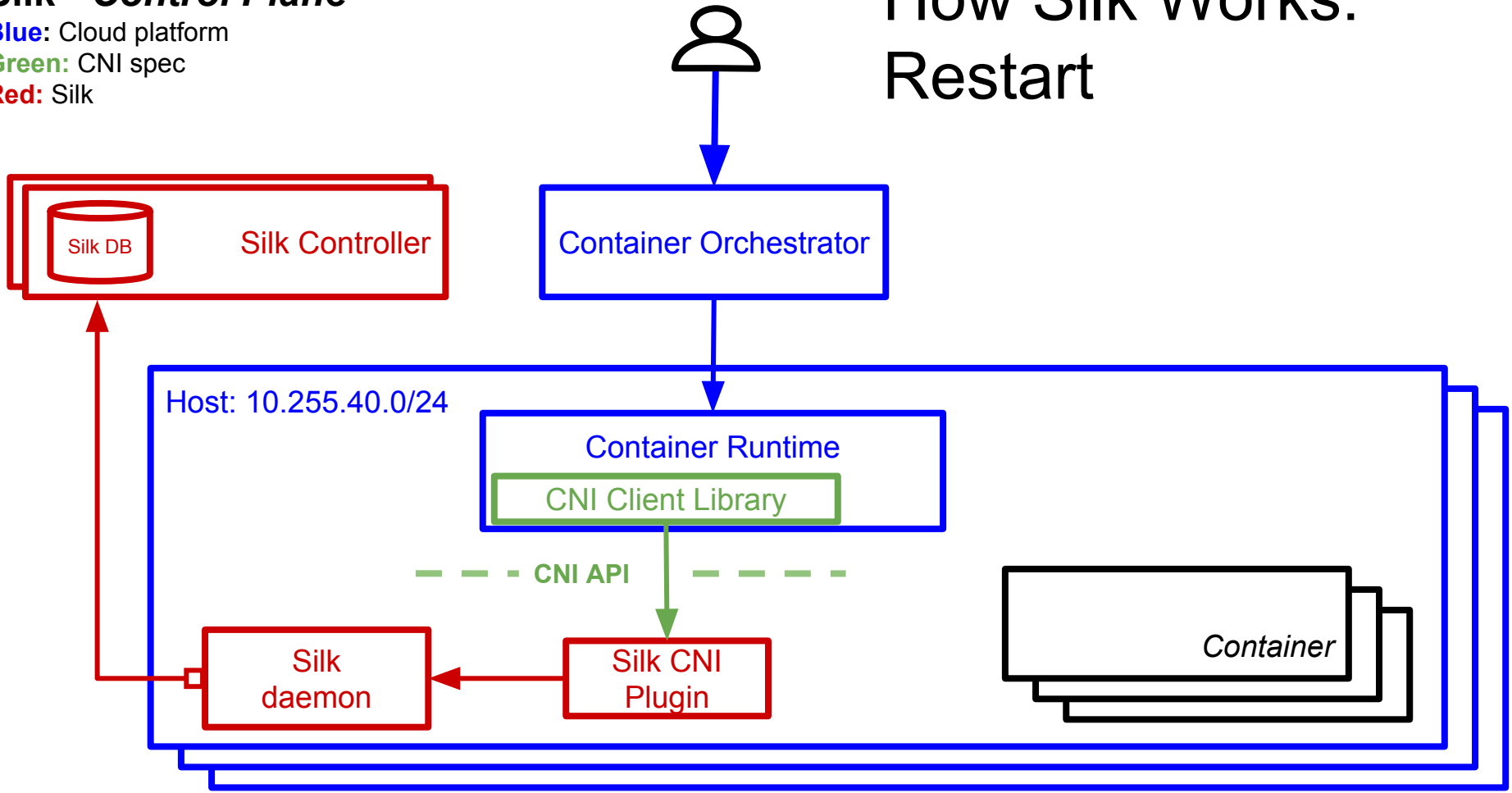
How Silk Works: Restart



Silk - *Control Plane*

Blue: Cloud platform
Green: CNI spec
Red: Silk

How Silk Works: Restart



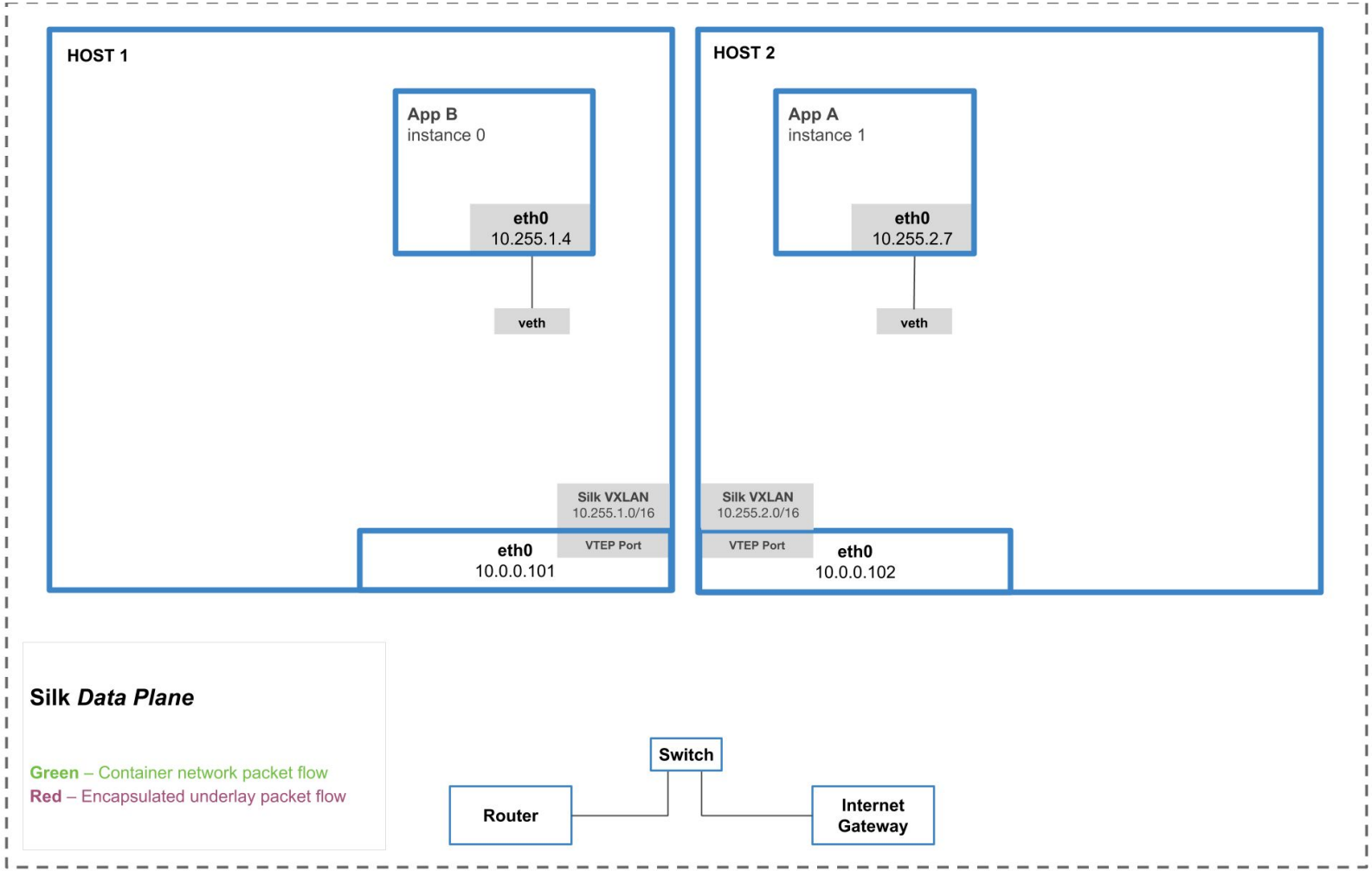
Reliability & Stability

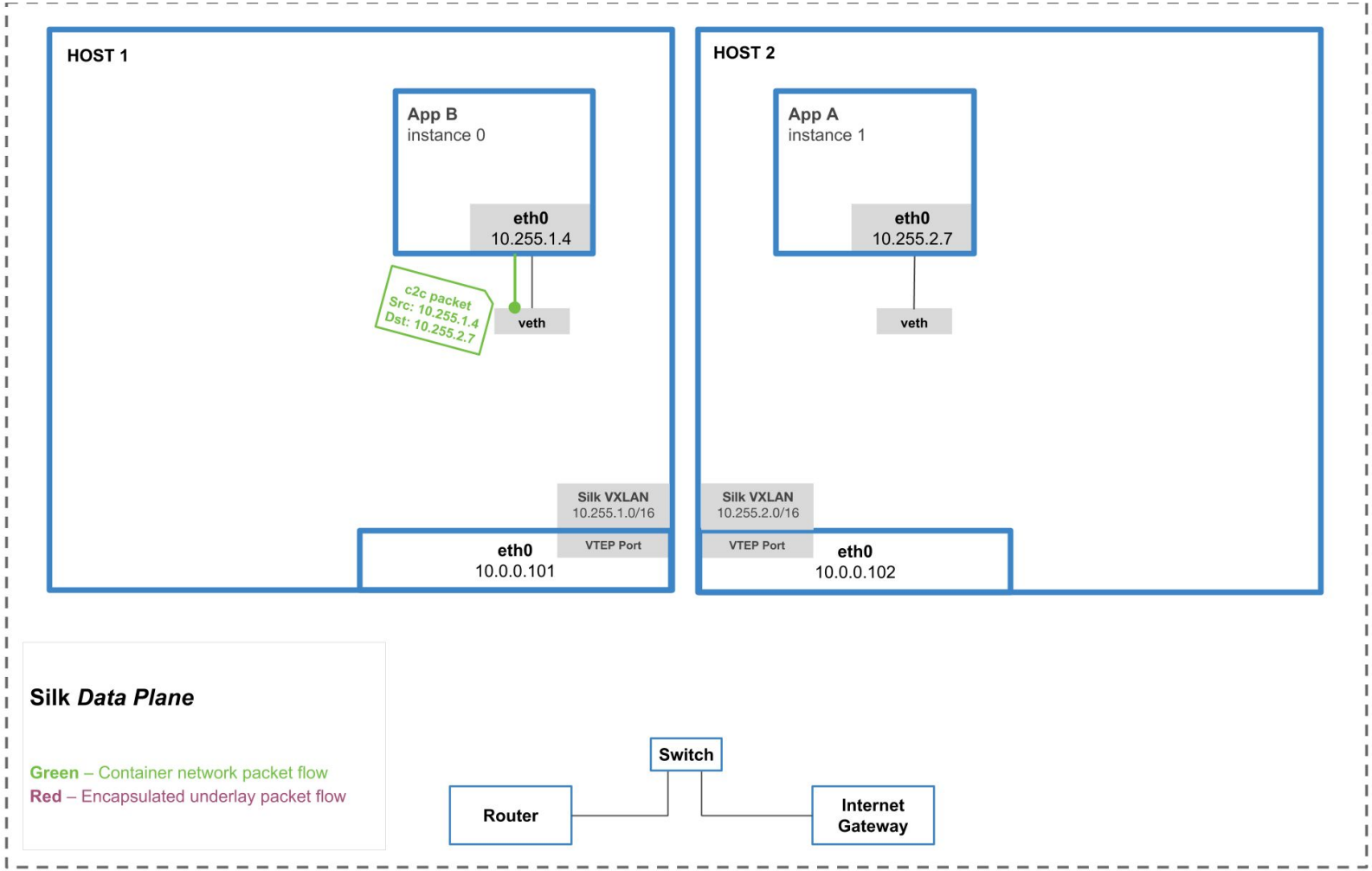
A lease is a subnet given to a host vm.

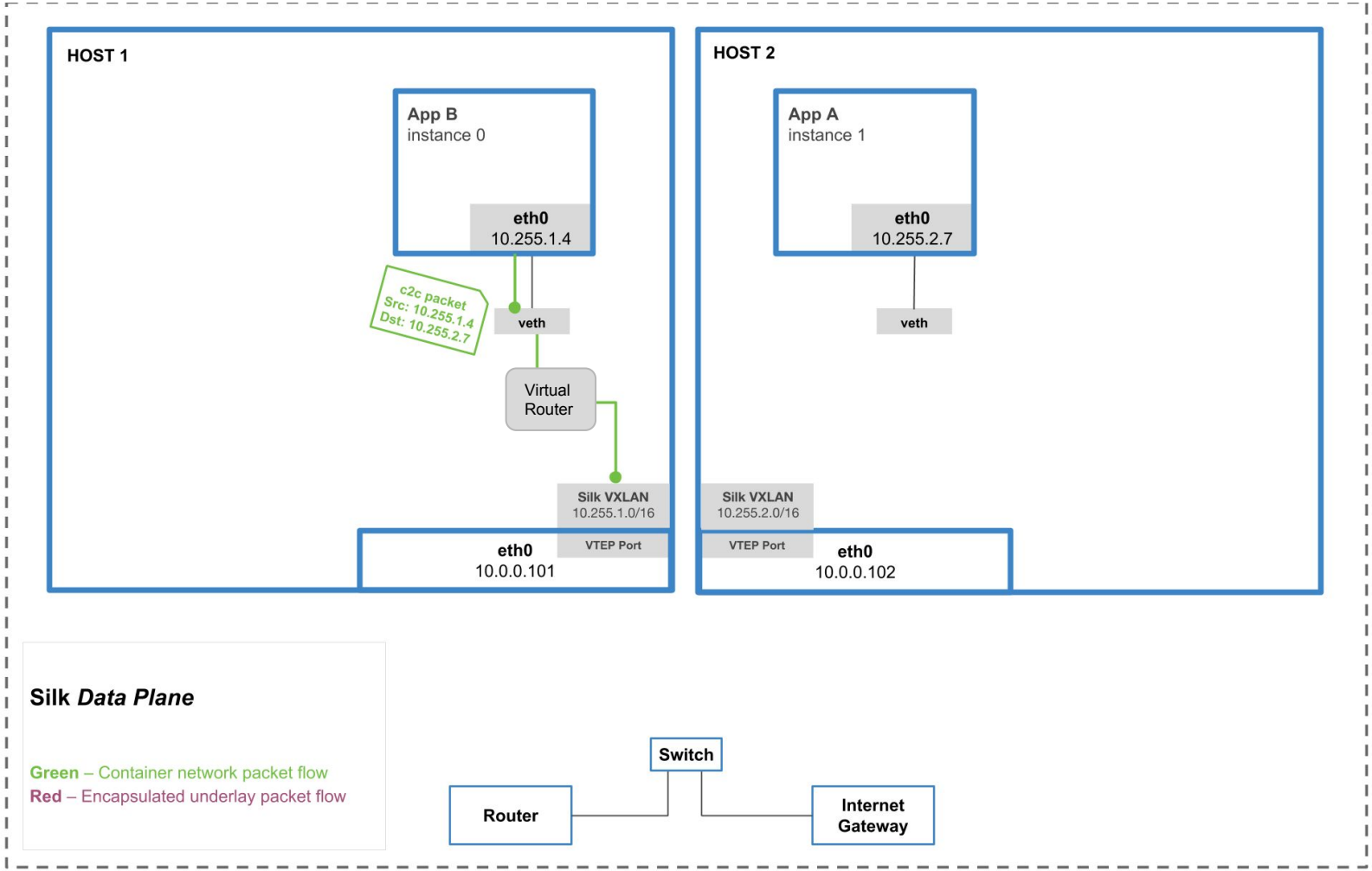
- Lease is given only if no containers on cell
- Network add will fail if daemon does not have a lease
- Lease is kept for as long as possible
- Add ability for teardown that can be utilized by CF

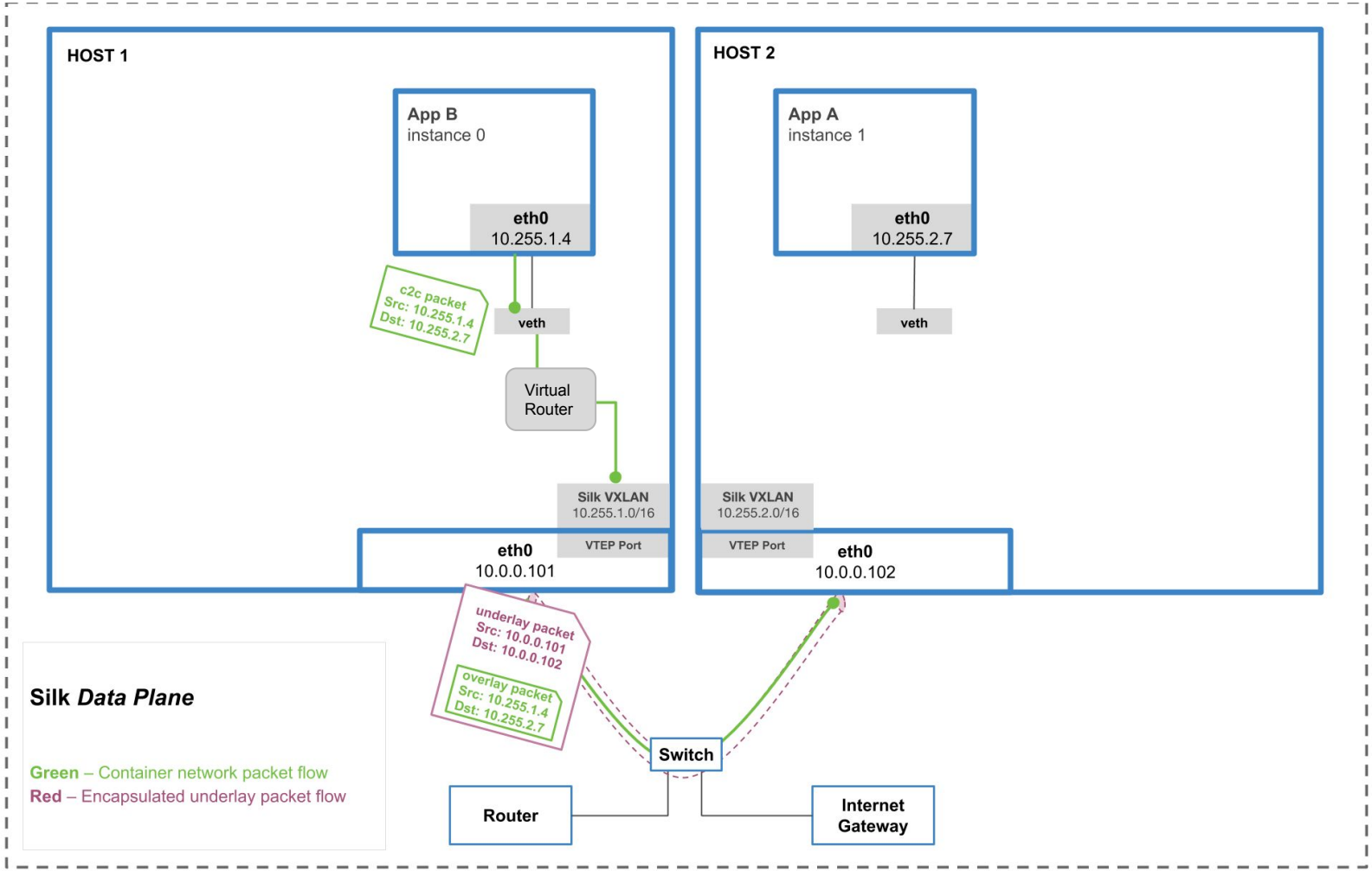
Network Setup

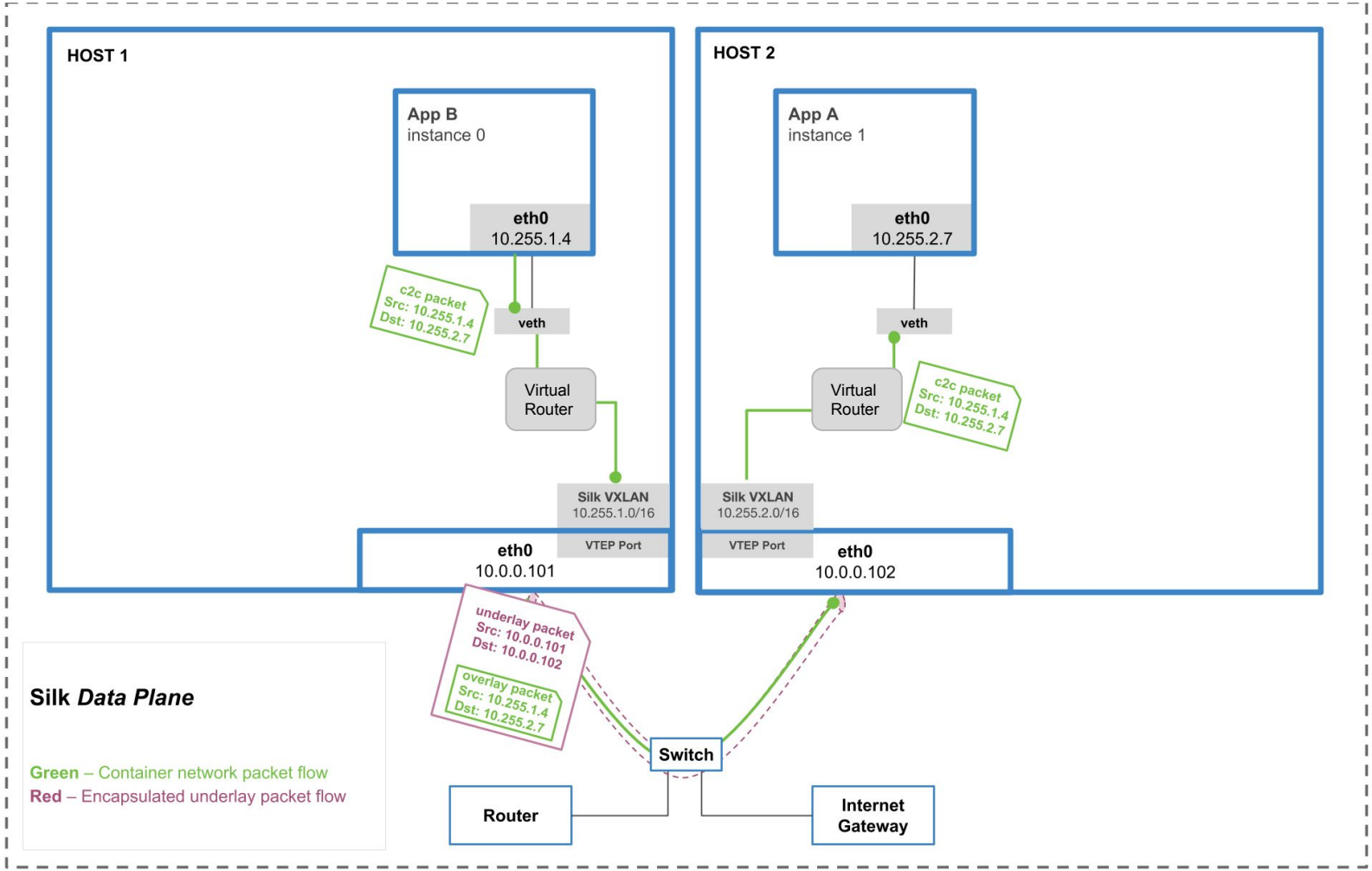
- IAAS agnostic: use VXLAN for encapsulation

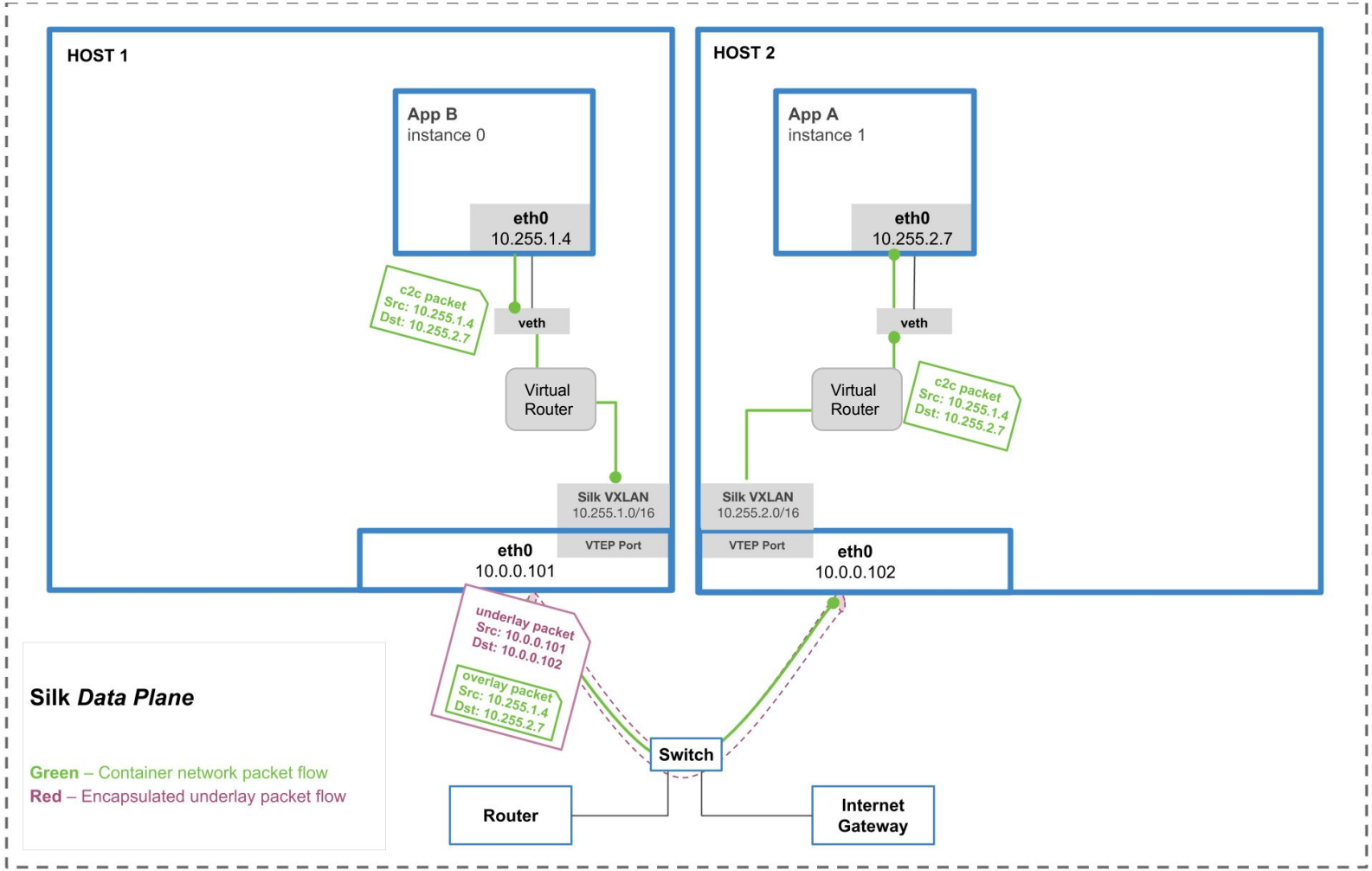

HOST 1
App B
instance 0
eth0
10.255.1.4
veth
Silk VXLAN
10.255.1.0/16
VTEP Port
eth0
10.0.0.101
HOST 2
App A
instance 1
eth0
10.255.2.7
veth
Silk VXLAN
10.255.2.0/16
VTEP Port
eth0
10.0.0.102
Silk *Data Plane*
Green – Container network packet flow
Red – Encapsulated underlay packet flow
Switch
Router
Internet Gateway


HOST 1
App B
instance 0
eth0
10.255.1.4
c2c packet
Src: 10.255.1.4
Dst: 10.255.2.7
veth
Silk VXLAN
10.255.1.0/16
VTEP Port
eth0
10.0.0.101
HOST 2
App A
instance 1
eth0
10.255.2.7
veth
Silk VXLAN
10.255.2.0/16
VTEP Port
eth0
10.0.0.102
Silk Data Plane
Green – Container network packet flow
Red – Encapsulated underlay packet flow
Switch
Router
Internet Gateway


HOST 1
App B
instance 0
eth0
10.255.1.4
c2c packet
Src: 10.255.1.4
Dst: 10.255.2.7
veth
Virtual
Router
Silk VXLAN
10.255.1.0/16
VTEP Port
eth0
10.0.0.101
HOST 2
App A
instance 1
eth0
10.255.2.7
veth
Silk VXLAN
10.255.2.0/16
VTEP Port
eth0
10.0.0.102
Silk Data Plane
Green – Container network packet flow
Red – Encapsulated underlay packet flow
Switch
Router
Internet
Gateway


HOST 1
App B
instance 0
eth0
10.255.1.4
c2c packet
Src: 10.255.1.4
Dst: 10.255.2.7
veth
Virtual
Router
Silk VXLAN
10.255.1.0/16
VTEP Port
eth0
10.0.0.101
HOST 2
App A
instance 1
eth0
10.255.2.7
veth
Silk VXLAN
10.255.2.0/16
VTEP Port
eth0
10.0.0.102
underlay packet
Src: 10.0.0.101
Dst: 10.0.0.102
overlay packet
Src: 10.255.1.4
Dst: 10.255.2.7
Switch
Router
Internet
Gateway
Silk Data Plane
Green – Container network packet flow
Red – Encapsulated underlay packet flow


HOST 1
App B
instance 0
eth0
10.255.1.4
c2c packet
Src: 10.255.1.4
Dst: 10.255.2.7
veth
Virtual
Router
Silk VXLAN
10.255.1.0/16
VTEP Port
eth0
10.0.0.101
HOST 2
App A
instance 1
eth0
10.255.2.7
veth
Virtual
Router
c2c packet
Src: 10.255.1.4
Dst: 10.255.2.7
Silk VXLAN
10.255.2.0/16
VTEP Port
eth0
10.0.0.102
underlay packet
Src: 10.0.0.101
Dst: 10.0.0.102
overlay packet
Src: 10.255.1.4
Dst: 10.255.2.7
Switch
Router
Internet
Gateway
Silk Data Plane
Green – Container network packet flow
Red – Encapsulated underlay packet flow


HOST 1
App B
instance 0
eth0
10.255.1.4
c2c packet
Src: 10.255.1.4
Dst: 10.255.2.7
veth
Virtual
Router
Silk VXLAN
10.255.1.0/16
VTEP Port
eth0
10.0.0.101
HOST 2
App A
instance 1
eth0
10.255.2.7
veth
Virtual
Router
c2c packet
Src: 10.255.1.4
Dst: 10.255.2.7
Silk VXLAN
10.255.2.0/16
VTEP Port
eth0
10.0.0.102
underlay packet
Src: 10.0.0.101
Dst: 10.0.0.102
overlay packet
Src: 10.255.1.4
Dst: 10.255.2.7
Switch
Router
Internet
Gateway
Silk Data Plane
Green – Container network packet flow
Red – Encapsulated underlay packet flow

Network Setup

- IAAS agnostic: use VXLAN for encapsulation
- Single, flat L3 network from container perspective

Agenda

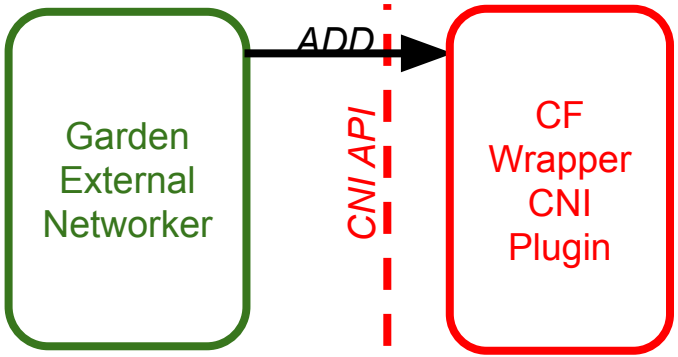
- What is CNI
- What is Cloud Foundry
- Cloud Foundry + CNI
- Main Motivations to Build Silk CNI
- Building Silk
- **What's Next?**
- Key Takeaways

The Future of Silk

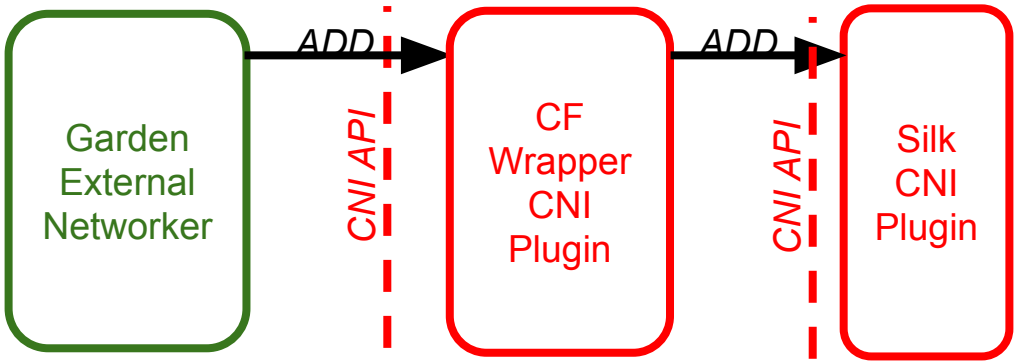
- CNI Chaining

CNI Chaining

CNI Chaining

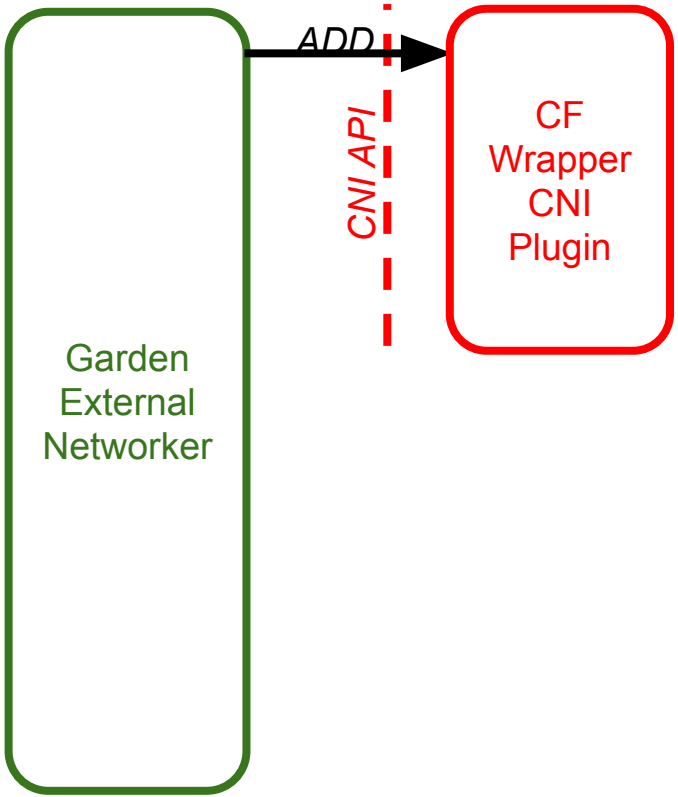


CNI Chaining

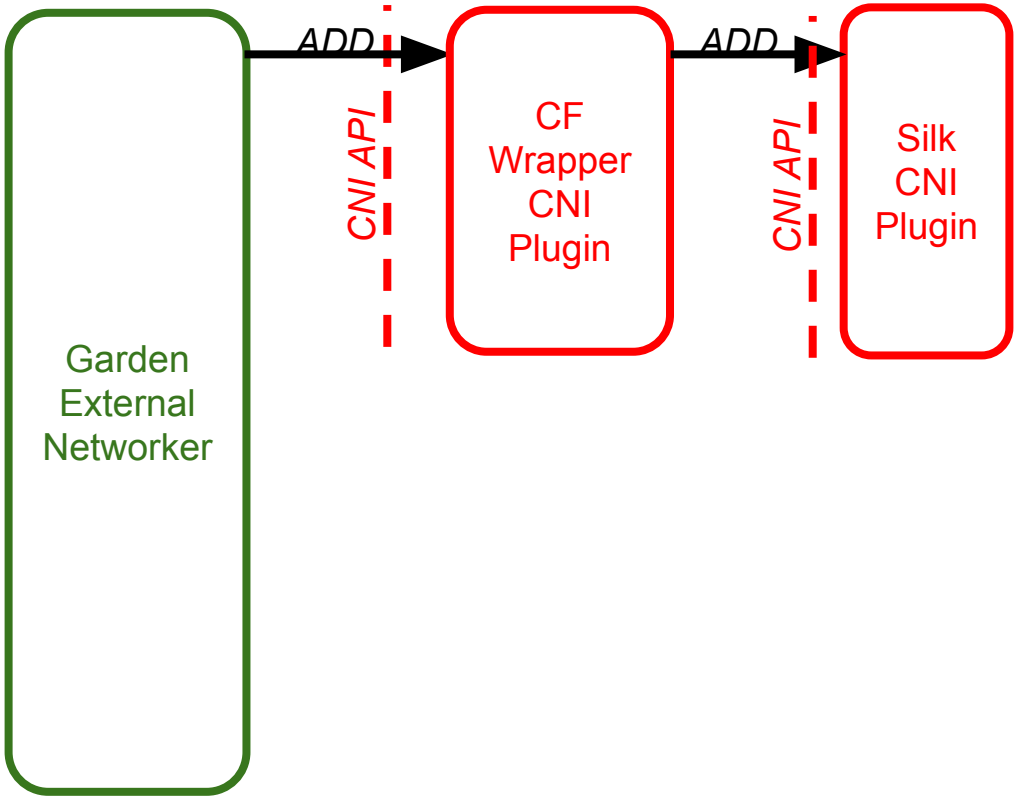


CNI Chaining

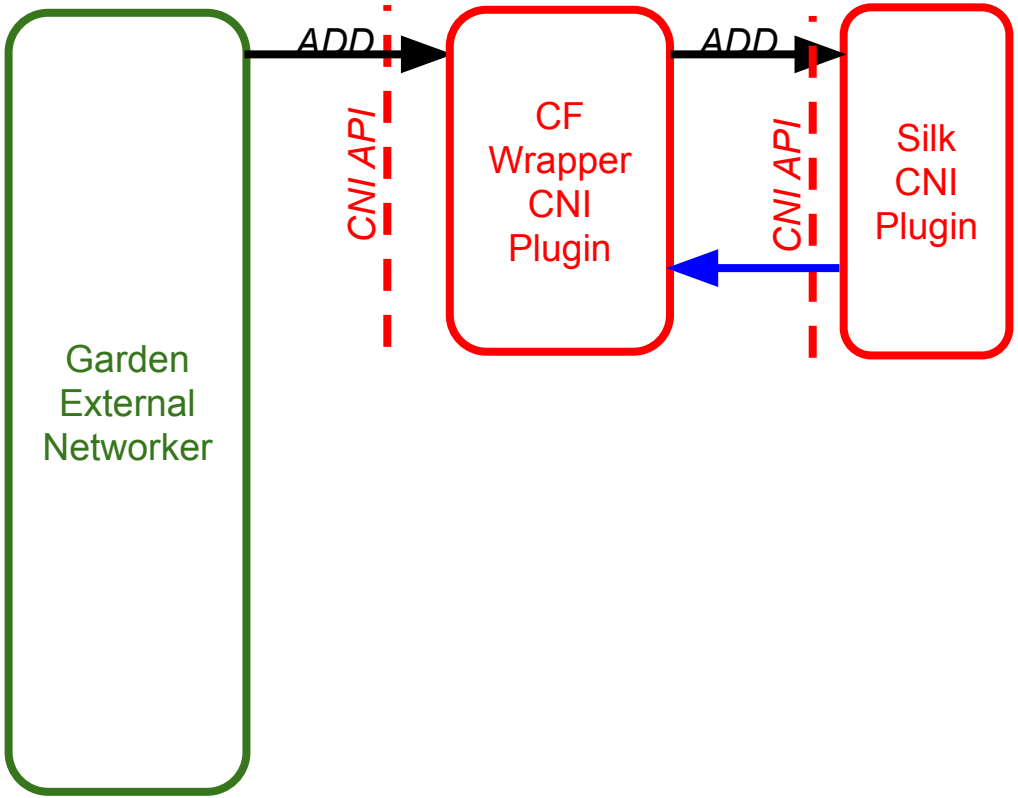
CNI Chaining



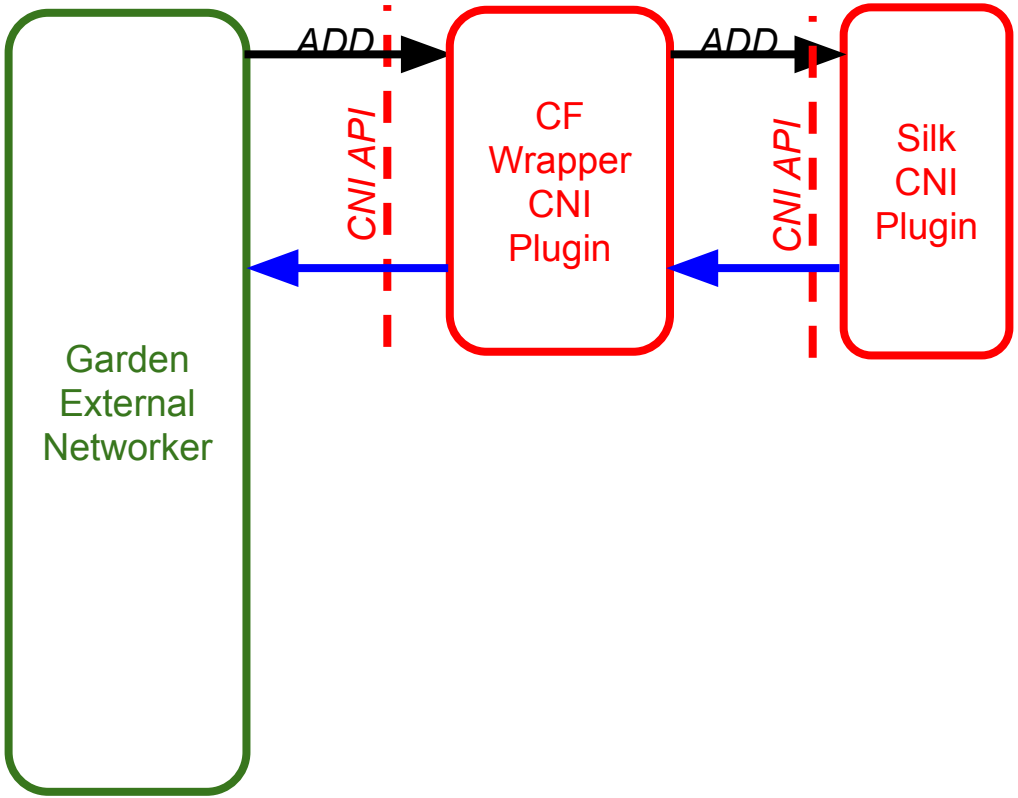
CNI Chaining



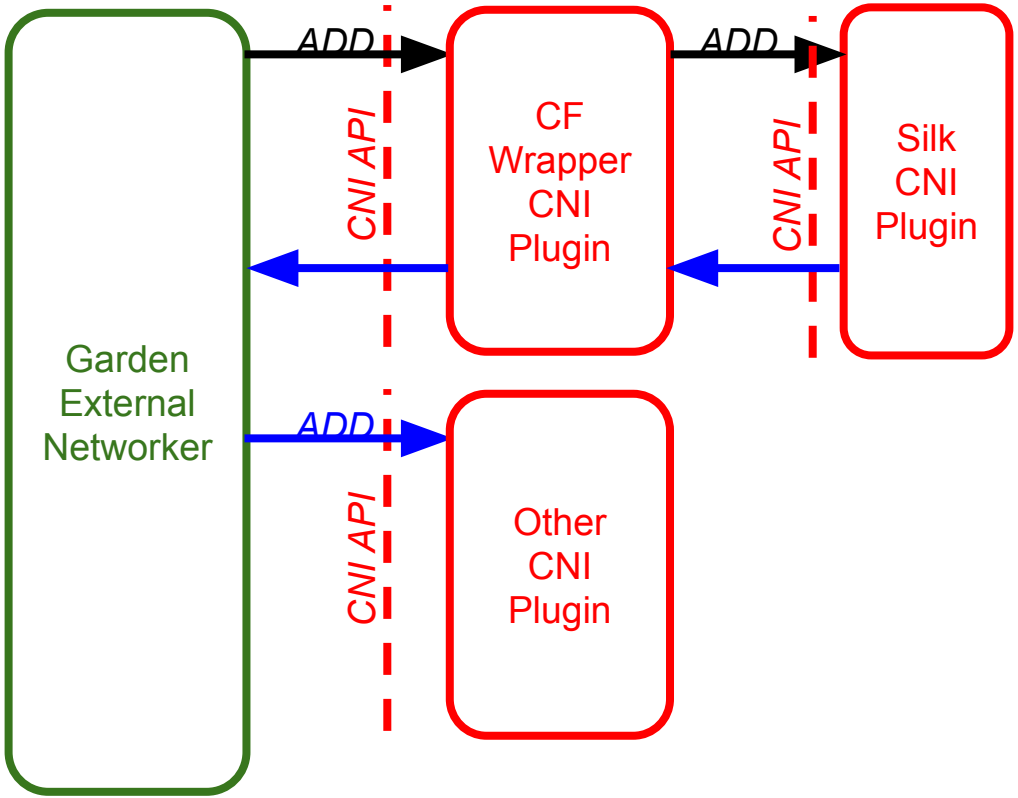
CNI Chaining



CNI Chaining

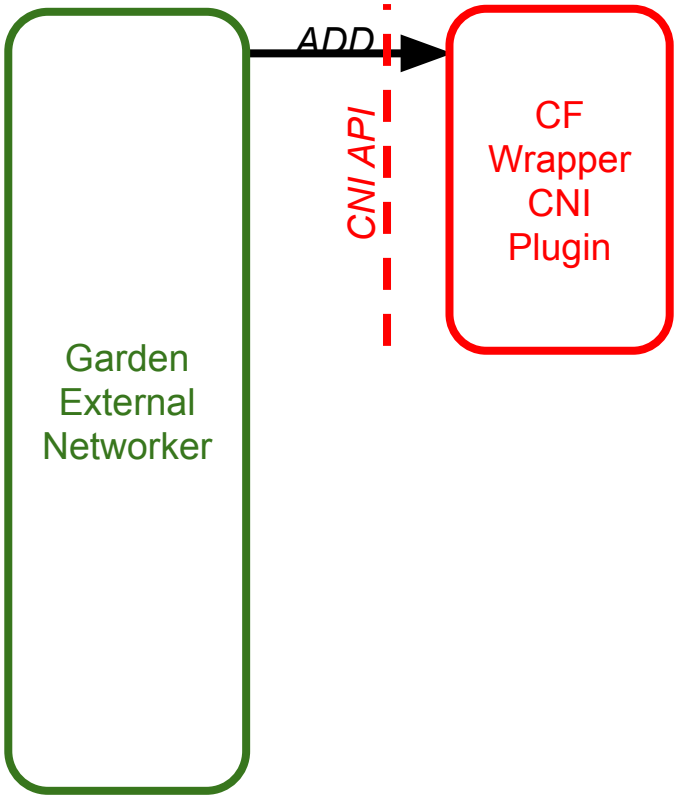


CNI Chaining

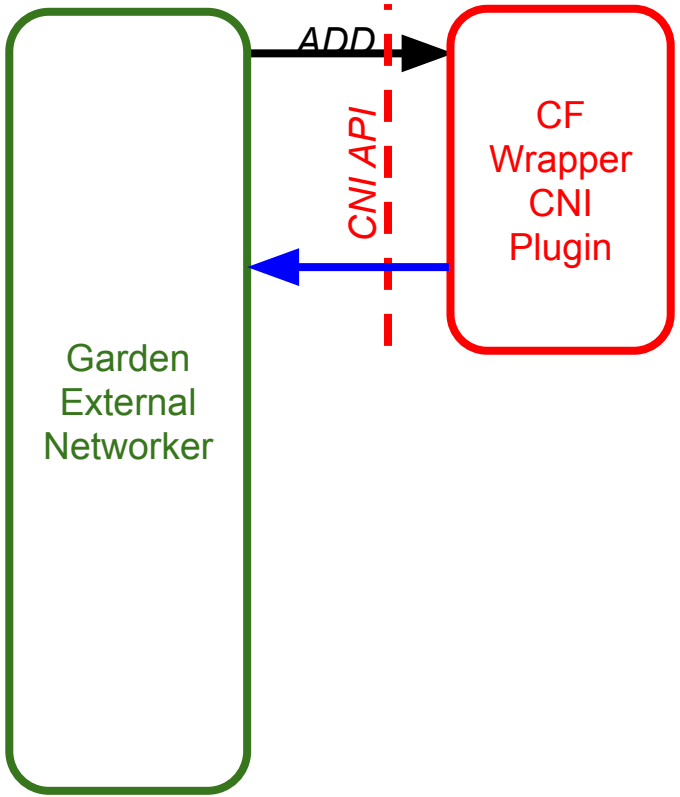


CNI Chaining

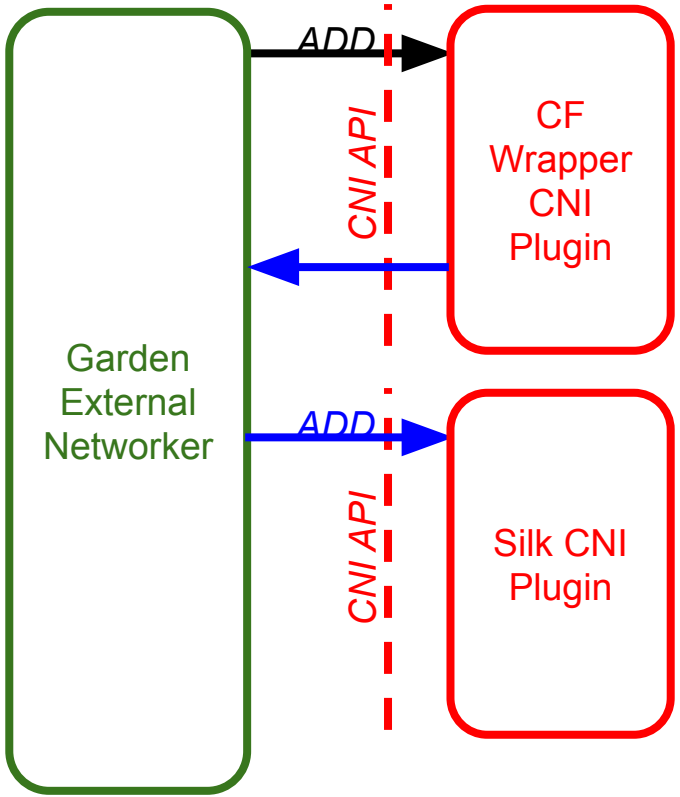
CNI Chaining



CNI Chaining



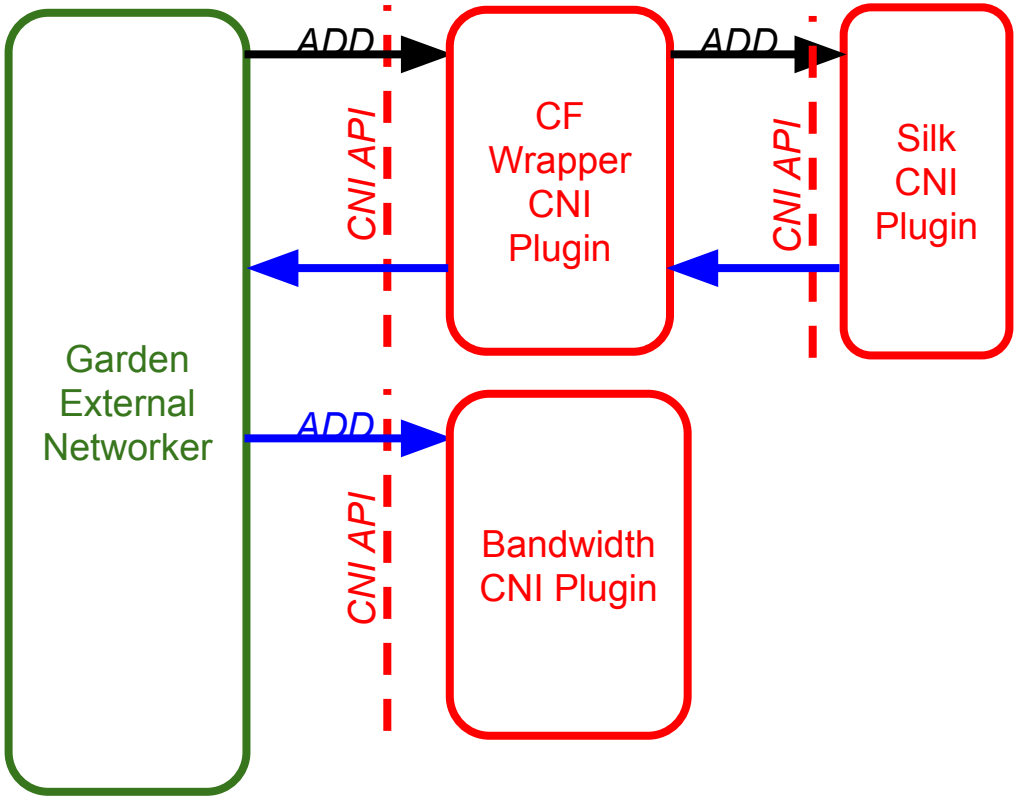
CNI Chaining



The Future of Silk

- CNI Chaining
  - Bandwidth Limiting Plugin

CNI Chaining



The Future of Silk

- CNI Chaining
  - Bandwidth Limiting Plugin
- Silk used by other container runtimes

Agenda

- What is CNI
- What is Cloud Foundry
- Cloud Foundry + CNI
- Main Motivations to Build Silk CNI
- Building Silk
- What's Next?
- **Key Takeaways**

Key Takeaways

- CNI provides a simple interface for pluggable networking
- Look to CNI repo for list of plugins before re-inventing the wheel
- Silk CNI is a plug-in built to meet availability and reliability requirements in a dynamic environment
- CNI specification v0.3.0 enables new use cases with plugin chaining

KubeCon

CloudNativeCon

Europe 2018

Thank you!

Angela Chin, achin@pivotal.io
Usha Ramachandran, uramachandran@pivotal.io