

Peut-on savoir vers quel pod une requête précise est routée par un Service Kubernetes ?

Fonctionnement du load balancing Kubernetes

- Le routage des requêtes par un Service Kubernetes (ClusterIP, NodePort, LoadBalancer) repose sur des mécanismes internes (iptables, IPVS, etc.) gérés par kube-proxy.
- À chaque requête, le choix du pod cible parmi la liste des endpoints est fait **de façon dynamique** (souvent round-robin ou random selon la configuration du cluster)^{[1] [2]}.
- Ce choix s'effectue au niveau du nœud qui reçoit la requête, via la pile réseau du système, et **aucune trace native n'est conservée** sur le cheminement d'une requête donnée.

Peut-on interroger le load balancer pour connaître la cible d'une requête donnée ?

Non, il n'existe pas de commande ou d'API Kubernetes permettant de savoir, a posteriori ou en temps réel, vers quel pod précis une requête a été routée^{[1] [2]}.

Le load balancing est stateless: chaque requête entrante est traitée indépendamment, et le routage se fait à la volée selon les règles du moment.

Solutions pour tracer le routage réel d'une requête

Si vous souhaitez savoir quel pod a effectivement traité une requête :

- **Ajouter une information d'identification dans la réponse**
Modifiez votre application pour inclure le nom du pod (hostname, variable d'environnement, etc.) dans la réponse HTTP. Cela permet de voir immédiatement quel pod a répondu à chaque requête^[3].
- **Analyser les logs applicatifs**
Ajoutez un identifiant unique à chaque requête (par exemple un header HTTP ou un trace-id), puis consultez les logs des pods pour retrouver lequel l'a traitée^[3].
- **Utiliser un service mesh ou des outils d'observabilité**
Des solutions comme Istio, Linkerd, ou des outils de tracing (Jaeger, Zipkin) permettent de tracer le cheminement d'une requête à travers le cluster, mais nécessitent une instrumentation supplémentaire^[4].
- **Sniffer le trafic réseau**
En dernier recours, il est possible d'utiliser des outils comme tcpdump sur les nœuds pour observer le routage, mais cela reste complexe et peu recommandé en production^[4].

Pourquoi cette limitation ?

- Le load balancing Kubernetes est conçu pour la performance et la simplicité, pas pour la traçabilité au niveau de chaque requête.
- Le choix du pod cible est fait par des règles réseau (iptables/IPVS) qui ne conservent aucune trace du routage individuel^{[1] [2]}.

En résumé :

Il n'est pas possible d'interroger le load balancer Kubernetes pour savoir, pour une requête précise, quel pod a été ciblé. Pour le savoir, il faut instrumenter l'application, utiliser des headers ou des outils de tracing, ou consulter les logs applicatifs^{[1] [2] [3]}.

✱✱

1. <https://stackoverflow.com/questions/52745577/if-i-have-a-public-load-balancer-how-does-direct-service-to-service-communication-work>
2. <https://learnk8s.io/kubernetes-services-and-load-balancing>
3. <https://stackoverflow.com/questions/44550109/how-to-find-kubernetes-pod-that-is-handling-the-request/44551665>
4. https://www.reddit.com/r/kubernetes/comments/fxyr4j/how_to_sniff_all_pods_httphttps_requests_inside_a/