Les réseaux de conteneurs
Fonctionnalité & défis

12 Décembre 2024

Cloud Native Lorient

Qui suis-je

Alexis La Goutte

Consultant réseaux & Sécurité

- (Net|Secu)Ops qui aide les Ops & qui fait du code depuis plus 15 ans
- Papa d'un cluster de 3 filles
- Contributeur à (trop) plein de projet OSS
- Certifié K8S / NSX / Cilium...
- A@alagoutte

Agenda

- Virtualisation du réseau
- Load Balancer
- Pare-feu/Firewall
- Introduction à Cilium

3

Et vous ?

- Qui est dev ?
- Qui est Ops ?
- Qui est Net|SecuOps ?

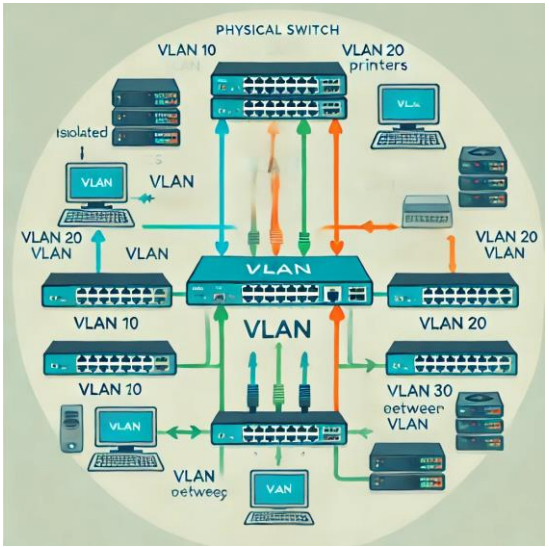
Virtualisation du réseau

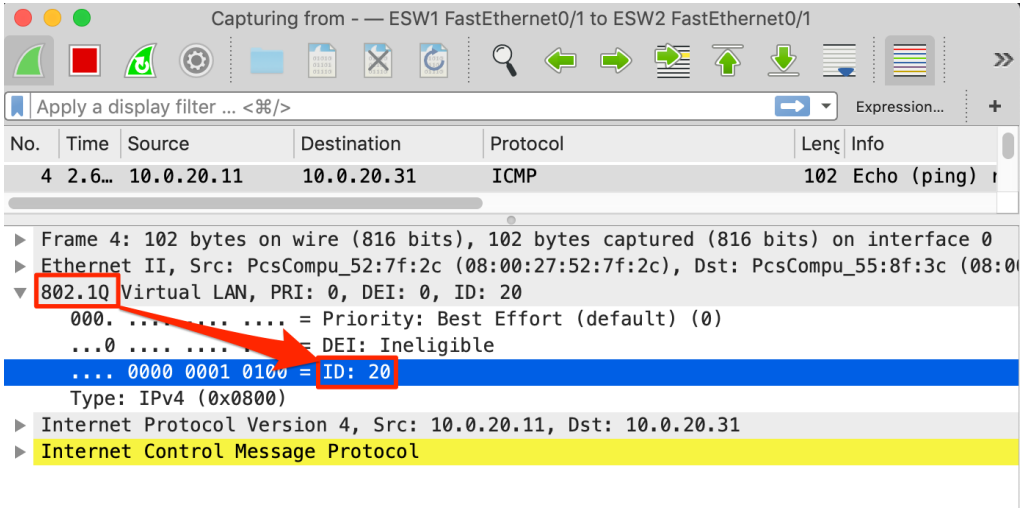
Virtualisation du réseau (Histoire)

- Vlan
- SDN « Hardware » (VXLAN/BGP/EVPN)
- SDN « Logiciel » (NSX)

Vlan (alias 802.1Q)

- Permet la séparation « logique » des réseaux IP via un « Tag »





7

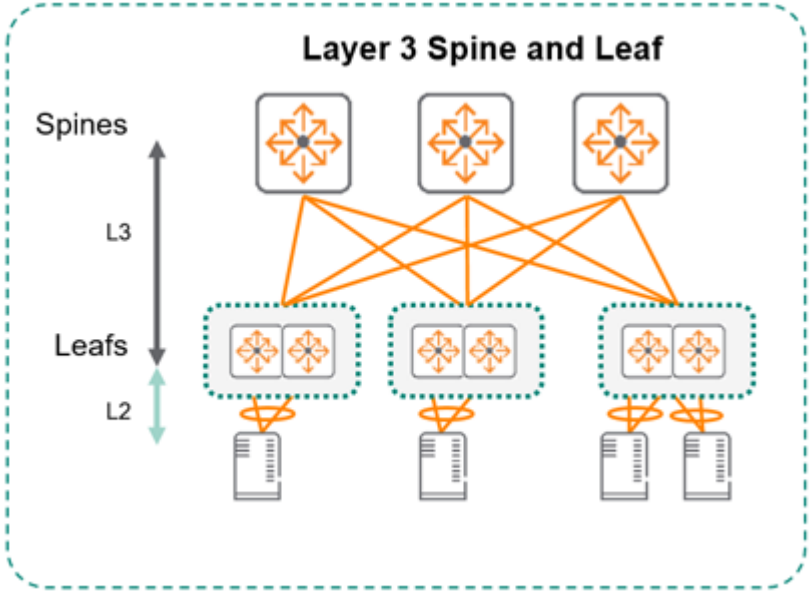
SDN « Hardware » (VXLAN/BGP/EVPN)

- Fabric **Leaf/Spine**

  avec du routage L3 (underlay),
  des interfaces TEP

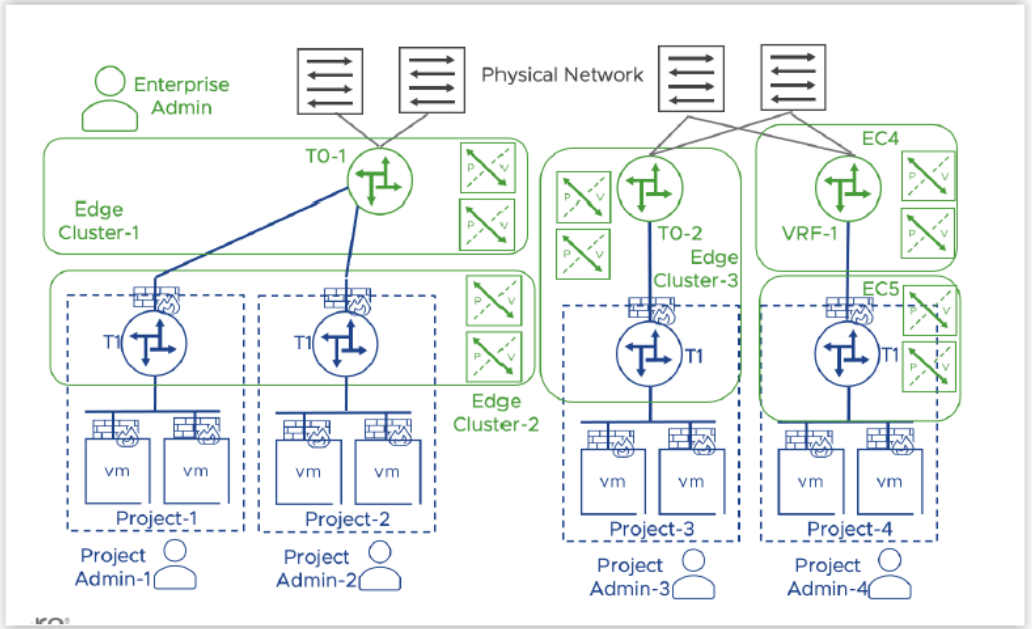
  et du BGP EVPN.. (pour échanger les table ARP) avec Route Reflector

  Pour les **GROS** réseaux (DC...)



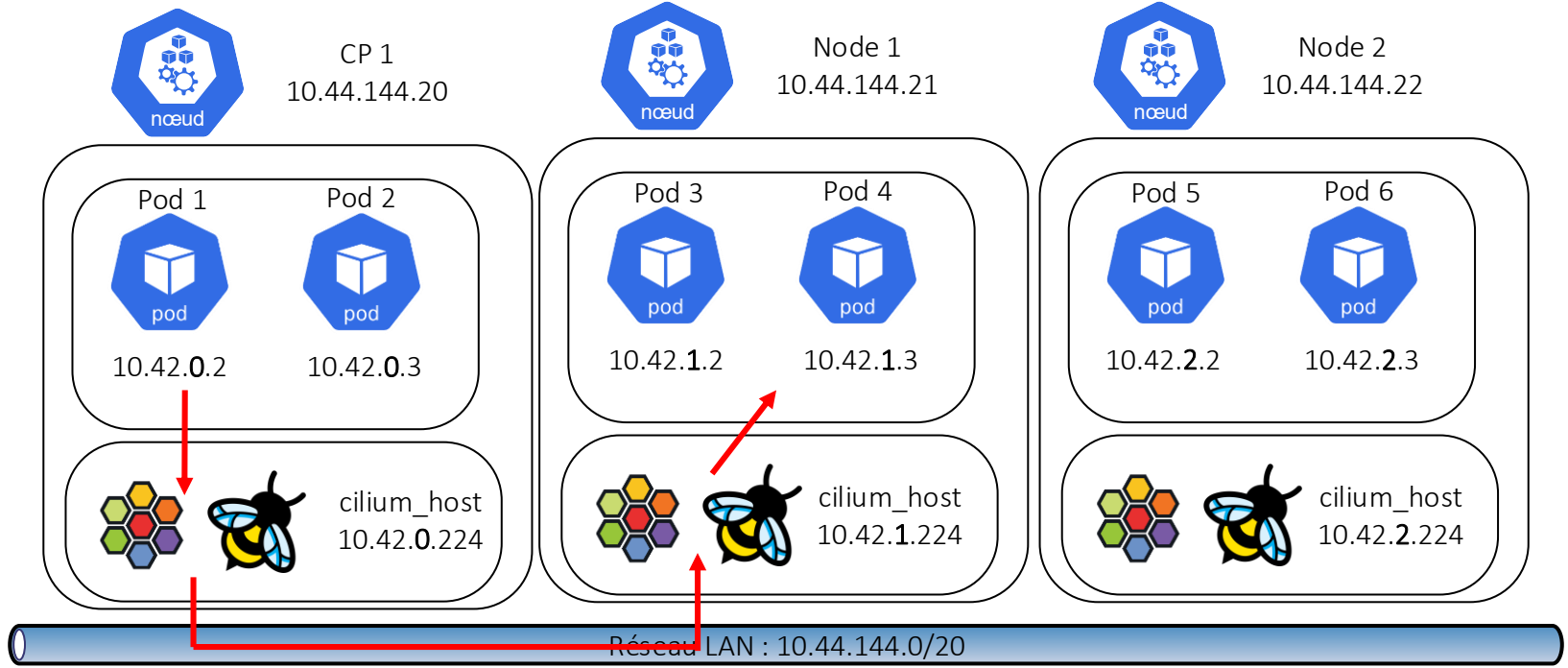
NSX (by VMware)

- Virtualisation du réseau logiciel (dans les Hyperviseurs)
- Routeur distribué directement dans les hosts
- VM (Edge T0/T1) pour la connexion au monde physique



K8S : Virtualisation by Design

- Chaque node est un routeur avec une plage IP allouée



K8S : Virtualisation by Design (cluster-cidr)

```
alagoutte@ALG-RKE2-CILIUM-CP:~$ kubectl exec -it -n kube-system cilium-kkrxj -c cilium-agent -- cilium-dbg bpf tunnel lis
TUNNEL       VALUE
10.42.3.0    10.44.144.23:0
10.42.4.0    10.44.144.24:0
10.42.2.0    10.44.144.22:0
10.42.1.0    10.44.144.21:0
alagoutte@ALG-RKE2-CILIUM-CP:~$
```

```
alagoutte@ALG-RKE2-CILIUM-CP:~$ ip route
default via 10.44.155.254 dev ens33 proto static
10.42.0.0/24 via 10.42.0.244 dev cilium_host proto kernel src 10.42.0.244
10.42.0.244 dev cilium_host proto kernel scope link
10.42.1.0/24 via 10.42.0.244 dev cilium_host proto kernel src 10.42.0.244 mtu 1450
10.42.2.0/24 via 10.42.0.244 dev cilium_host proto kernel src 10.42.0.244 mtu 1450
10.42.3.0/24 via 10.42.0.244 dev cilium_host proto kernel src 10.42.0.244 mtu 1450
10.42.4.0/24 via 10.42.0.244 dev cilium_host proto kernel src 10.42.0.244 mtu 1450
10.44.144.0/20 dev ens33 proto kernel scope link src 10.44.144.20
```

```
cilium:
  ipam:
    mode: cluster-pool
    operator:
      clusterPoolIPv4MaskSize: 24
      clusterPoolIPv4PodCIDRList:
        - 10.44.0.0/16
```

Load Balancing

Load Balancing (ou Reverse Proxy ?)

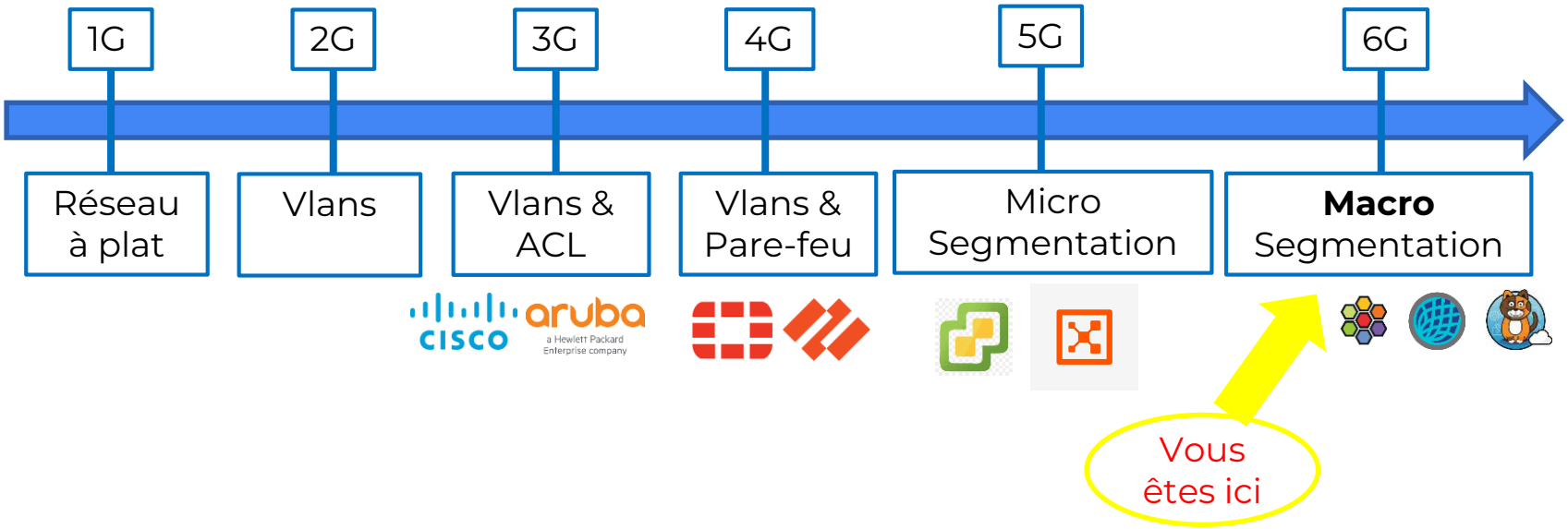
- Hardware
  - F5 Big-IP, Citrix ADC / NetScaler...

- Software
  - HAProxy, Nginx, Traefik

- Cloud
  - AWS ELB, Azure Load Balancer, GCP Load Balancing

Pare-feu

Etat des Lieux



K8S : Network Policy

- Policy As Code
- Applicable sur :
  - Flux entrant (Ingress)
  - Flux Sortant (Egress)
- Limité au L4

```yaml
apiVersion: networking.k8s.io/v1
kind: NetworkPolicy
metadata:
  name: test-network-policy
  namespace: default
spec:
  podSelector:
    matchLabels:
      role: db
  policyTypes:
  - Ingress
  - Egress
  ingress:
  - from:
    - ipBlock:
        cidr: 172.17.0.0/16
        except:
        - 172.17.1.0/24
    - namespaceSelector:
        matchLabels:
          project: myproject
    - podSelector:
        matchLabels:
          role: frontend
    ports:
    - protocol: TCP
      port: 6379
  egress:
  - to:
    - ipBlock:
        cidr: 10.0.0.0/24
    ports:
    - protocol: TCP
      port: 5978
```


16

CILIUM pour les nuls

Cilium pour les nuls

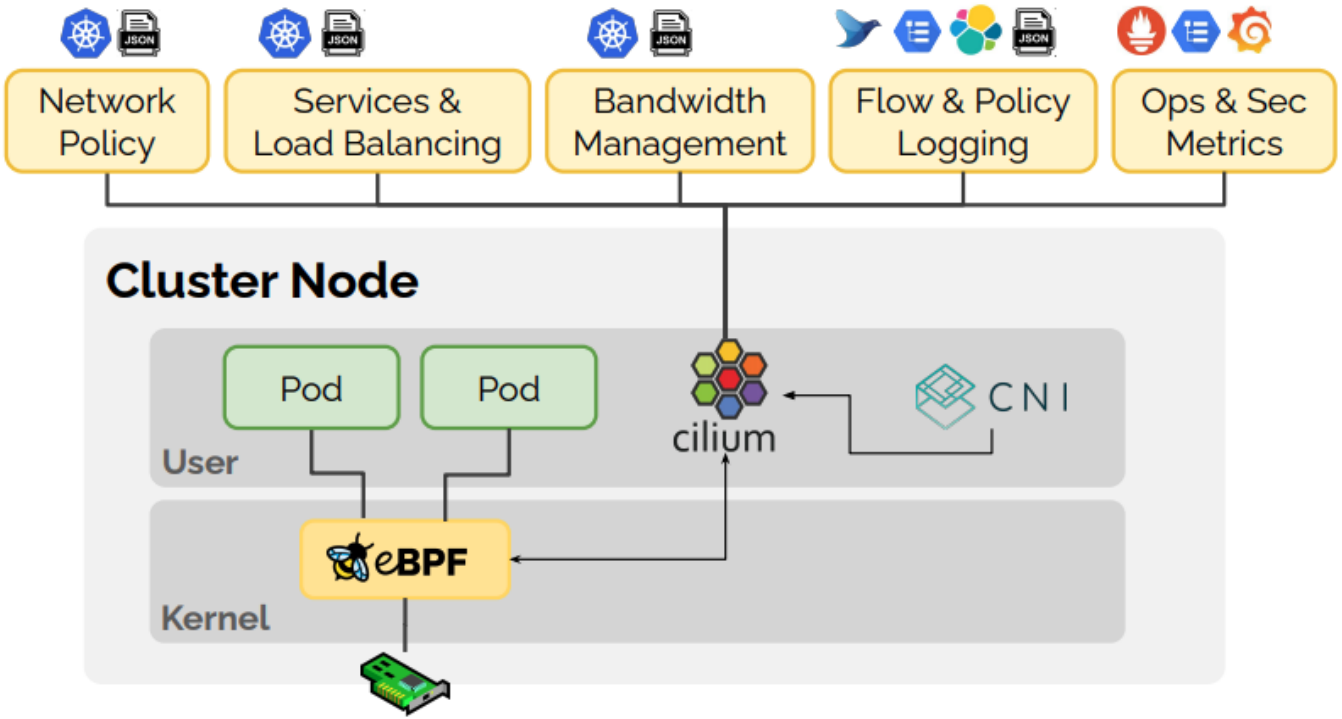
- Cilium est une CNI (Container Network Interface) qui permet :
  - **La sécurisation des microservices** : Définir des règles de sécurité (L4 à L7) granulaires entre les services.
  - **La performance réseau optimisée** : Utilisation d'**eBPF** pour minimiser les latences et la consommation de ressources.
  - **L'observabilité des flux réseau** : Utilisation d'**Hubble** pour comprendre le comportement réseau dans un cluster Kubernetes.

- C'est la seule CNI "Graduated" à la CNCF; elle est utilisée chez des
  - Cloud public (Google / GKE via Anthos), Azure (AKS), Amazon (EKS-A)
  - Cloud Privées : Adobe, Bell, Datadog, Ikea, Deezer...

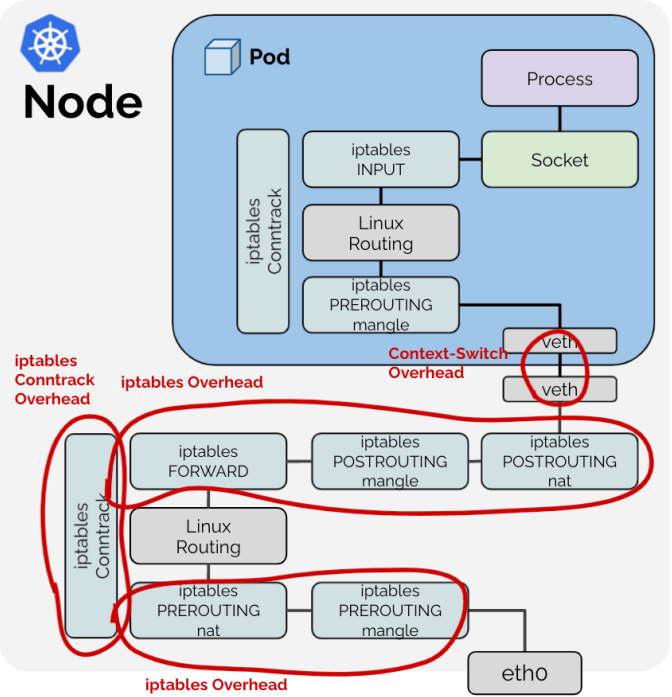
HUBBLE

18

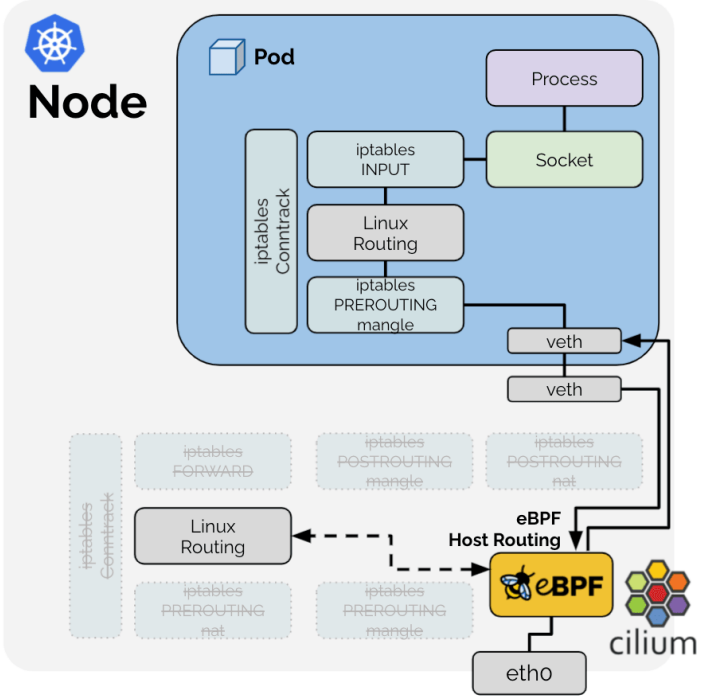
Cilium



Cilium : Inside

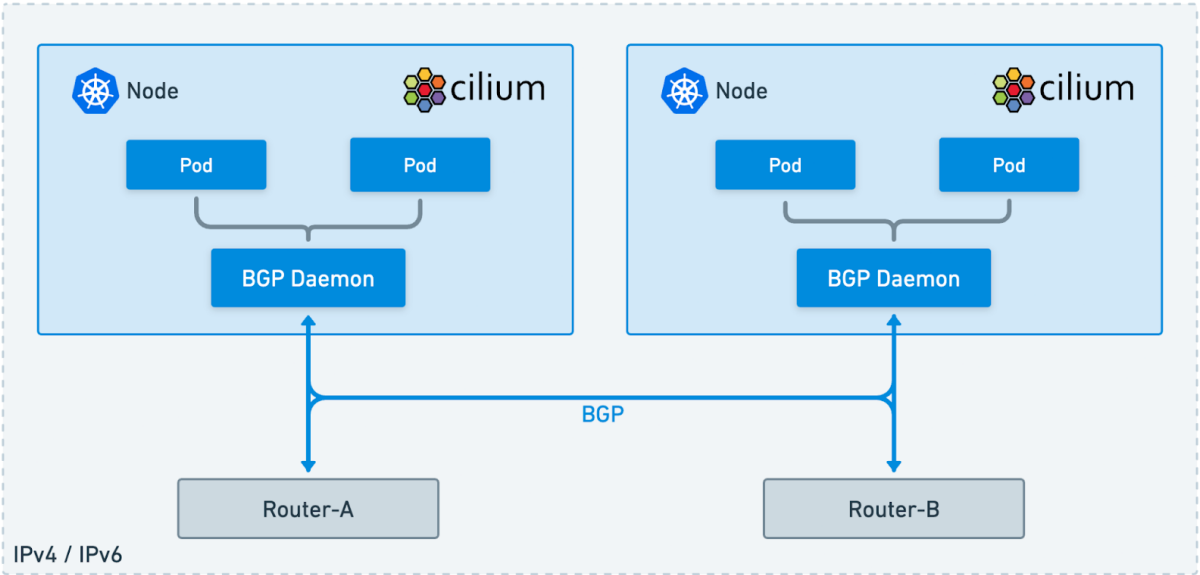


Standard Container Networking



Cilium eBPF Container Networking

Load Balancing : Cilium avec BGP 1/3



Load Balancing : Cilium avec BGP 2/3

```yaml
apiVersion: "cilium.io/v2alpha1"
kind: CiliumBGPPeeringPolicy
metadata:
  name: blue-peering-policy
spec:
  nodeSelector:
    matchLabels:
      bgp-policy: blue
  virtualRouters:
  - localASN: 64512
  exportPodCIDR: true
  neighbors:
  - peerAddress: '10.0.0.1'
    peerASN: 64512
  - peerAddress: '10.0.0.2'
    peerASN: 64512
```

```yaml
apiVersion: "cilium.io/v2alpha1"
kind: CiliumLoadBalancerIPPool
metadata:
  name: "pool-blue"
spec:
  blocks:
    - cidr: "192.0.2.0/24"
  serviceSelector:
    matchLabels:
      color: blue
```

```yaml
apiVersion: v1
kind: Service
metadata:
  name: service-blue
  namespace: blue
  labels:
    color: blue
spec:
  type: LoadBalancer
  ports:
  - port: 80
```


Load Balancing : Cilium avec BGP 3/3

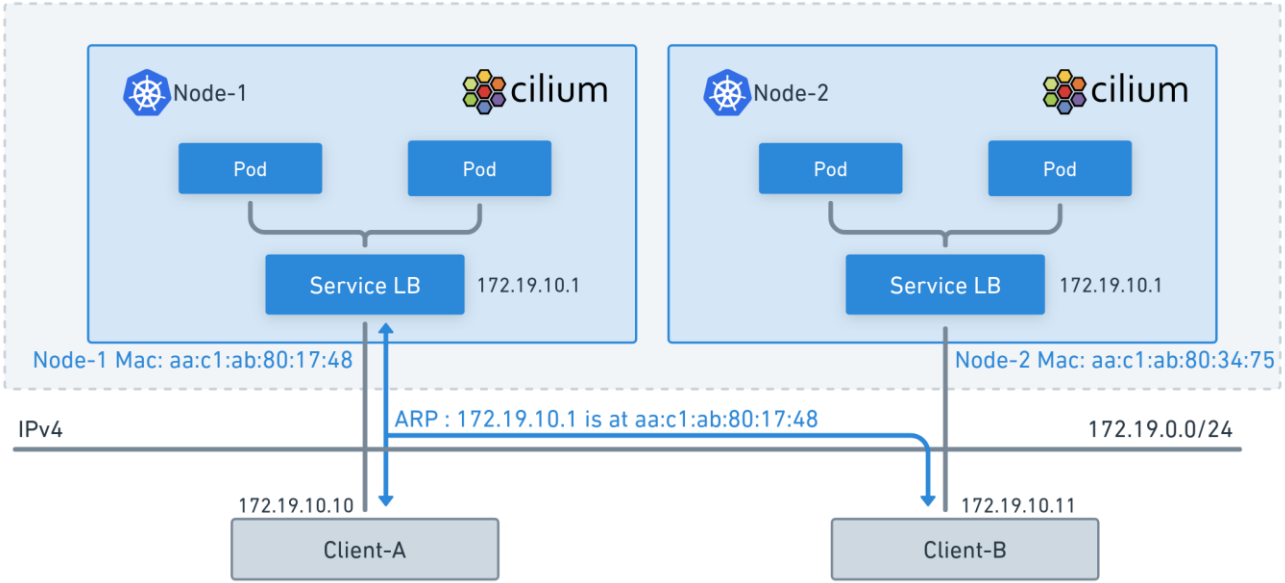
```
root@server:~# kubectl get svc -n blue
NAME           TYPE           CLUSTER-IP     EXTERNAL-IP   PORT(S)        AGE
service-blue   LoadBalancer   10.96.243.64   192.0.2.0     80:32527/TCP   3m17s
```

```
root@server:~# docker exec -it clab-bgp-cplane-devel-tor vtysh -c 'show bgp ipv4'
BGP table version is 9, local router ID is 172.0.0.1, vrf id 0
Default local pref 100, local AS 65000
Status codes:  s suppressed, d damped, h history, * valid, > best, = multipath,
               i internal, r RIB-failure, S Stale, R Removed
Nexthop codes: @NNN nexthop's vrf id, < announce-nh-self
Origin codes:  i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

   Network          Next Hop            Metric LocPrf Weight Path
*> 192.0.2.0/32     172.0.0.2                              0 65001 i

Displayed  1 routes and 1 total paths
root@server:~#
```

Load Balancing: Cilium avec mon L2: 1/3



Load Balancing: Cilium avec mon L2 2/3

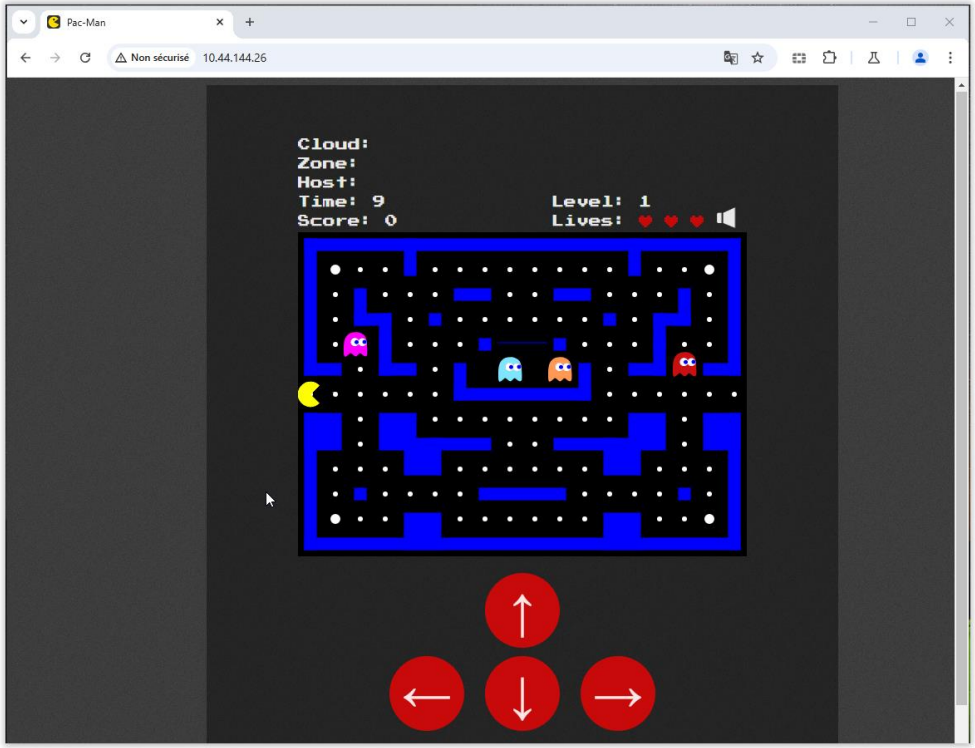
```
kubeProxyReplacement: strict
l2announcements:
  enabled: true
devices: {eth0, net0}
externalIPs:
  enabled: true
```

```
apiVersion: "cilium.io/v2alpha1"
kind: CiliumL2AnnouncementPolicy
metadata:
  name: cilium-lb-all-services
  namespace: kube-system
spec:
  loadBalancerIPs: true
```

```
apiVersion: "cilium.io/v2alpha1"
kind: CiliumLoadBalancerIPPool
metadata:
  name: cilium-lb-ipam
  namespace: kube-system
spec:
  blocks:
    - start: "10.44.144.26"
      stop: "10.44.144.29"
```

```
alagoutte@ALG-RKE2-CILIUM-CP:~$ kubectl get svc -n pacman
NAME      TYPE           CLUSTER-IP      EXTERNAL-IP    PORT(S)        AGE
mongo     ClusterIP      10.43.121.148   <none>         27017/TCP      44d
pacman    LoadBalancer   10.43.246.232   10.44.144.26   80:30398/TCP   44d
```

Cilium avec mon L2 3/3



Cilium : Macro Segmentation

- Utilisation de Cilium Network Policy
  Possibilité de faire des « policy » L7 HTTP/DNS....

```yaml
apiVersion: cilium.io/v2
kind: CiliumNetworkPolicy
metadata:
  namespace: default
  name: dns
spec:
  endpointSelector: {}
  egress:
    - toEndpoints:
        - matchLabels:
            io.kubernetes.pod.namespace: kube-system
            k8s-app: kube-dns
      toPorts:
        - ports:
            - port: "53"
              protocol: ANY
          rules:
            dns:
              - matchPattern: "*"
```

```yaml
apiVersion: "cilium.io/v2"
kind: CiliumNetworkPolicy
metadata:
  name: "rule1"
spec:
  description: "L7 policy to restrict access to specific HTTP call"
  endpointSelector:
    matchLabels:
      org: empire
      class: deathstar
  ingress:
    - fromEndpoints:
        - matchLabels:
            org: empire
      toPorts:
        - ports:
            - port: "80"
              protocol: TCP
          rules:
            http:
              - method: "POST"
                path: "/v1/request-landing"
```


Cilium visiblité (hubble)

What you can't do with network policies (at least, not yet)

As of Kubernetes 1.31, the following functionality does not exist in the NetworkPolicy API, but you

- The ability to log network security events (for example connections that are blocked or accepted).

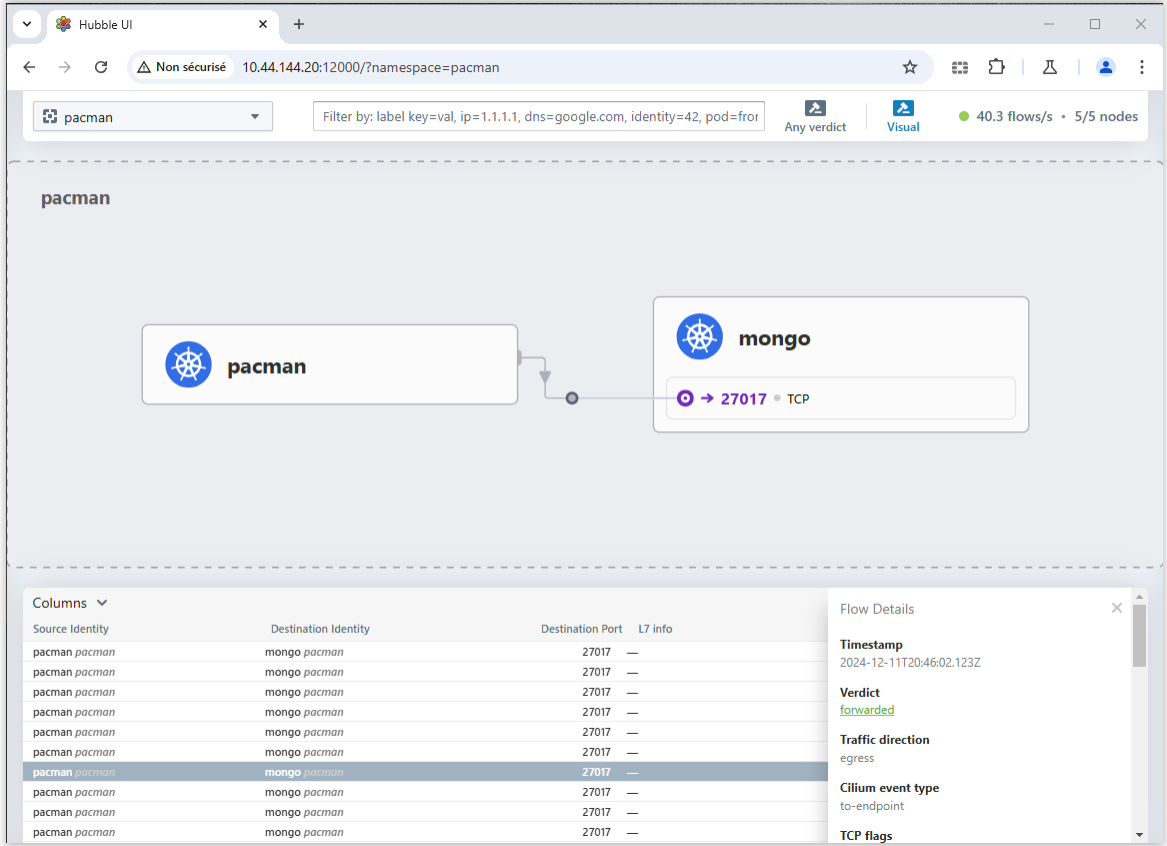
Cilium visibilité (CLI)

 # hubble observe

```
alagoutte@ALG-RKE2-CILIUM-CP:~$ hubble status
Healthcheck (via localhost:4245): Ok
Current/Max Flows: 20,475/20,475 (100.00%)
Flows/s: 48.32
Connected Nodes: 5/5
alagoutte@ALG-RKE2-CILIUM-CP:~$ hubble observe -n pacman
Dec 11 20:41:44.112: 10.42.2.188:54816 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: ACK, FIN)
Dec 11 20:41:46.977: pacman/pacman-599d78464b-jtk5q:38956 (ID:9421) -> pacman/mongo-84cd97647c-whp5f:27017 (ID:19988) to-endpoint FORWARDED (TCP Flags: ACK, PSH)
Dec 11 20:41:46.977: pacman/pacman-599d78464b-jtk5q:38956 (ID:9421) <- pacman/mongo-84cd97647c-whp5f:27017 (ID:19988) to-endpoint FORWARDED (TCP Flags: ACK)
Dec 11 20:41:51.935: 127.0.0.1:33092 (world) <> pacman/mongo-84cd97647c-whp5f (ID:19988) pre-xlate-rev TRACED (TCP)
Dec 11 20:41:54.109: 10.42.2.188:50760 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: SYN)
Dec 11 20:41:54.109: 10.42.2.188:50760 (host) <- pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-stack FORWARDED (TCP Flags: SYN, ACK)
Dec 11 20:41:54.109: 10.42.2.188:50760 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: ACK)
Dec 11 20:41:54.109: 10.42.2.188:50760 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: ACK, PSH)
Dec 11 20:41:54.109: 10.42.2.188:50762 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: SYN)
Dec 11 20:41:54.109: 10.42.2.188:50762 (host) <- pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-stack FORWARDED (TCP Flags: SYN, ACK)
Dec 11 20:41:54.109: 10.42.2.188:50762 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: ACK)
Dec 11 20:41:54.109: 10.42.2.188:50762 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: ACK, PSH)
Dec 11 20:41:54.110: 10.42.2.188:50762 (host) <- pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-stack FORWARDED (TCP Flags: ACK, PSH)
Dec 11 20:41:54.110: 10.42.2.188:50760 (host) <- pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-stack FORWARDED (TCP Flags: ACK, PSH)
Dec 11 20:41:54.110: 10.42.2.188:50760 (host) <- pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-stack FORWARDED (TCP Flags: ACK, FIN)
Dec 11 20:41:54.111: 10.42.2.188:50762 (host) <- pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-stack FORWARDED (TCP Flags: ACK, FIN)
Dec 11 20:41:54.111: 10.42.2.188:50762 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: ACK, FIN)
Dec 11 20:41:54.111: 10.42.2.188:50760 (host) -> pacman/pacman-599d78464b-jtk5q:8080 (ID:9421) to-endpoint FORWARDED (TCP Flags: ACK, FIN)
Dec 11 20:41:56.981: pacman/pacman-599d78464b-jtk5q:38956 (ID:9421) -> pacman/mongo-84cd97647c-whp5f:27017 (ID:19988) to-endpoint FORWARDED (TCP Flags: ACK, PSH)
Dec 11 20:41:56.981: pacman/pacman-599d78464b-jtk5q:38956 (ID:9421) <- pacman/mongo-84cd97647c-whp5f:27017 (ID:19988) to-endpoint FORWARDED (TCP Flags: ACK)
```

Cilium visibilité (UI)

 # cilium hubble ui



HUBBLE

Cilium Enterprise avec Support

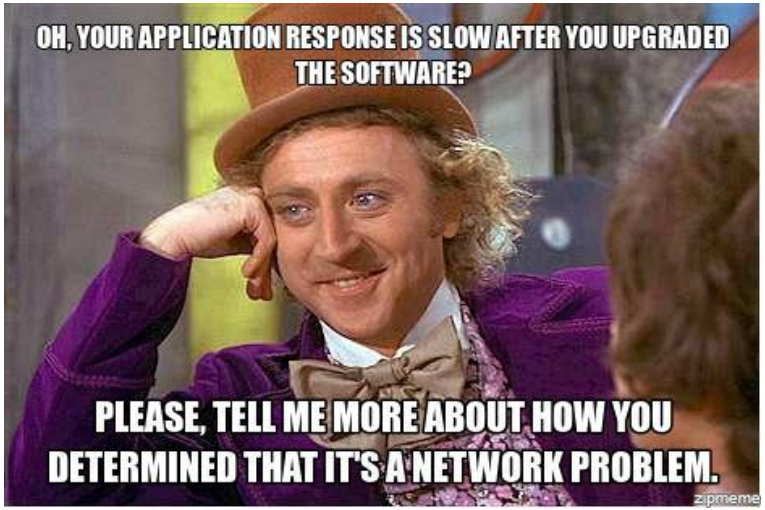
- Version "Enterprise" disponible
- Feature avancée
  - SR6
  - Egress HA
  - Advanced Policy Troubleshooting UI / Editor
  - Security Visibility and Enforcement via Tetragon (Support L7/TLS...)
  - BFD
  - ...
- Support 24x7 avec SLA
- ...

En attendant Noël

https://labs-map.isovalent.com/holidays/

Conclusion...

Conclusion....



- Aidez vos Net/SecOps

- Choisissiez une bonne CNI

Merci