
KubeCon
CloudNativeCon
OPEN SOURCE SUMMIT
China 2019

From Secure Container to Secure Service

Xu Wang & Fupan Li, Ant Financial

Back to KubeCon NA 2018

We summarized the progress.

We talked about the overhead and other issues.

And we predicted the secure containers is going to production in 2019.

Additional Background

Long tutorial in KubeCon NA 2018 by Lei & Me

Hands-on: K8s + containerd + Kata Containers

Deck and Video:

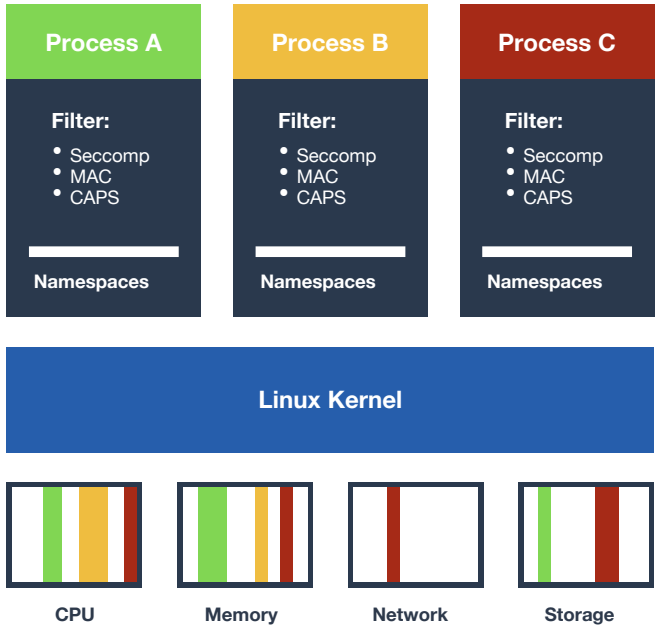
https://kccna18.sched.com/event/GrZN/tutorial-katacontainers-the-hard-way-kubernetes-containerd-katacontainers-lei-zhang-alibaba-xu-wang-hyperhq-limited-seating-available-see-description-for-details

*"The only real **solution to security** is to admit that bugs happen, and then mitigate them by having multiple layers."*

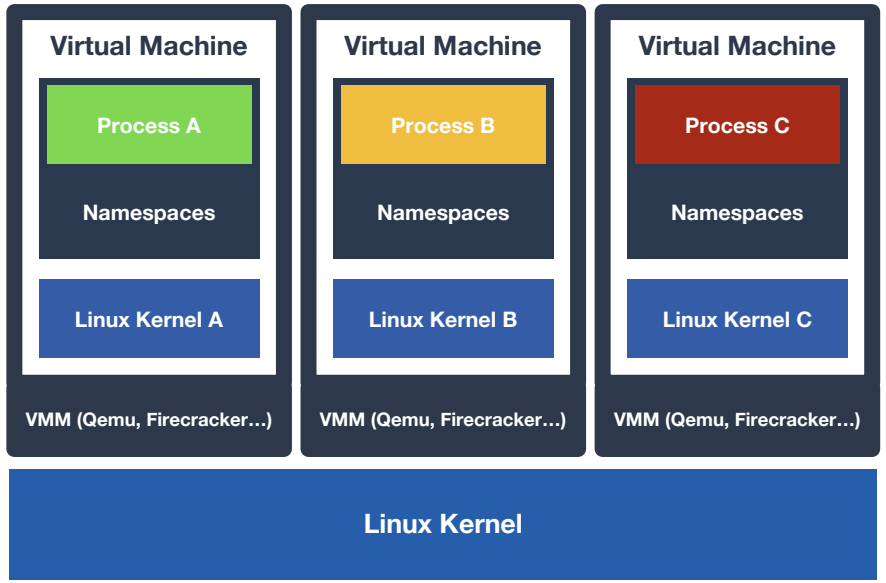
---Linus Torvalds (LinuxCon NA 2015, Seattle)

Container Runtimes on Linux

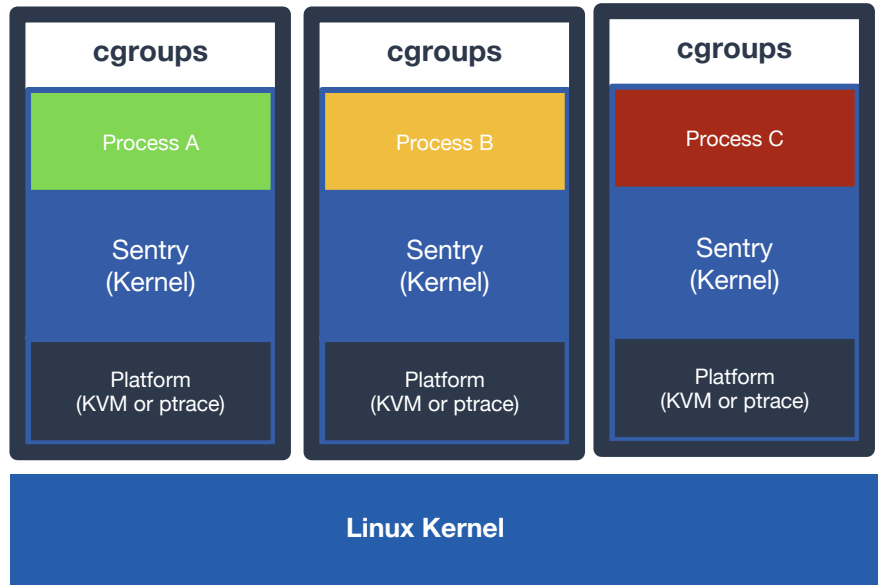
Linux Containers
By Process Isolation



Kata Containers
(Secure Container)

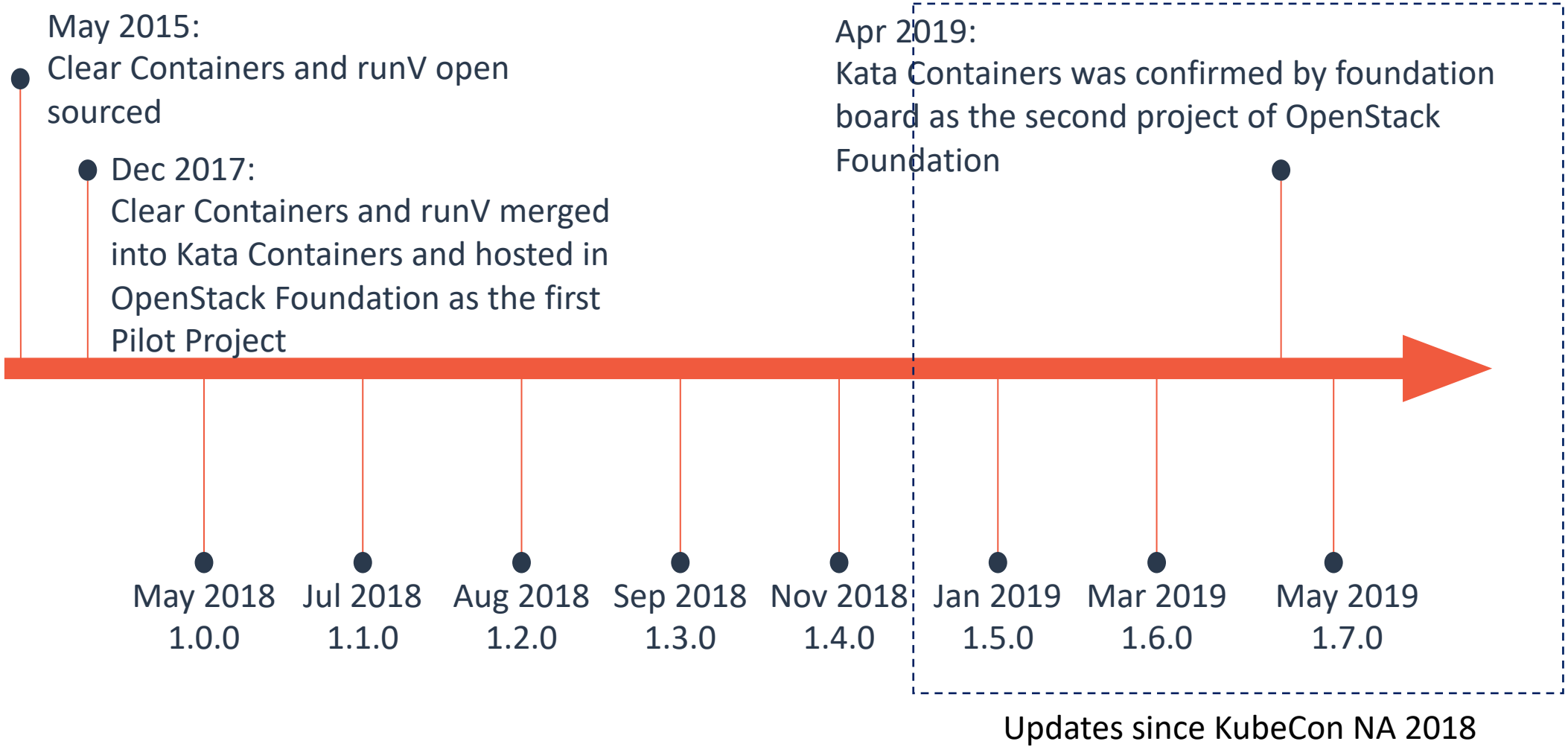


gVisor
(Secure Container)

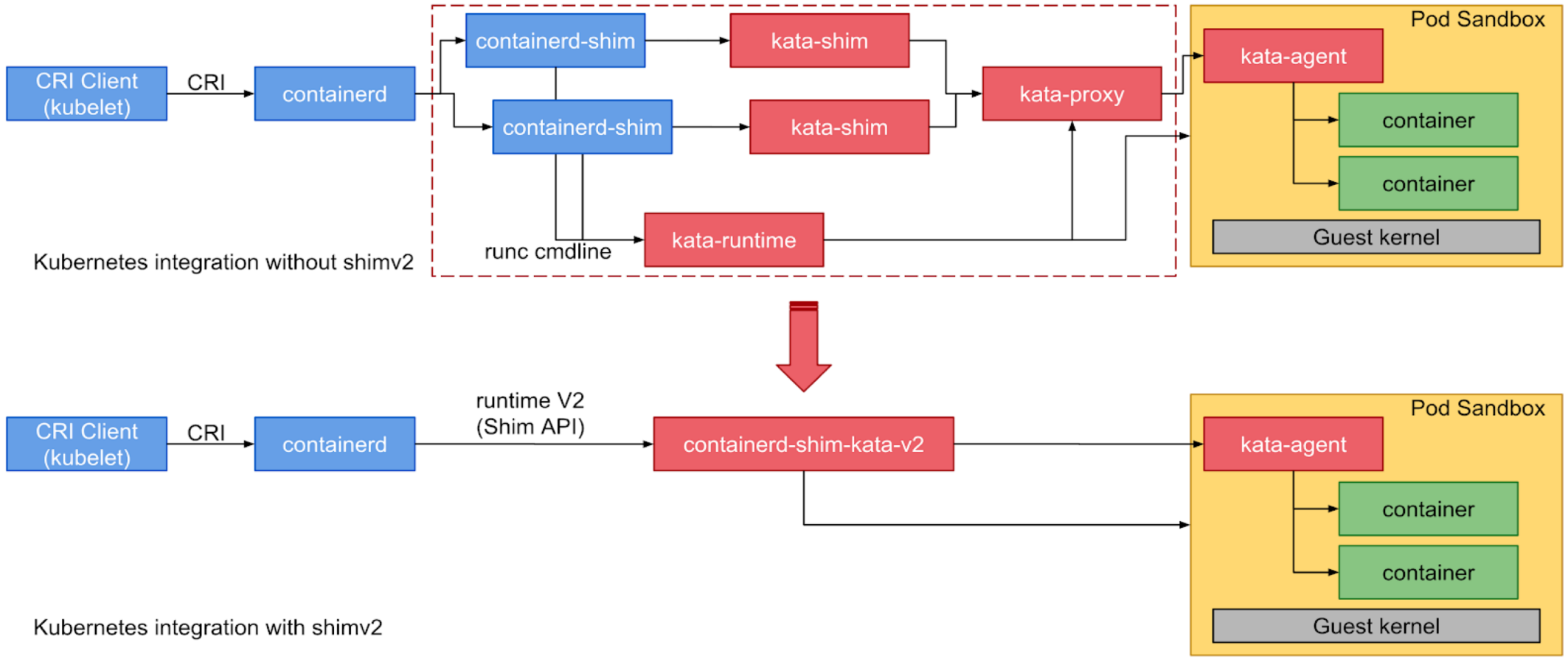


- Independent kernel for each POD sandbox
- Resource Isolation + Security Isolation

A Brief History of Kata Containers



Shim v2 Support in Kata 1.5



Eliminated 2N+1 helper processes

FireCracker Support in Kata 1.5

Firecracker

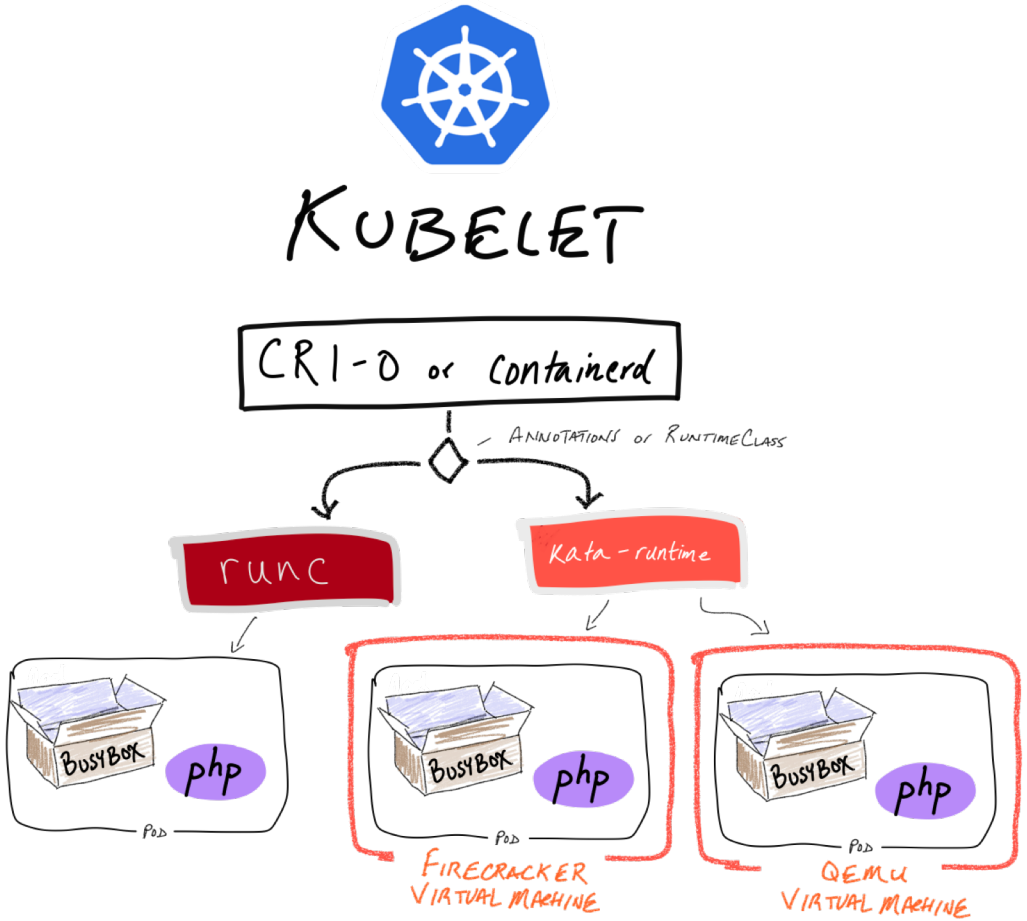
- Open sourced by AWS - Nov 2018
- From their GitHub page:

> *"Firecracker has a minimalist design. It excludes unnecessary devices and guest-facing functionality to reduce the memory footprint and attack surface area of each microVM. This improves security, decreases the startup time, and increases hardware utilization."*

Kata + Firecracker integration status

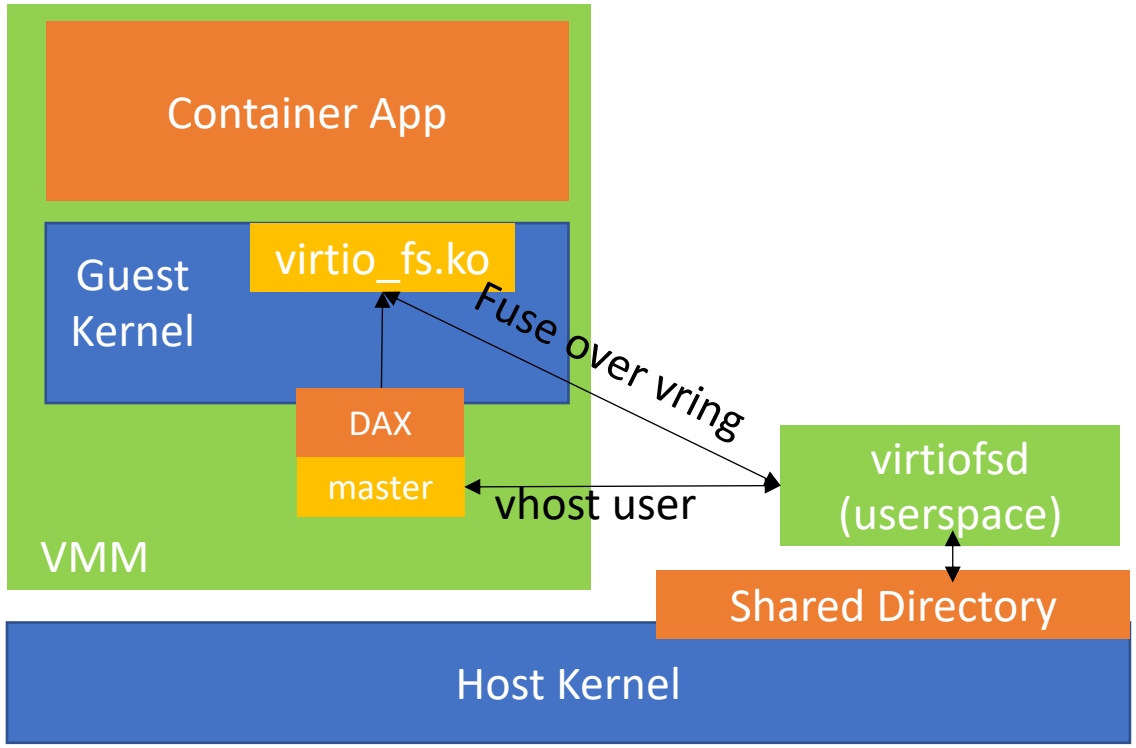
- With minimal design of the VMM, there are limitations when using Kata+Firecracker:
  - No filesystem sharing with host
  - No hardware device support
  - No dynamic resizing of the guest (vCPU/memory hotplug)

Work with Kubernetes RuntimeClass



- On each node, you can run workloads which will utilize runc, kata-qemu and kata-firecracker.
- You can select your method of isolation on a per-workload (per-pod) basis

Virtio-fs Support in Kata 1.7



virtio-fs

- Origin from RedHat
- Based on fuse, better POSIX compatibility
- VirtIO based, native design for virtualization (not another network FS)
- With DAX, better performance and lower memory overhead in guests
- Userspace virtiofs daemon, more flexible

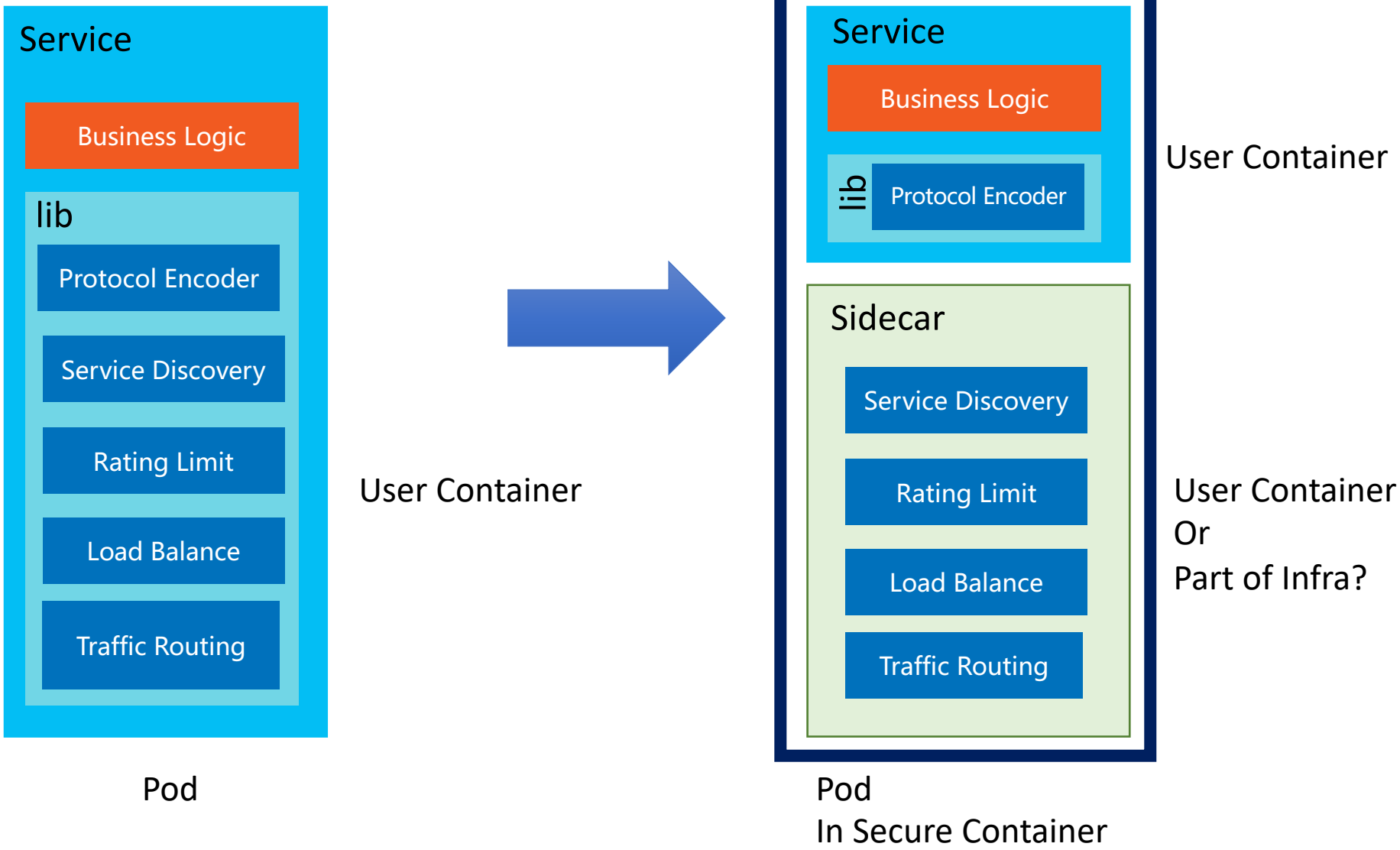
Summary of the Progress

- Better integration with Kubernetes
- Less memory overhead
- Improvements on filesystem sharing

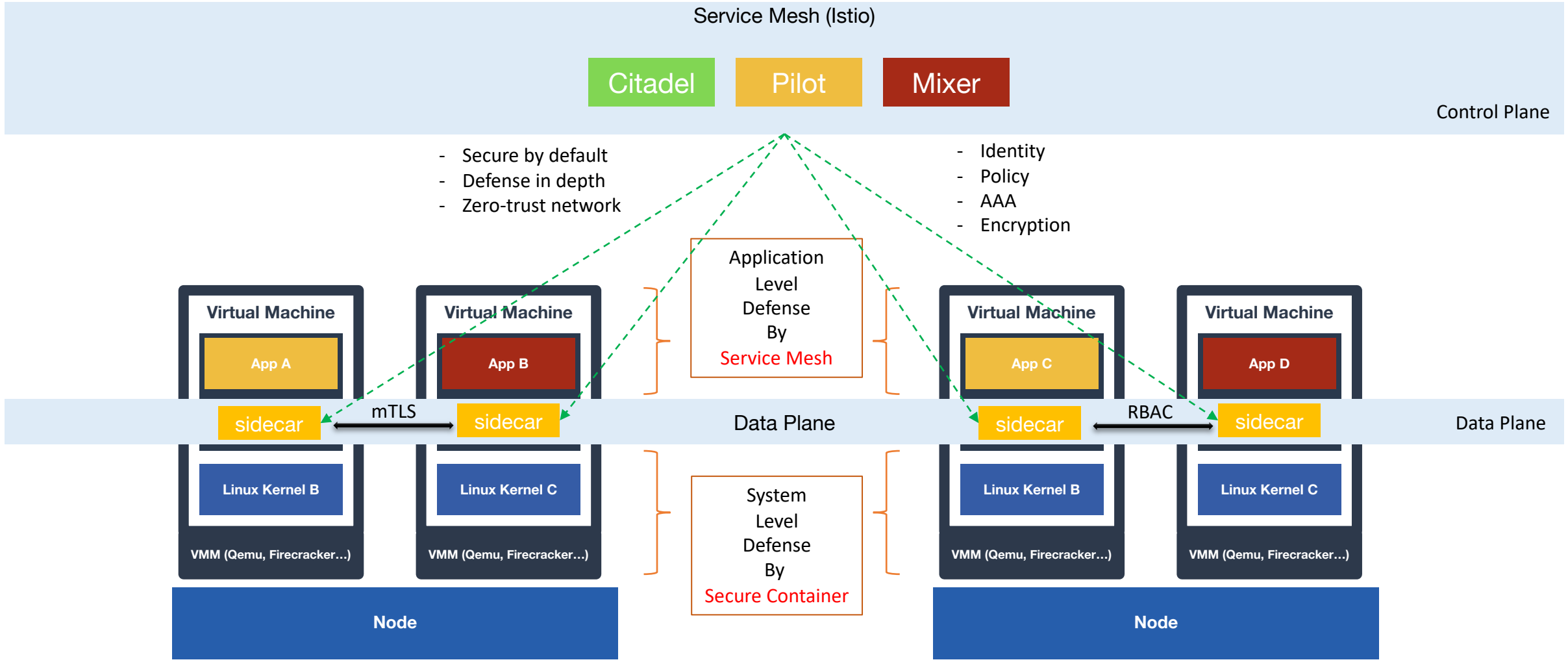
Well, Security is an End-to-End Issue

We need not only secure container runtime, but secure services in Financial Scenarios.

ServiceMesh:
Evolution of the Financial Grade Infrastructure

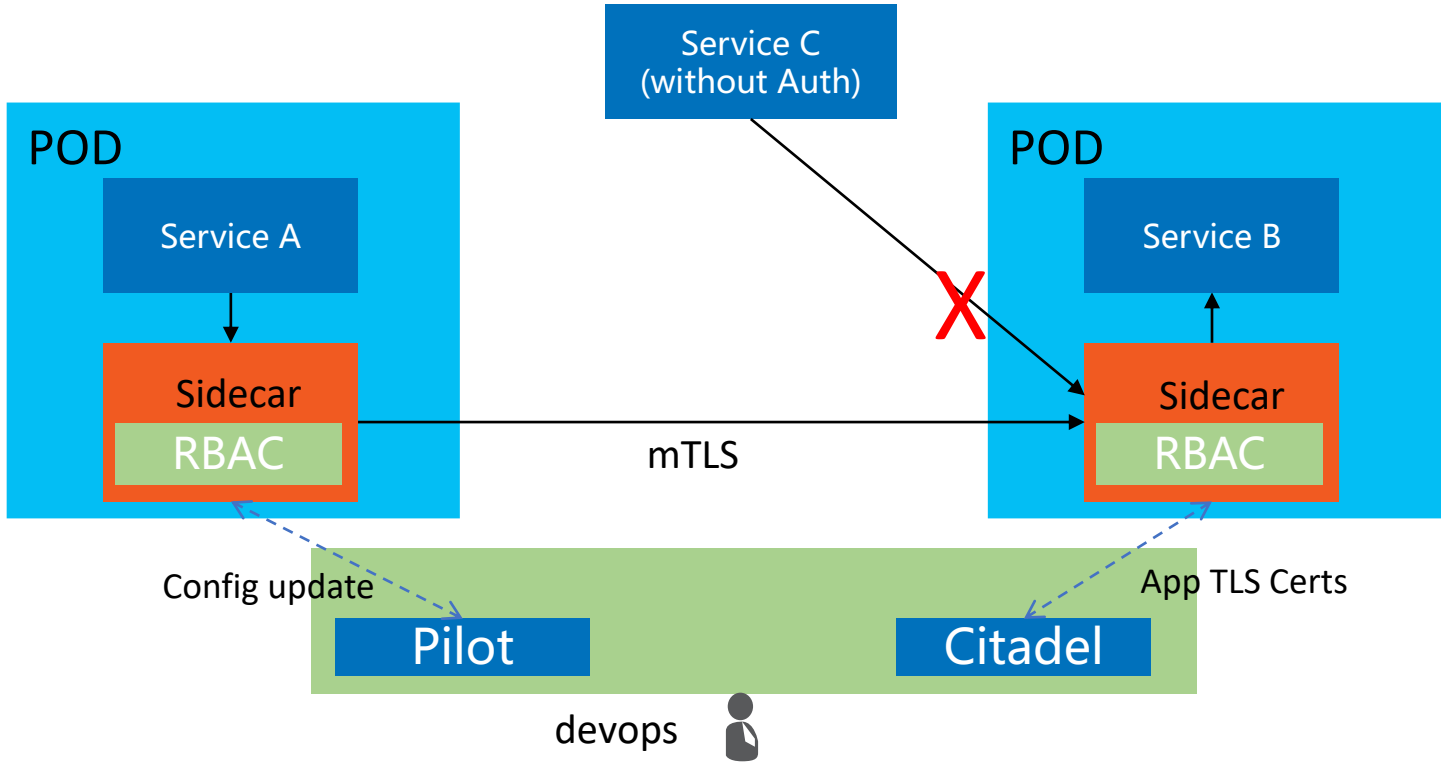


Service Mesh + Kata Containers



Demo: Kata + Service Mesh

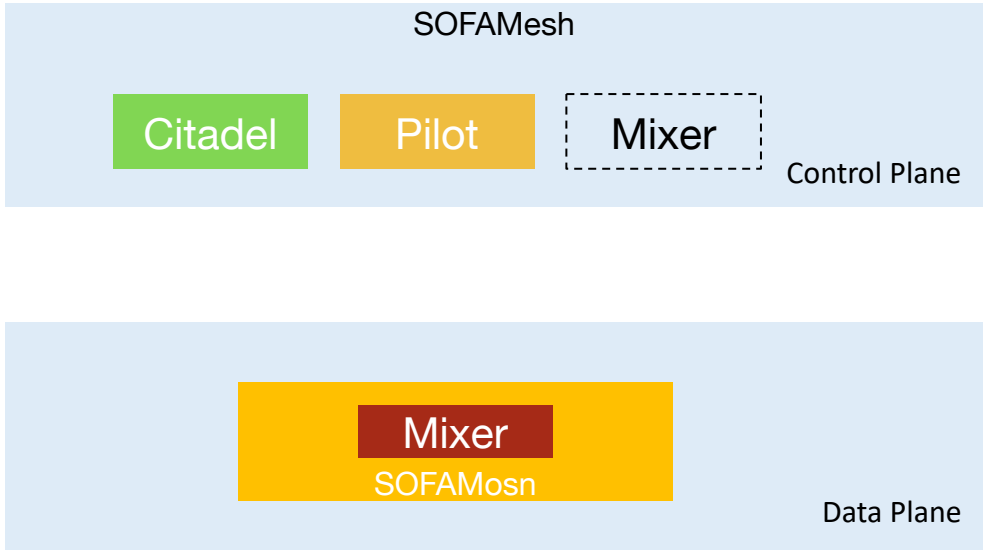
- ServiceMesh Security Mechanisms and Kata Containers
  - Enforce mTLS Data Plane for Kata + Istio (video)
    https://istio.io/docs/tasks/security/authn-policy/#namespace-wide-policy
  - Enable RBAC for ingress traffic for Kata + Istio (video)
    https://istio.io/docs/tasks/security/authz-http/#enforcing-namespace-level-access-control



SOFAMesh: Service Mesh Practice in Ant Financial

SOFAMesh

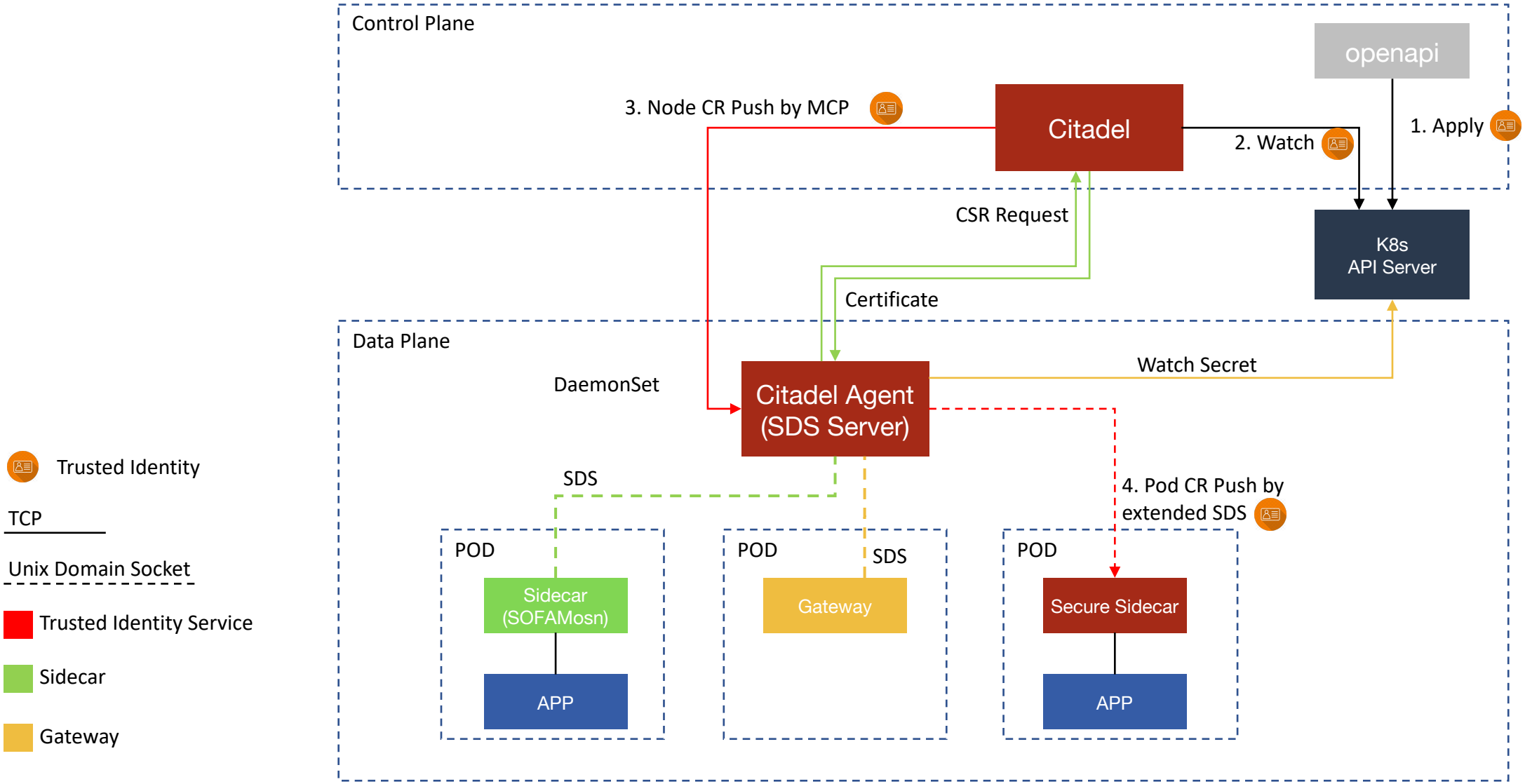
- **Large-scale** Service Mesh Practice
- Based on Istio, with improvements and extensions
  - SOFAMosn (in golang) as sidecar to replace envoy
  - Migrate mixer to data plane for performance
  - Improve Pilot for more flexible service discovery
  - Performance improvement of Pilot
- Support RPC：SOFARPC/Dubbo/HSF
- Verified in Ant Financial, and feed back to community
- Open Source: https://github.com/sofastack/sofa-mesh



SOFAMosn

- Not only Service Mesh Sidecar in SOFAMesh
- But also：API Gateway，Ingress Gateway
- Support envoy xDS v2 API
- Open Source: https://github.com/sofastack/sofa-mosn

Practice: Trusted Identity Service



The Next Step

- Current:
  - Kata works with Istio / SOFAMesh
- In the Future:
  - Mesh sidecar optimization in Kata Context (w/ eBPF etc.)
    - And Interoperability with non-kata containers
  - Resource isolation between mesh sidecar and user containers


KubeCon
CloudNativeCon
OPEN SOURCE SUMMIT
China 2019
谢谢！ Thank You