Datacenter network at Vinted

Rytis Lebedys
24th September 2025

Vinted

Agenda

Introduction (Topology, Scale)

Managing the network (Ansible, Inspec Tests)

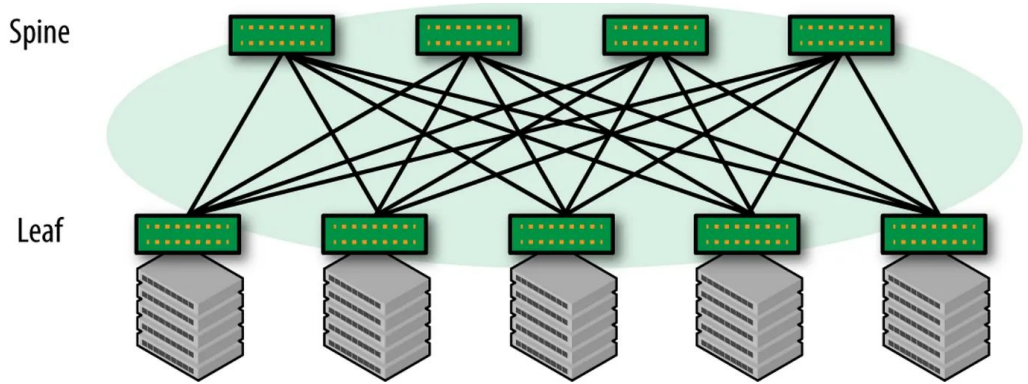
Network services (Kubernetes, Anycast, EVPN)

Network Operating Systems (Cumulus, SONiC)

Introduction

Network Topology Overview

- CLOS topology
- 3 locations
- 150+ racks
- Almost 200 network devices



Managing the network

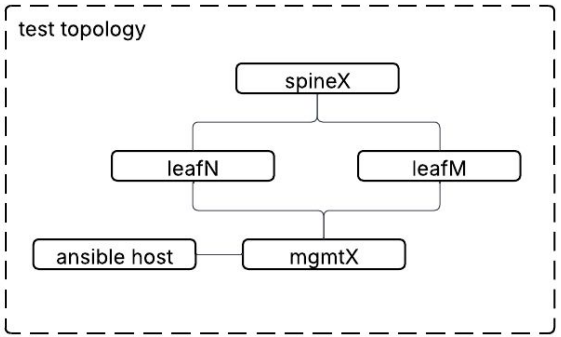
Network as code

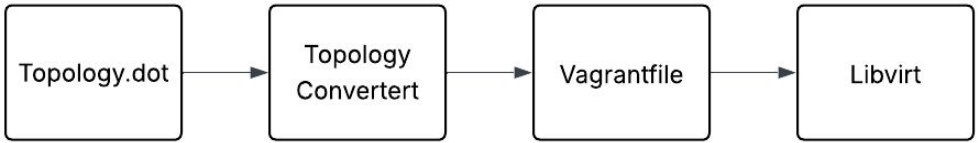
- Interfaces
- BGP
- Permissions
- Prometheus exporters
- DHCP relay

Automated tests - topology creation

```
//
// network links
//

// leaf -> spine: ports swp1..58
"dc1-leaf01":"swp1" -- "dc1-spine01":"swp1"
"dc1-leaf01":"swp2" -- "dc1-spine02":"swp1"
"dc2-leaf01":"swp1" -- "dc2-spine01":"swp1"
"dc2-leaf01":"swp2" -- "dc2-spine02":"swp1"
"dc3-leaf01":"swp1" -- "dc3-spine01":"swp1"
"dc3-leaf01":"swp2" -- "dc3-spine02":"swp1"
```

- Topology definition in graphviz DOT syntax
- Interconnections between VMs by UDP unicast tunnels

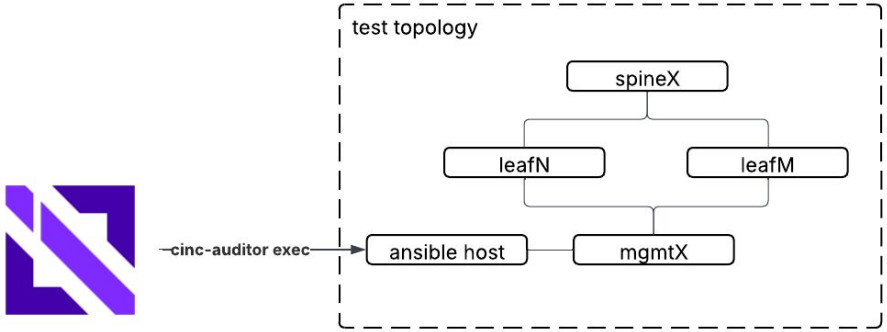




https://gitlab.com/cumulus-consulting/tools/topology_converter

Automated tests - cinc-auditor

- Cinc-auditor runs tests on ansible-host VM in test topology
- Tests utilise ssh, http, etc to get required state of the topology

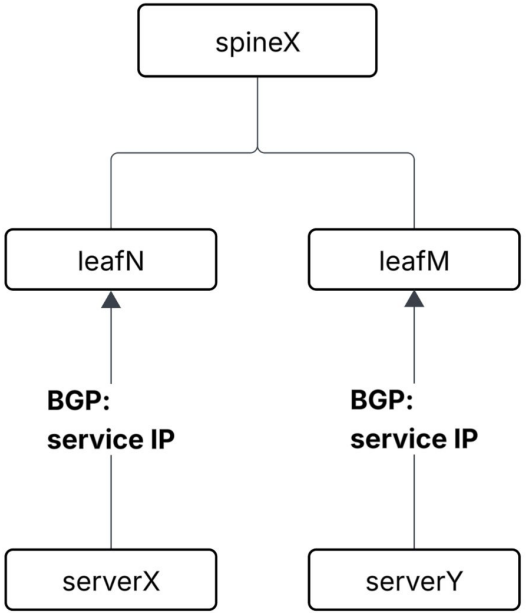



```
helpers.exits_dc1.each do |switch|
  describe net_bgp_route(switch, route: '172.17.4.26/26', vrf: 'external') do
    it { should_not exist }
  end

  describe net_bgp_route(switch, route: '172.17.4.23/32', vrf: 'external') do
    it { should exist }
    its('next_hop_count') { should eq 3 }
    its('next_hops') { should include '172.17.1.3' }
    its('next_hops') { should include '172.17.2.3' }
    its('next_hops') { should include '172.17.3.3' }
  end

  describe net_bgp_route(switch, route: '172.17.10.28/32', vrf: 'external') do
    it { should_not exist }
  end
end
```

Network services

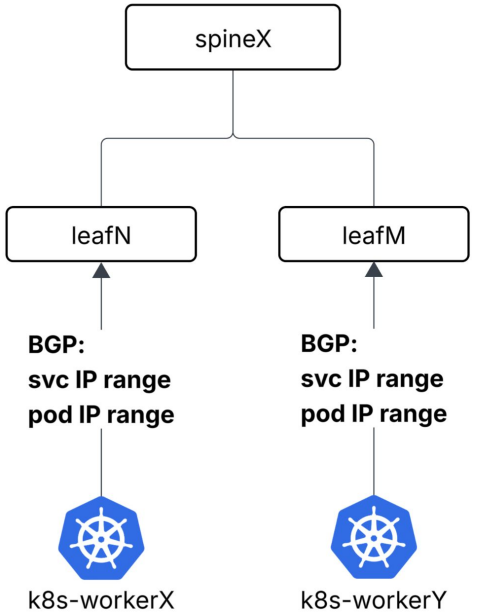
Network services - anycast for service HA

- Service HA is implemented by multiple servers advertising service IP via BGP
- In case one server goes down, it stops advertising service IP and traffic shifts to other servers
- We use this for NAT, front reverse proxies, PXE, etc.



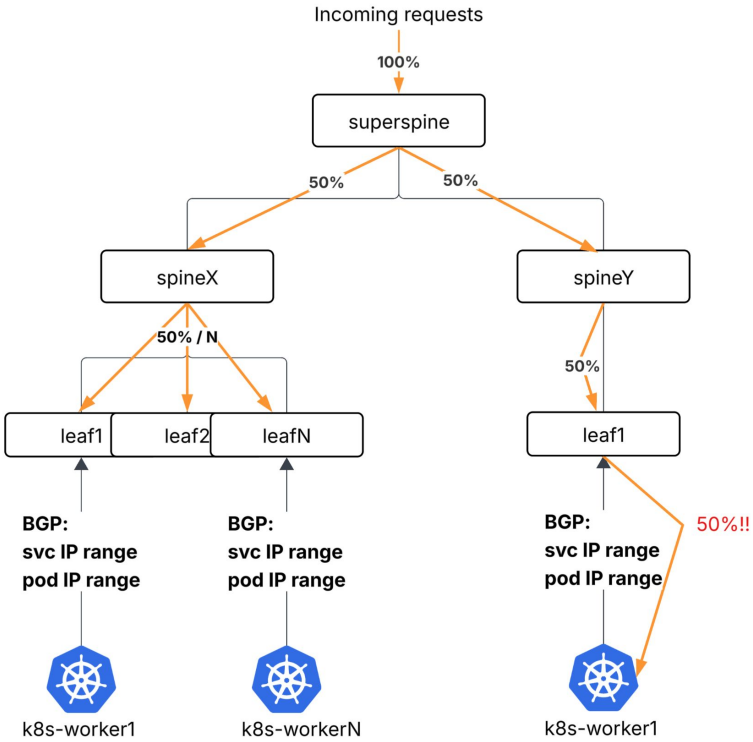
Network services - kubernetes (1)

- Kubernetes workers integrate into network over BGP with leaf switch in rack
- Advertises svc IP range (anycast) and pod IP range (unique per worker)

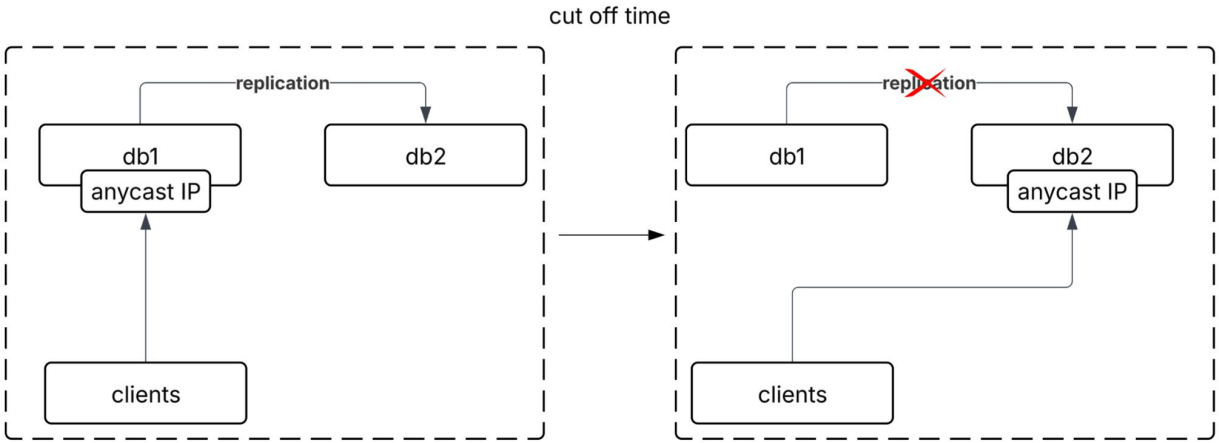


Network services - kubernetes (2)

- We had only few kubernetes worker nodes connected to spineY
- 50/50 traffic split consumed all available NIC traffic on workers connected to spineY
- It resolved when we started using Istio with envoy as fronts receiving incoming traffic - it uses pod IPs as upstreams instead of svc IP.



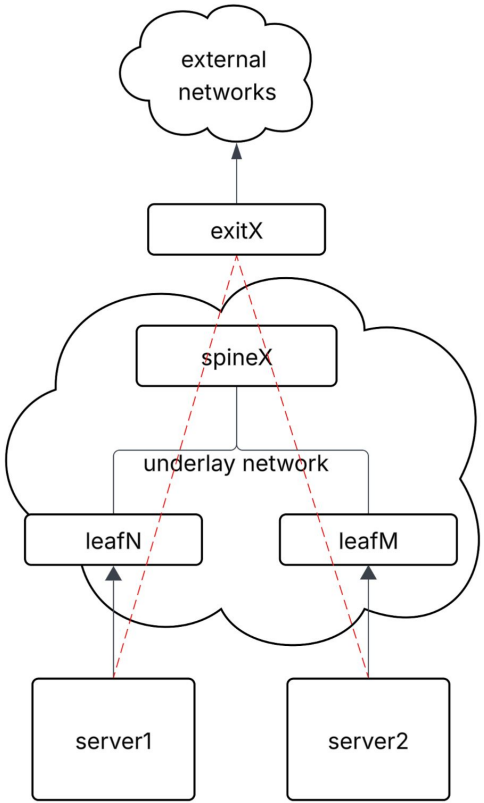
Network services - database sharding



- Anycast allowed us to have virtual IP for the database sharding in a L3 network
- In L2 network we would just move the IP to the other host and make sure ARP records got updated to the new MAC

Network services - EVPN

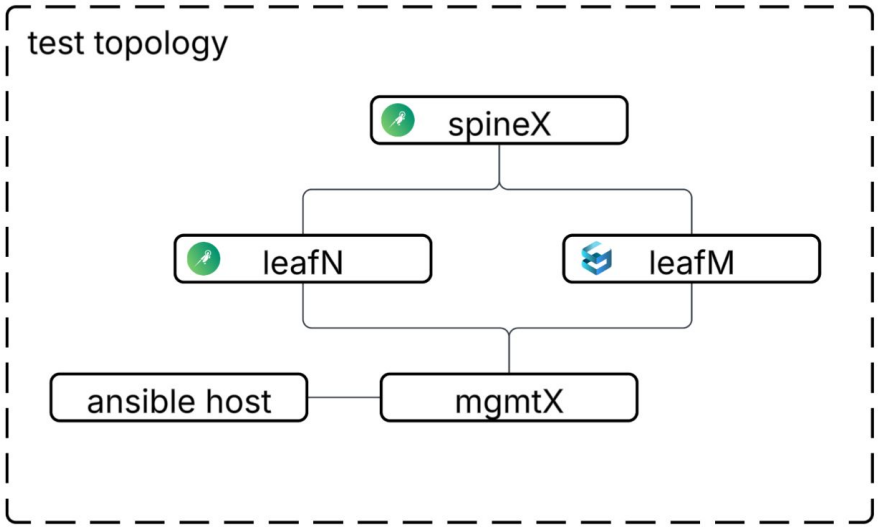
- EVPN for external network
- server1 and server2 connected to exitX as layer 2
- VTEPs are on leaf and exit
- External network in any rack/leaf
- Complex to troubleshoot
- Decommissioned at Vinted



Network Operating Systems

Network Operating Systems - Cumulus and SONiC

- Cumulus supported multiple hardware manufacturers until it got acquired by Nvidia
- We use SONiC instead of Cumulus for new hardware
- SONiC image - community version built in house
- We use both in our testing env - cumulus VX and sonic VS



Summary

Summary

- Network as code - from the beginning
- Automated tests and easy to set up test environments
- As simple as possible - no unnecessary functionality
- Still no dedicated network team at Vinted

Questions?