Безопасность от виртуальных машин до контейнеров и обратно

Дмитрий Евдокимов
Founder&CTO Luntry

Мона Архипова
Независимый эксперт

Industrial++ 2024

О нас

Founder & CTO Luntry

Опыт в ИБ более 15 лет

Докладчик: BlackHat, HITB, ZeroNights, HackInParis, SAS, OFFZONE, PHDays и др.

Автор Telegram-канала “k8s (in)security”
Автор курса “Cloud Native Sec в K8s”

Независимый эксперт

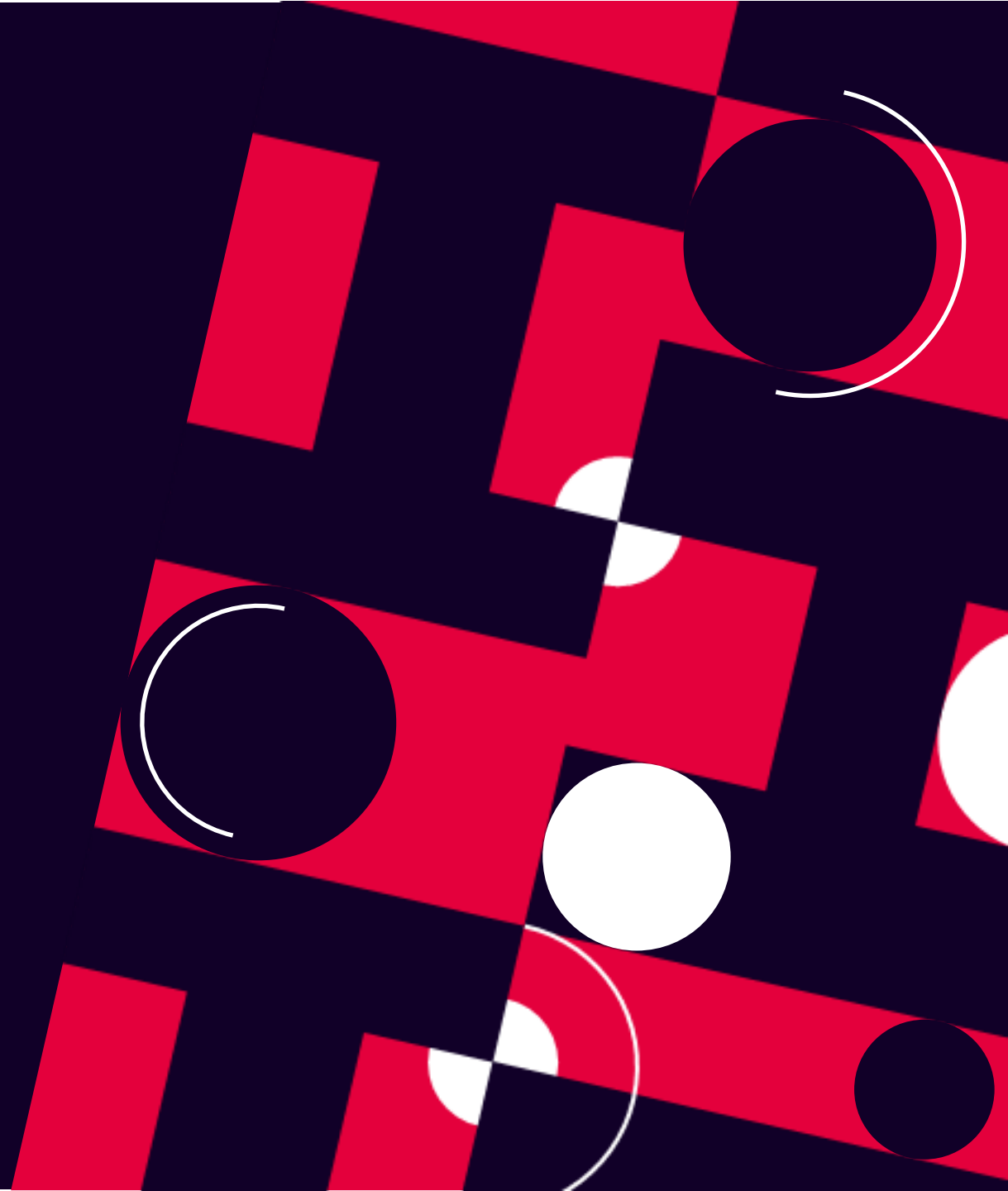
Более 17 лет в ИТ, из них более 12 – еще и ИБ

Докладчик и член программных комитетов ведущих отраслевых конференций

vCISO/vCIO для стартапов и SMB
Все еще играющий тренер

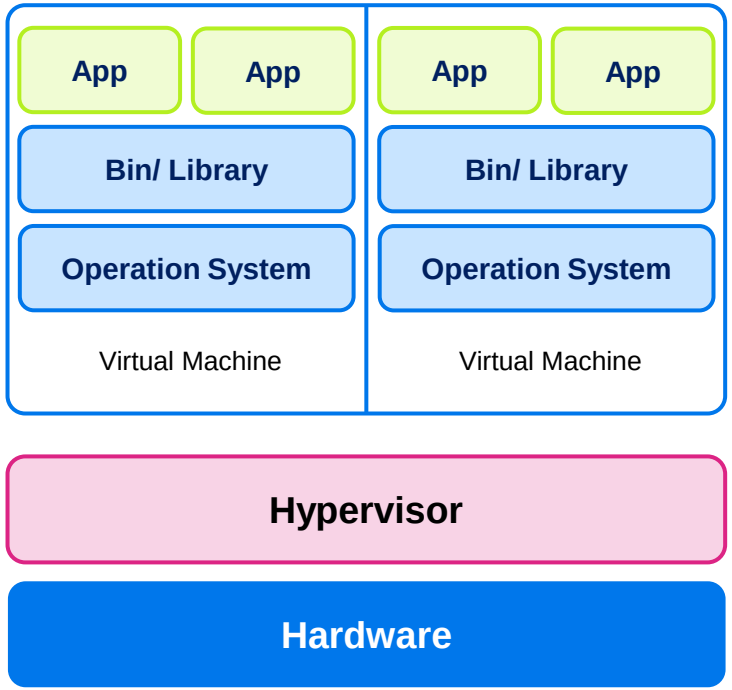
Виртуализация vs контейнеризация

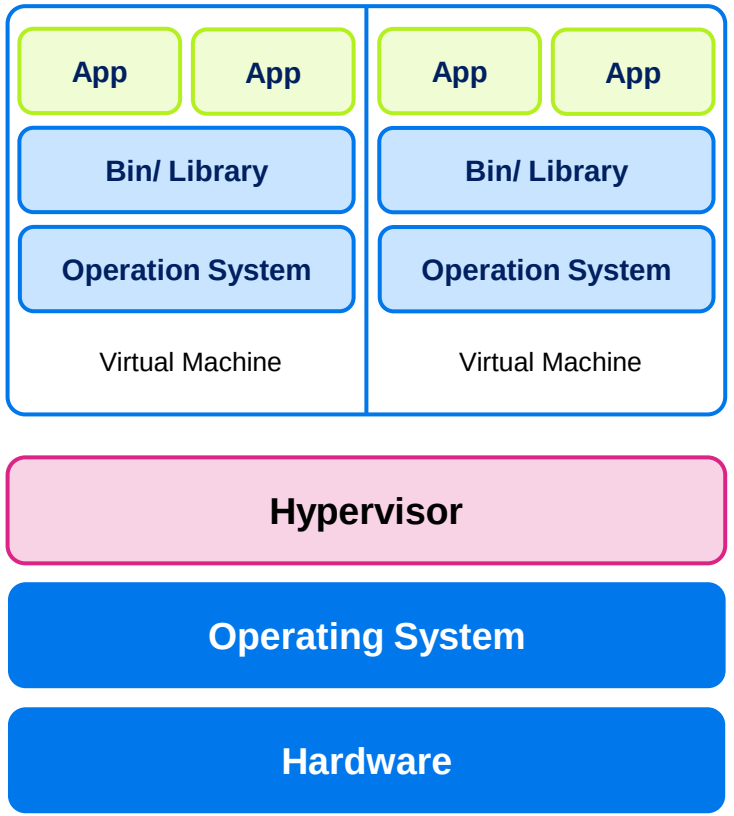
База



Industrial++ 2024

Виртуализация





IBM

vmware

KVM

Microsoft Hyper-V

Xen Project

Industrial++ 2024

Что такое виртуальная машина?

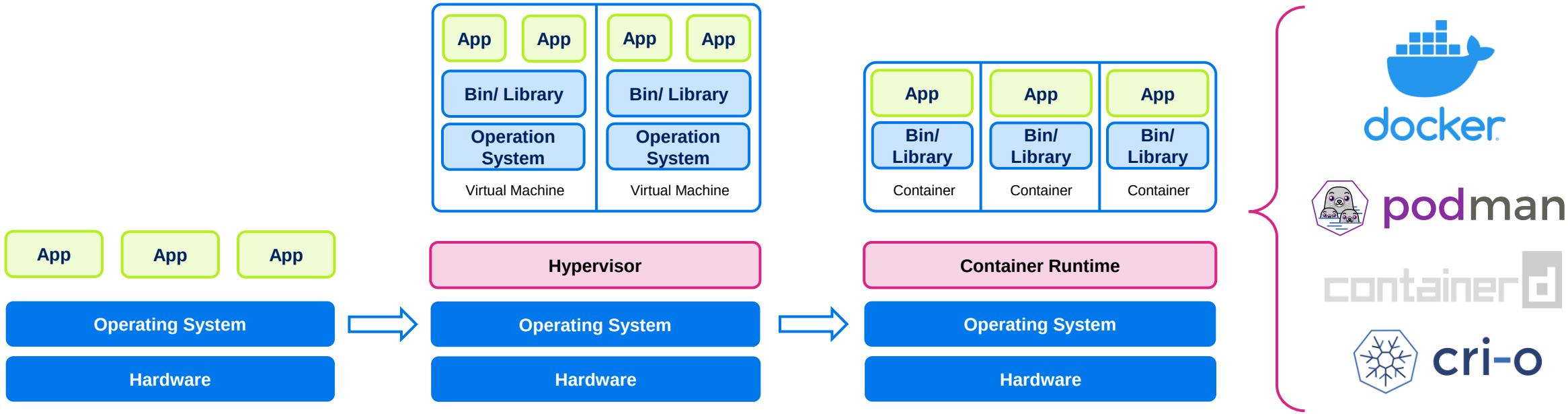
VM – это полноценная «копия» физического оборудования:

- Полная аппаратная виртуализация
  - X86
  - RISC
- На уровне ОС
  - Динамическая рекомпиляция
  - Паравиртуализация/портирование
  - Перехват guest-вызовов

Регуляторика по виртуализации

- В силу зрелости – технология понятна, принимается в рамках как локальных, так и глобальных требований регуляторов и при добровольных сертификациях
- Приказ ФСТЭК России от 18 февраля 2013 г. N 21
- ГОСТ Р 56938-2016 "Защита информации. Защита информации при использовании технологий виртуализации. Общие положения» (01.06.2017)
- Требования по безопасности информации к средствам виртуализации (выписка). Утверждены приказом ФСТЭК России от 27 октября 2022 г. N 187 (КИИ, АСУ ТП, ПД)

Контейнеризация



Традиционное развертывание

Виртуализированное развертывание

Контейнеризированное развертывание

Что такое контейнер?

Container – это Linux process с определёнными свойствами/ограничениями:

- что можно увидеть: namespaces (pid, user, uts, ipc, net, mnt), pivot_root (+ image)
- что можно делать: Capabilities, seccomp, LSMs
- что можно использовать: Control group (процессор, память, устройства, …)

```
   8554 ?        Sl   414:20 /usr/bin/containerd-shim-runc-v2 -namespace k8s.io -id d4c22c8f5248619f
   8602 ?        Ss     0:00  \_ /pause
2407498 ?        Ss   141:25  \_ /usr/bin/tini -- /usr/local/bin/docker-entrypoint -c /etc/filebeat.
2407509 ?        Sl   11158:26      \_ filebeat -c /etc/filebeat.yml -e
   8700 ?        Sl   400:29 /usr/bin/containerd-shim-runc-v2 -namespace k8s.io -id 487fb543a56f1455
   8721 ?        Ss     0:00  \_ /pause
   9699 ?        Ssl  302:20  \_ /bin/node_exporter
2332096 ?        Sl   363:31 /usr/bin/containerd-shim-runc-v2 -namespace k8s.io -id 331319558688f52c
2332117 ?        Ss     0:00  \_ /pause
2228218 ?        Ss     0:00  \_ nginx: master process nginx -g daemon off;
2228254 ?        S      0:00      \_ nginx: worker process
2228255 ?        S      0:00      \_ nginx: worker process
```

Регуляторика по контейнеризации

Выписка из Требований по безопасности информации, утвержденных приказом ФСТЭК России от 4 июля 2022 г. N 118

- Сертифицированная ОС
- Сертифицированное средство контейнеризации
- Сертифицированное наложенное средство безопасности

УТВЕРЖДЕНЫ
приказом ФСТЭК России
от 4 июля 2022 г. № 118

Требования по безопасности информации к средствам контейнеризации (выписка)

5. Настоящие Требования включают требования по безопасности информации, предъявляемые к:

уровню доверия средства контейнеризации;

хостовой операционной системе, в среде которой функционирует средство контейнеризации;

составу функций безопасности средства контейнеризации;

изоляции контейнеров средством контейнеризации;

выявлению уязвимостей в образах контейнеров;

проверке корректности конфигурации контейнеров;

контролю целостности контейнеров и их образов в средстве контейнеризации;

регистрации событий безопасности в средстве контейнеризации;

управлению доступом в средстве контейнеризации;

идентификации и аутентификации пользователей в средстве контейнеризации;

централизованному управлению образами контейнеров и контейнерами в средстве контейнеризации.

Оркестрация

Есть ли разница?

Оркестрация

ANSIBLE

puppet

Progress Chef

CFEngine

SALTSTACK

VMware Aria™

Hewlett Packard Enterprise

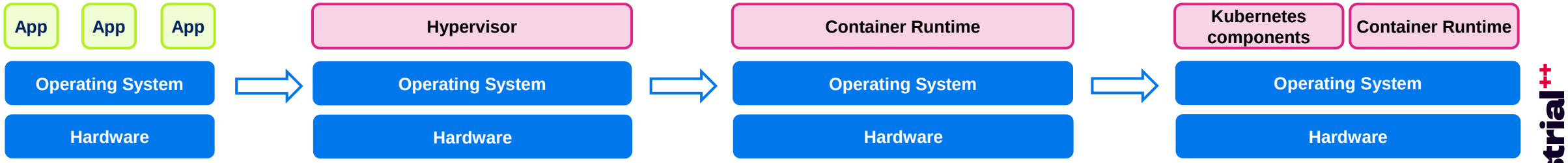
Microsoft System Center Orchestrator

docker SWARM

MESOS

HashiCorp Nomad

kubernetes



Традиционное развертывание

Виртуализированное развертывание

Контейнеризированное развертывание

Kubernetes-развертывание

Industrial++ 2024

Защита/изоляция данных

Industrial++ 2024

Гипервизоры и ВМ

- Установка с нуля с последующей конфигурацией
- Предсобранный образ

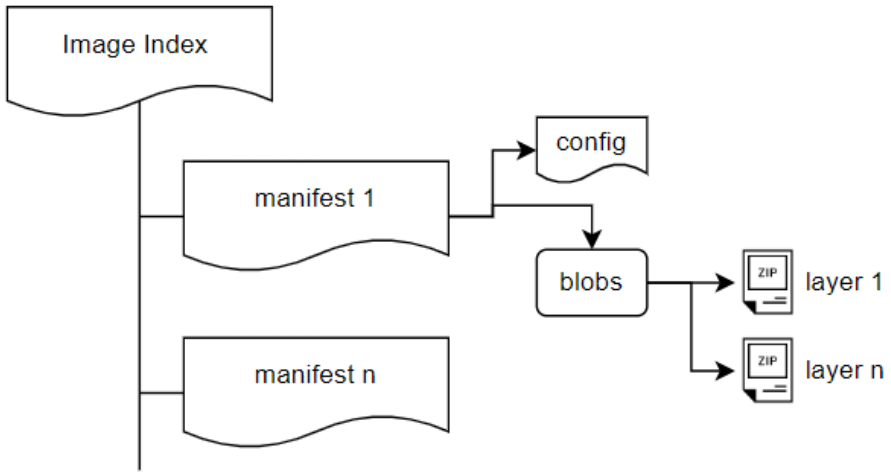
В обоих случаях:
полноценная эмуляция физического сервера и ОС.
В том числе устаревших.

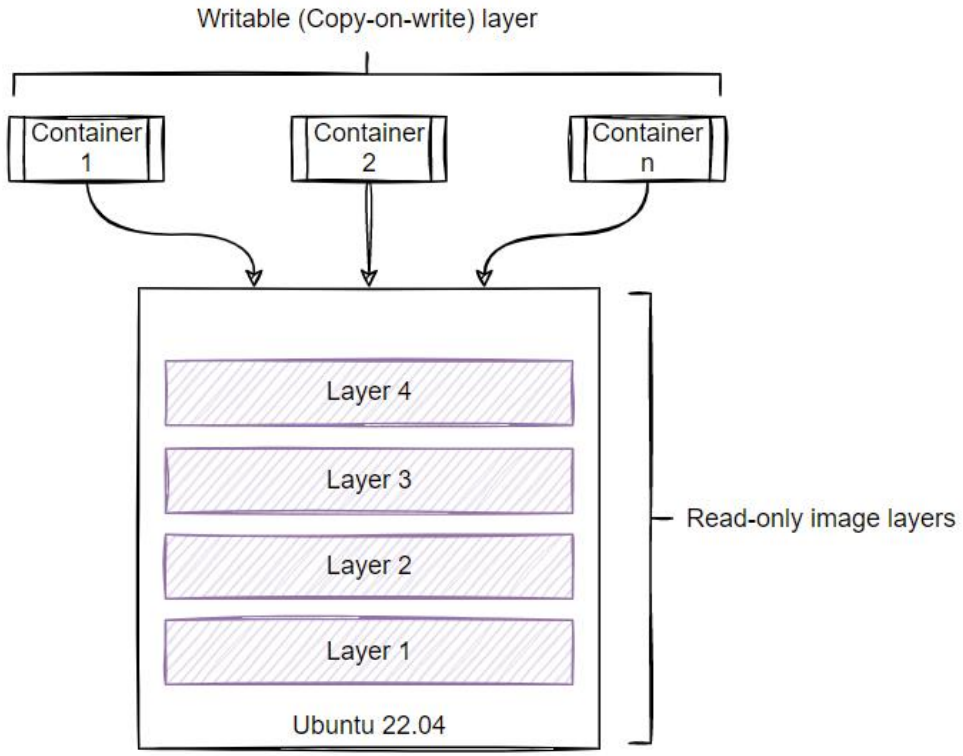
- Формат и локация хранения данных для каждой ВМ определяется на уровне гипервизора.
- Данные ВМ отделены друг от друга

Что такое образ контейнера?

Container Image – это неизменяемый пакет файлов операционной системы, кода приложения и любых зависимостей приложения

- Union File System
  - OverlayFS как реализация
- OCI image-спецификация





Защита и изоляция похожи

Защита образа и данных в нем:

- защита image registry для контейнеров/источника установки образа для ВМ
- поиск конфиденциальной информации и секретов в слоях образа контейнера/проверка предсобранного образа ВМ
- подпись и шифрование образа

Защита данных обрабатываемых в образе:

- изоляция одного контейнера от другого/изоляция по умолчанию в ВМ
- все доступно с хостовой ОС/все доступно с гипервизора

Защита/изоляция сети

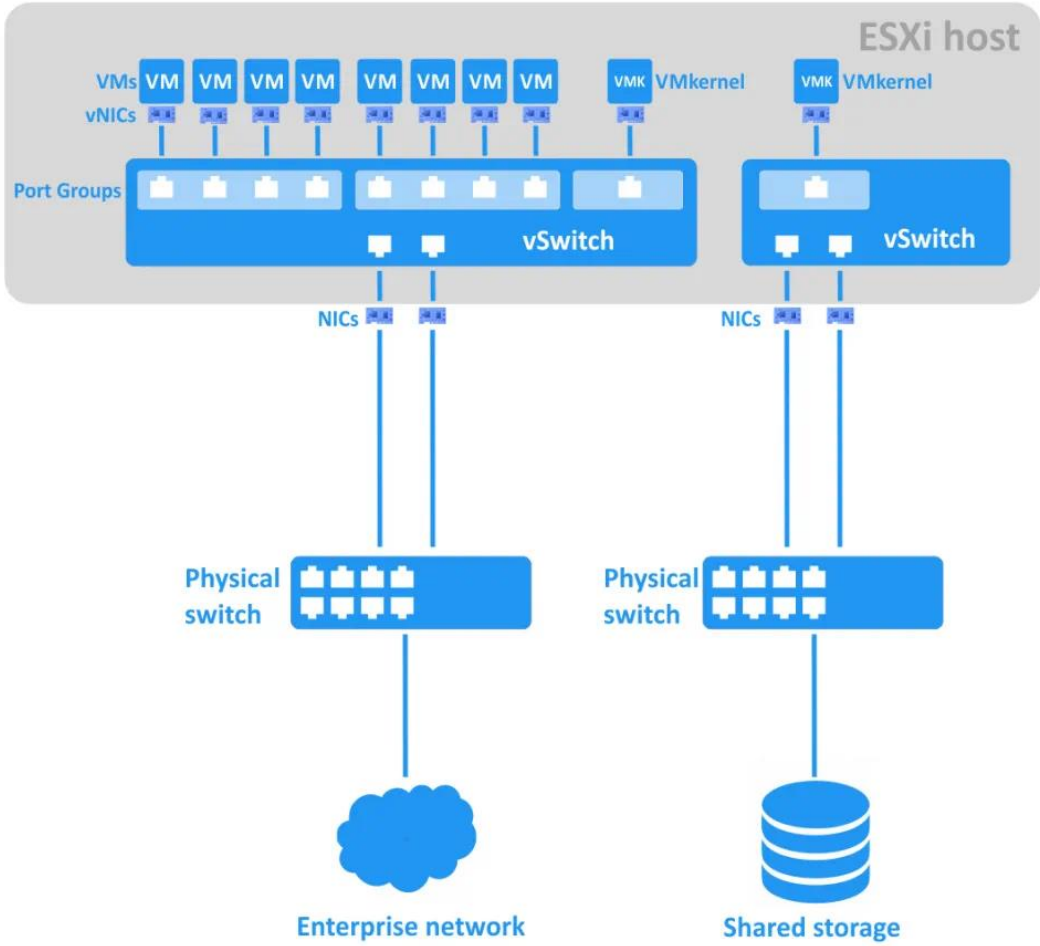
Industrial++ 2024

ВМ внутри гипервизора

- Полная аналогия с физическим сервером – Bridge/PrivateBridge, NAT, localhost
- Как правило, адреса и DNS статичны
- Защита уровня хоста/сети аналогична любому другому физическому устройству
- Шифрование – уровня бизнес-приложений либо на штатных механизмах гостевой ОС

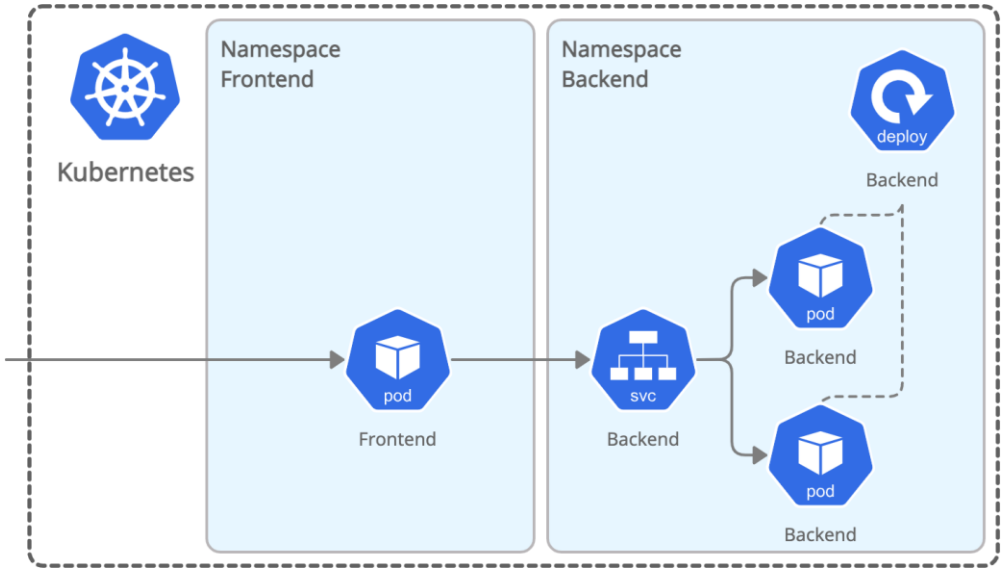
Сегментация

- Виртуальные свитчи/роутеры
- Виртуальное вендорское сетевое оборудование (но не для всех)
- Физическое оборудование за пределами гипервизора
- Можно выбирать, где будет происходить терминация vLAN (VST-VGT-EST / dom0) при использовании



Сеть в Kubernetes для начинающих

- Сеть на базе CNI-плагинов
- Под капотом: iptables, nftables, eBPF, ...
- Сеть плоская
- У всех Pods есть IP
- PodCIDR на каждой Node
- Service для load-balancing
- DNS для service-discovery



IP-адрес меняется/переходит от запуска к запуску микросервиса!

Сетевая безопасность: сегментация

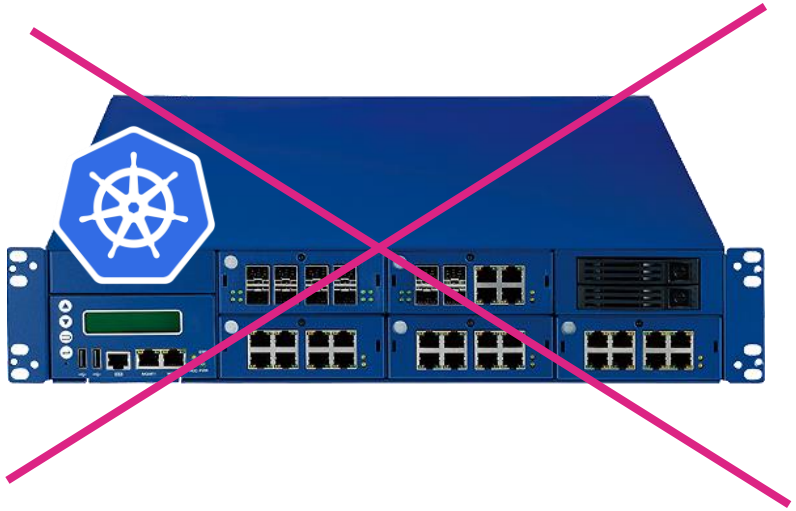
2 способа

CNI NetworkPolicy
(родной межсетевой экран Kubernetes)

- Native – формат политик от Kubernetes
- Custom – расширенный формат политик от разработчиков CNI

Service Mesh

- Sidecar proxy (service proxy)
- Shared proxy per node
- Shared proxy per service account (per node)
- Shared remote proxy with micro proxy
- eBPF Accelerated Per-Node Proxy
- Hybrid



Сетевая безопасность: контроль входящего и исходящего трафика

Для входящего трафика:

- Host-Based Firewall
- Ingress Gateway
- API Gateway

Для исходящего трафика:

- Host-Based Firewall
- CNI NetworkPolicy (Native, Custom)
- Service Mesh
- Конфигурацией NAT
- С помощью Egress Gateways

Сетевая безопасность: шифрование

3 способа

На уровне CNI-плагинов с помощью:

- IPSec
- Wireguard

На уровне Service Mesh с помощью:

- mTLS

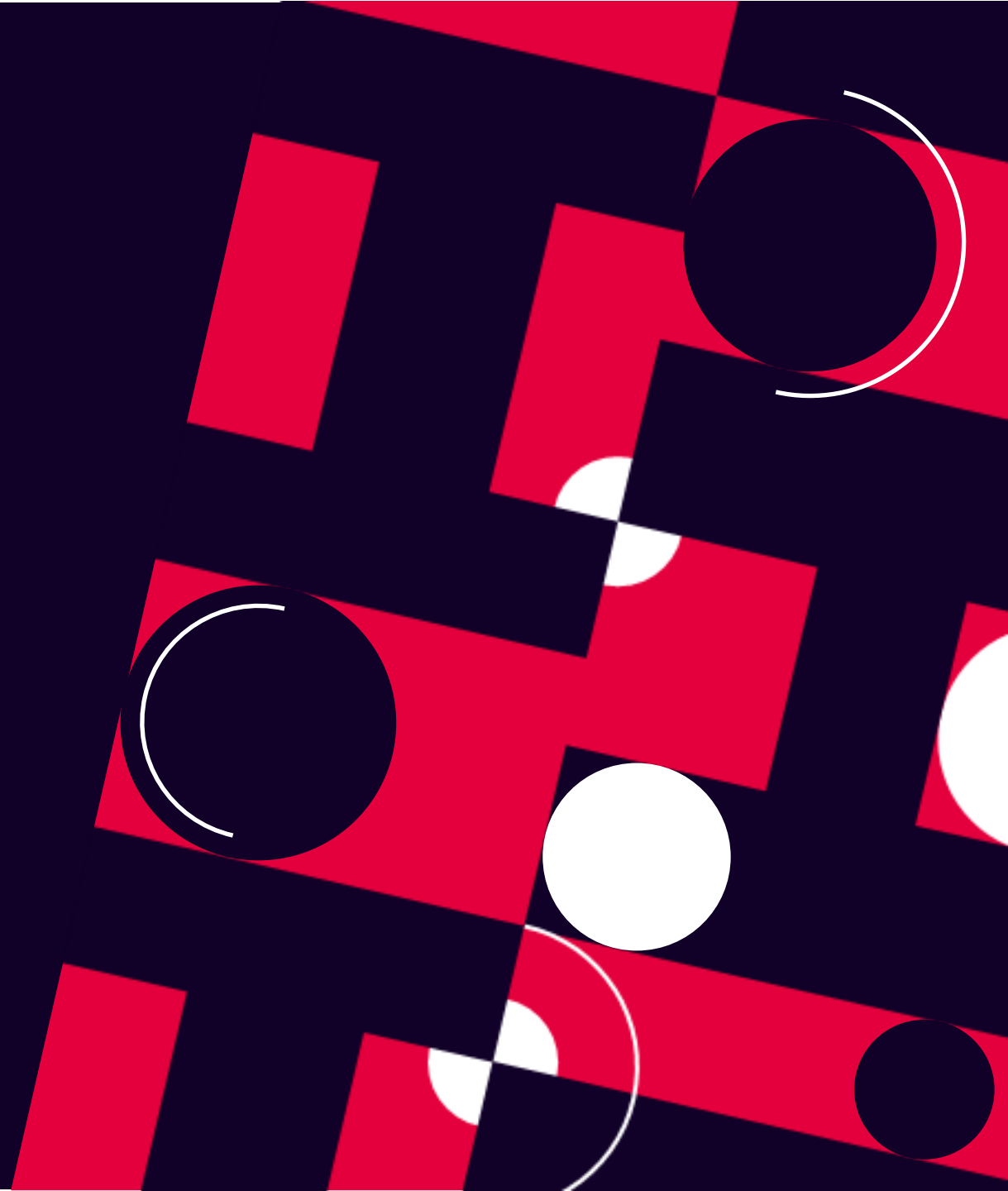
На уровне ваших приложений с помощью:

- SSL/TLS-протоколов

Кардинально разные на уровне сети

- Гранулярность доступа на стороне контейнеров
- Гибкость на стороне контейнеров
- Расширяемость на стороне контейнеров

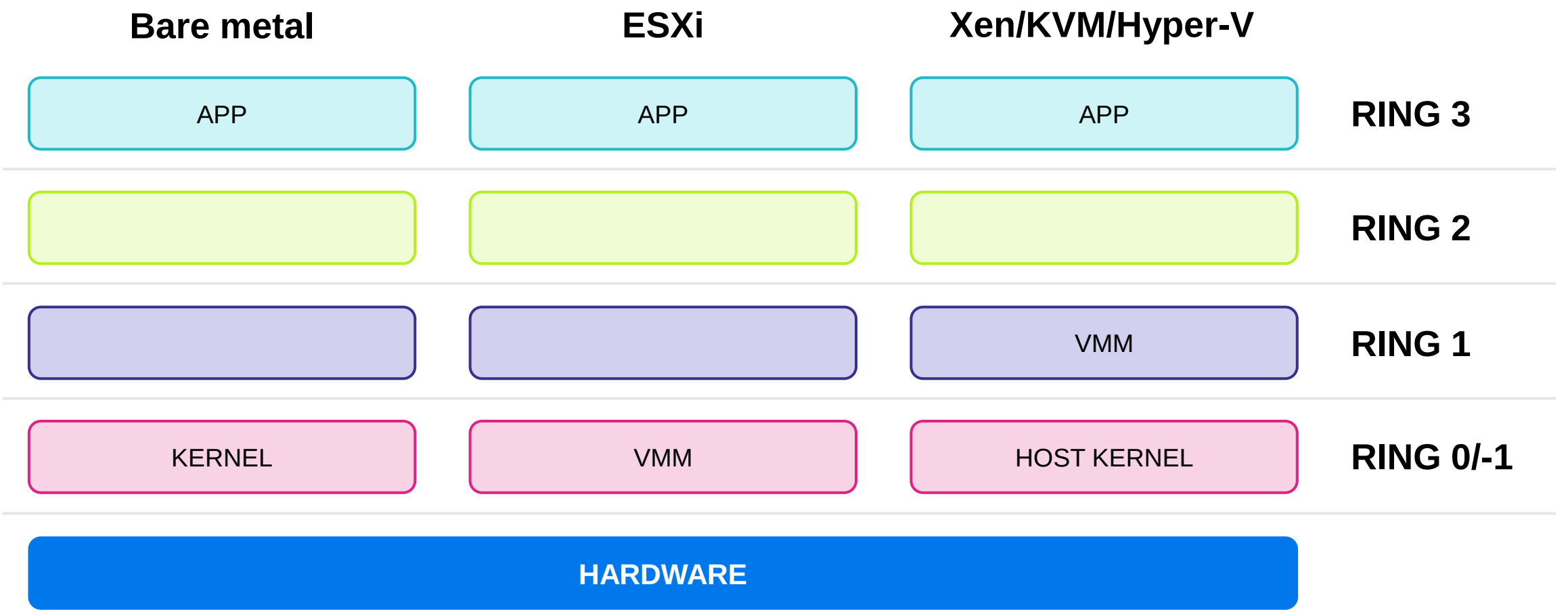
Защита/изоляция среды выполнения (runtime)



Industrial++ 2024

Гипервизор и ВМ

Стык аппаратной и программной виртуализации



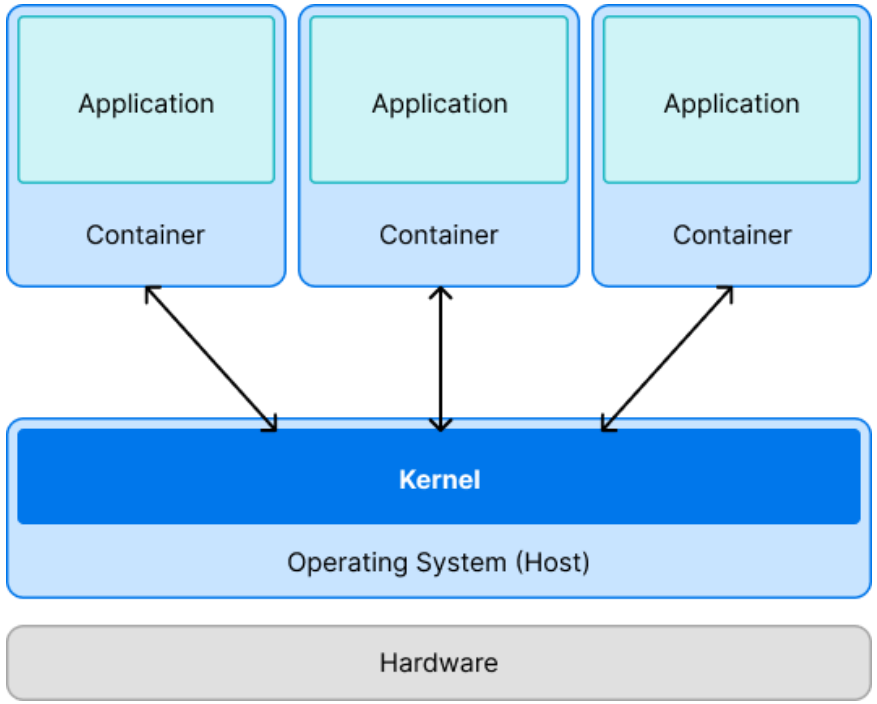
Контейнер и ядро хостовой ОС

В нативной контейнеризации системные вызовы обрабатываются в ядре хостовой ОС:

- runc (Go), crun (C), youki (Rust)

Эксплуатация уязвимости в ядре хостовой ОС приводит к побегу из контейнера:

- CVE-2023-3389
- CVE-2023-0461
- CVE-2022-0847
- …



Защита

Необходимо:

- усложнить запуск кода атакующего
- уменьшить поверхность атаки
- обновлять ядро хостовой ОС

Механизмы и подходы защиты:

- Seccomp-профиль
- AppArmor-профиль
- SeLinux-профиль
- SecurityContext
- Rootless-контейнеры
- Tiny/Slim/Distroless-образы
- Специализированные OS
- Альтернативные Runtimes

Применимо к native containers

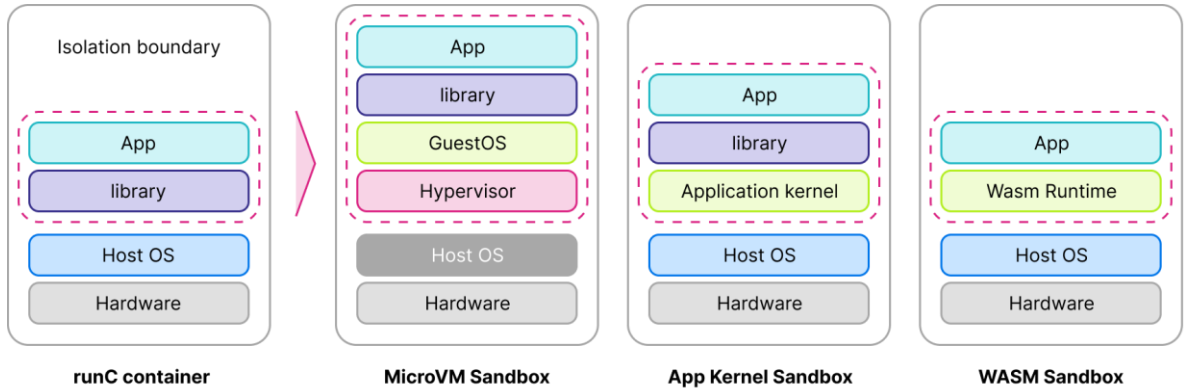
Альтернативные runtimes

WASM
- WasmEdge, Wasmtime, Wasmer, …

Sandbox/App kernel
- gVisor, Quark

MicroVM
- Kata containers

VM
- kubeVirt




```yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  labels:
    app: gvisor-nginx
  name: gvisor-nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      app: gvisor-nginx
  template:
    metadata:
      labels:
        app: gvisor-nginx
    spec:
      runtimeClassName: gvisor
      topologySpreadConstraints:
      - maxSkew: 1
        topologyKey: kubernetes.io/hostname
        whenUnsatisfiable: DoNotSchedule
        labelSelector:
          matchLabels:
            app: gvisor-nginx
      containers:
        - name: nginx
          image: nginx
          imagePullPolicy: IfNotPresent
      restartPolicy: Always
```

Подробнее: Кубик Runtime в конструкторе Kubernetes для безопасности

Сложность нарушения изоляции

Targets:

Target	Prize	Master of Pwn Points	Eligible for Add-on Prize
Oracle VirtualBox	$40,000	4	Yes
VMware Workstation	$80,000	8	Yes
VMware ESXi	$150,000	15	No
Microsoft Hyper-V Client	$250,000	25	Yes

Targets:

Target	Prize	Master of Pwn Points
containerd	$60,000	6
Docker Desktop	$60,000	6
Firecracker	$60,000	6
gRPC	$30,000	3

Источник: Данные с соревнований pwn2own 2024

Все перемешалось

- Гипервизоры ВМ предоставляют более сильную изоляцию по сравнению с контейнерами
- Контейнеры могут использовать множество механизмов безопасности
- Kubernetes-оркестратор позволяет жонглировать runtimes для контейнеров

Обслуживание

Industrial++
2024

Рутинные операции

- Обновления безопасности системы
- Развертывание общих сервисов
- Выкатка релизов собственных разработок
- Балансировка/масштабирование
- Capacity planning

Уровень подготовки

- Уровень погружения (администратор приложения)
- А если не облако?
- Проблемы, о которых не написано в поисковиках
- Обслуживание low-level

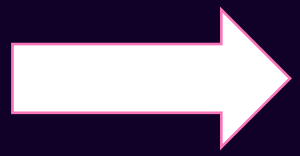
Заключение

Industrial++
2024

Итог

- ВМ и контейнеры имеют сильные и слабые стороны
- ВМ и контейнеры могут и хорошо уживаются вместе
- Оркестраторы стирают грань между ними
- Используйте каждое с умом
- Новые абстракции – это новые задачи и новые возможности
- Не бойтесь новых технологий: все новое – это хорошо забытое старое ;)

Спасибо за внимание!



Дмитрий Евдокимов

de@luntry.ru
@Qu3b3c

Мона Архипова

mona.arkhipova@gmail.com
@Mona_Sax

LUNTRY

Industrial++ 2024