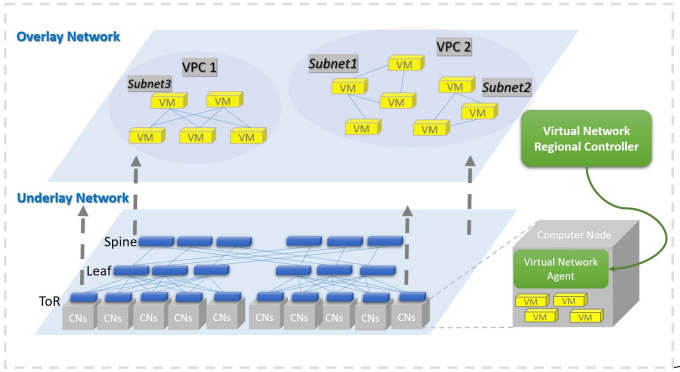
Hyperscale Cloud Network Systems

Dr. Liguang Xie
Email: lxie@futurewei.com
Sr. Principal Architect, Cloud Networking
Seattle Cloud Lab

12/02/2022

FUTUREWEI Technologies

Cloud Network System Research



DC Network Complexity

- Multiple layers of network virtualization and controllers
- Device heterogeneity
- Constant and frequent churns
- Unavoidable human errors

DC Network Scale

- Large machine pool (1 million of serves)
- Large size & number of VPCs
- Extra-high throughput (10s Tbps & 100s millions of active connections)

Huawei Cloud Global Infrastructure

- DCN
- Regional networks
- Inter-region backbone networks
- Local zones
- Edge networks

Research Areas

- Cloud Native Unified SDN Platform
- Cloud-Scale Programmable Data Plane
- Large Scale Cloud Emulation Platform
- SmartNIC Offloading & Hardware Acceleration
- Network Observability & Intelligence
- Mizar eBPF Data Plane
- Arion Cloud Gateway
- Alcor SDN
- Merak Emulator

FUTUREWEI Technologies

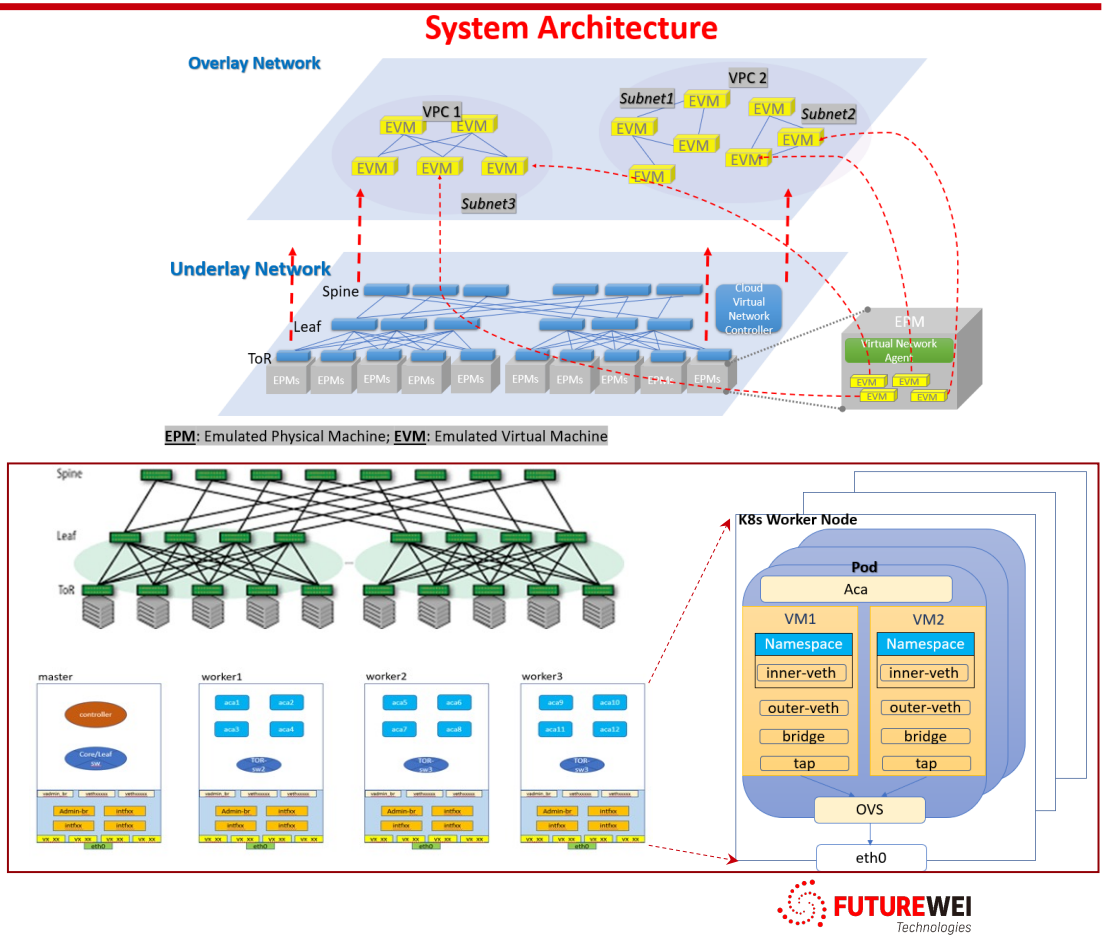
Merak – Cloud Emulation Platform

Goal & Vision

- Emulating large-scale, highly extensible and high-fidelity physical and virtual networks to achieve **100K** emulated physical machines and **10M** emulated virtual machines
- Light-weighted container and virtualization techniques for high density emulation, with a target of **50K** VMs per worker node & **10-100X** density bump
- On-demand configuration of physical and virtual networks and multi-tenancy
- Support verification and scale testing of underlay & overlay networks
- Support functional testing of new data plane techniques (e.g., eBPF/P4)

What we can emulate (as of now)

- Underlay/physical network
- Overlay/virtual network
- VM provisioning (w/o scheduling)
- Open benchmarking framework for testing different control plane and data plane targets
  - Virtual network control plane: Neutron, OVN, Alcor
  - Virtual network data plane: OVS, eBPF
  - Physical network control plane: Ryu, ONOS

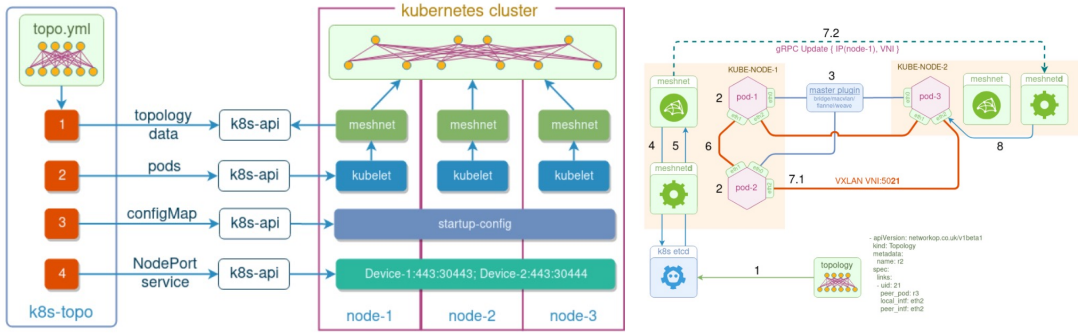


FUTUREWEI Technologies

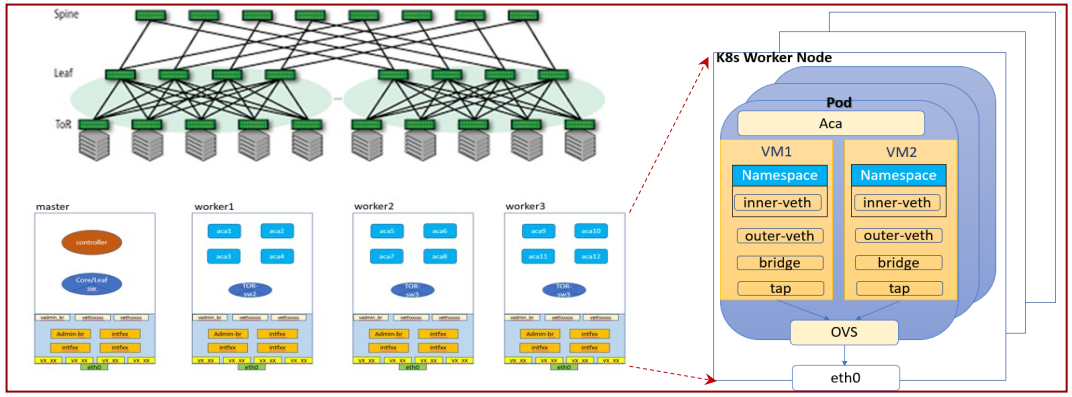
Merak – Cloud Emulation Platform (cont'd)

- **Highlight of Architecture**
  - Configurable and automated deployable underlay network topology via K8s/ MeshNet (multi-layers of switch/router, vhost and links)
  - Merak emulation core:
    - use container to emulate network devices and compute nodes
    - use namespace to emulate vm

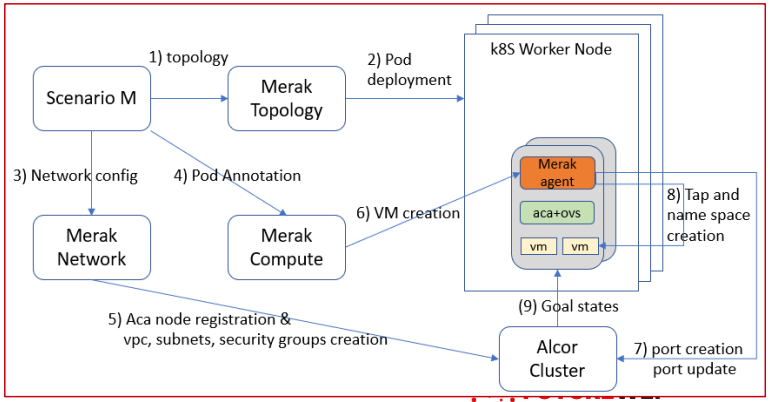
Underlay Network (K8s-topo => K8s API => Meshnet CNI)



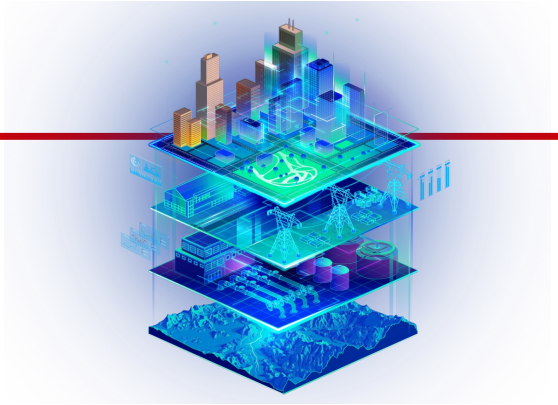
Overlay Network



Merak Software Architecture and Emulated VM Boot Workflow

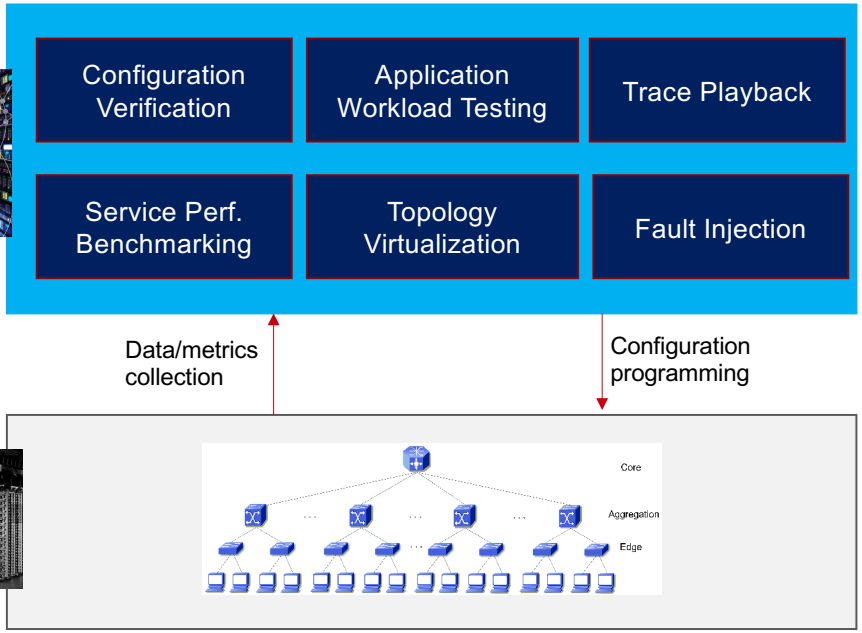


Work In Progress



Building **Cloud Digital Twin (CDT)**

- Extend CDT to IaaS/PaaS services including compute (VM/container scheduling) and storage
- Customer workload (AI/ML, Gaming, Web etc.) benchmarking platform
- Application/service configuration validation
- Fault injection…



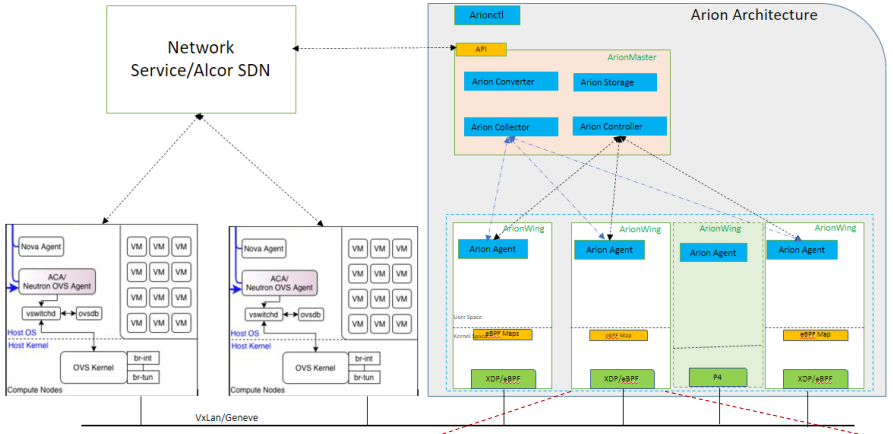
Potential Collaboration Areas

- Configuration verification & correction
- Traffic trace playback and fault injection
- Applications built on CDT
  - ✓ Data center energy
  - ✓ Cloud and data center security
- Much more…

Arion – Programmable Cloud Data Plane Platform

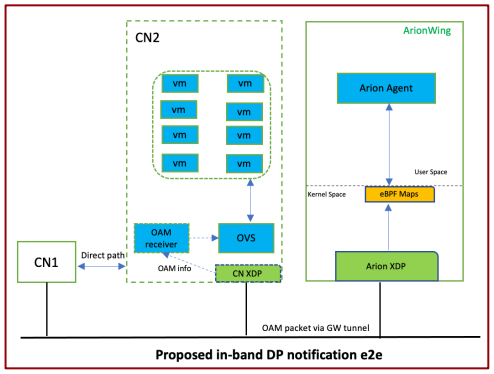
Goal and Vision

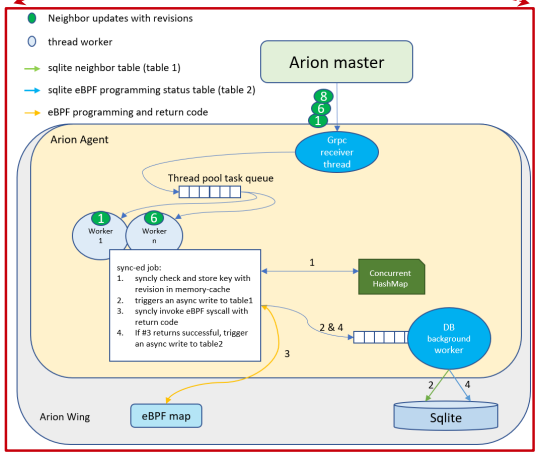
- 10Tbps cloud west-east gateways for large VPCs (1M VMs)
- Hybrid hardware and software codesign
  - Leverage advantage of eBFP/XDP & P4 Tofino switches
  - Reduce platform cost
- Supports fast network provisioning (scale up to 100M configurations) & watch
  - average r/w latency of single configuration 20us and 1ms
- Single machine throughput near line rate at 400Gbps, cluster throughout reaches 10Tbps
- Unify programmable interface of and simply control plane design



Architectural Features

- High-performance software-hardware codesign with unified P4-based programming interface
- Three-layer cache design (hot/warm/cold) combined with state reconciliation for ultra-low e2e programming latency (<20 us) at scale
- SmartNIC offloading of eBPF/XDP for high packet processing throughput (17+ Mpps)
- Work published on NSDI 2022 and IEEE ICPADS 2022





Thank You.

Open-Source Project:

1. Merak: https://github.com/futurewei-cloud/merak
2. Mizar: https://github.com/CentaurusInfra/mizar
3. Arion: https://github.com/futurewei-cloud/arion
4. Alcor: https://github.com/futurewei-cloud/alcor

Questions later?

Feel free to send email to lxie@futurewei.com