Building a Secure and Maintainable PaaS

eBPF Summit

Leveraging eBPF to Scale Security and Improve Platform Support

Bradley Whitfield

October 28, 2020

Dragon - Internal Platform as a Service

Requirements for Scaling Up

- ❑ Secure Network Isolation
- ❑ Network Visibility and Auditing
- ❑ Minimize maintenance and performance overhead
- ❑ Scale past iptables limits
- ❑ …

Network Security and Auditing

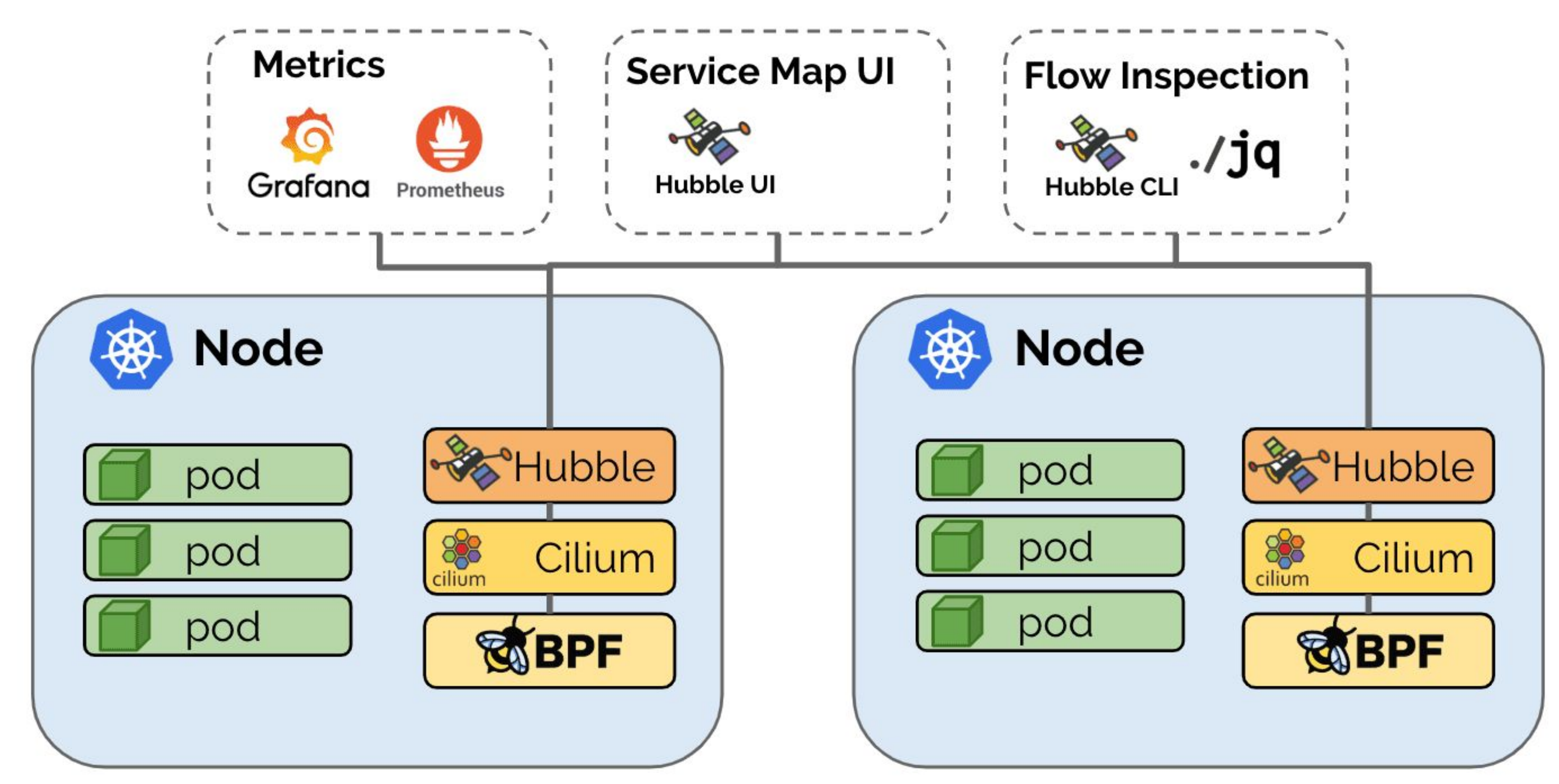
Scalability and Maintainability

Source: https://commons.wikimedia.org/wiki/File:Pictofigo-Scalability.png

Evaluating eBPF CNI Offerings

cilium

KubeCon
CloudNativeCon
North America 2019

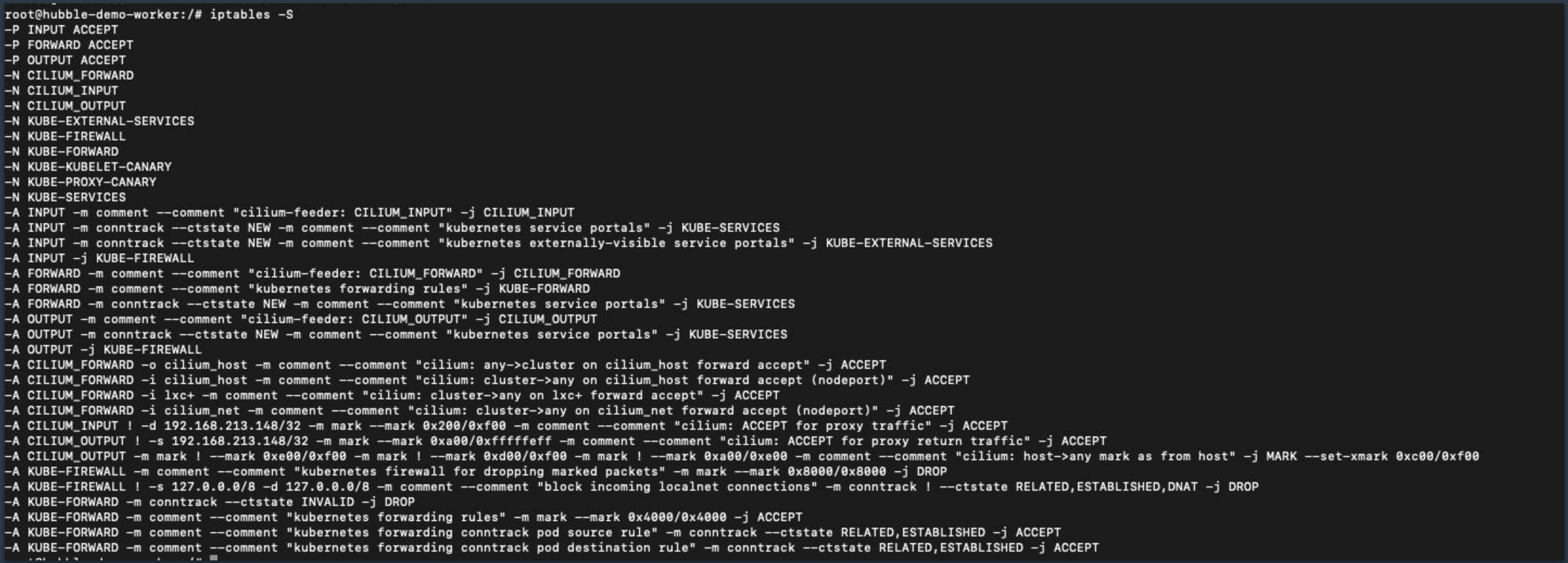

Metrics
Grafana
Prometheus
Service Map UI
Hubble UI
Flow Inspection
Hubble CLI
./jq
Node
pod
pod
pod
Hubble
cilium
Cilium
BPF
Node
pod
pod
pod
Hubble
cilium
Cilium
BPF

Evaluating Cilium and Hubble

Cilium Benefits

- ❑ Pod network filtering uses eBPF rather than iptables
- ❑ More flexible network policies
- ❑ Tools to help with network troubleshooting and policies
- ❑ Additional features like IPSec, Cluster Mesh, and more

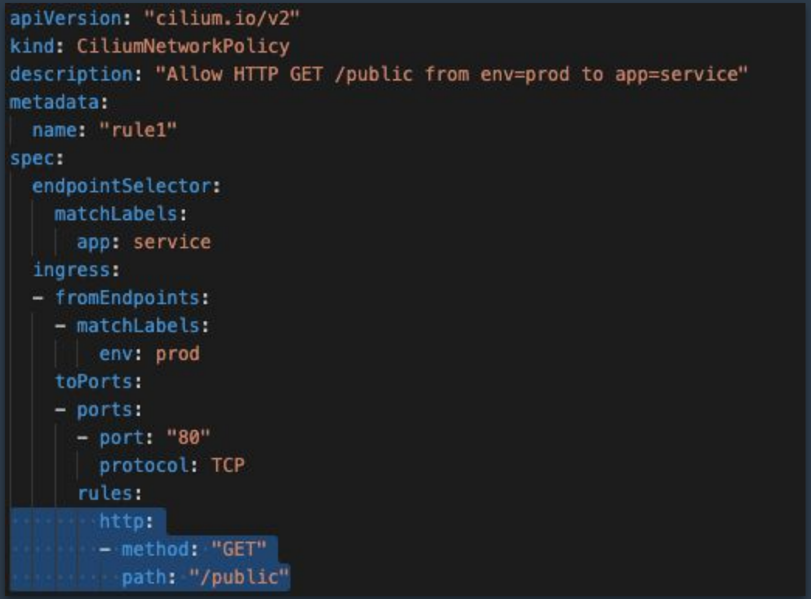
Reduced iptables Complexity



```
root@hubble-demo-worker:/# iptables -S
-P INPUT ACCEPT
-P FORWARD ACCEPT
-P OUTPUT ACCEPT
-N CILIUM_FORWARD
-N CILIUM_INPUT
-N CILIUM_OUTPUT
-N KUBE-EXTERNAL-SERVICES
-N KUBE-FIREWALL
-N KUBE-FORWARD
-N KUBE-KUBELET-CANARY
-N KUBE-PROXY-CANARY
-N KUBE-SERVICES
-A INPUT -m comment --comment "cilium-feeder: CILIUM_INPUT" -j CILIUM_INPUT
-A INPUT -m conntrack --ctstate NEW -m comment --comment "kubernetes service portals" -j KUBE-SERVICES
-A INPUT -m conntrack --ctstate NEW -m comment --comment "kubernetes externally-visible service portals" -j KUBE-EXTERNAL-SERVICES
-A INPUT -j KUBE-FIREWALL
-A FORWARD -m comment --comment "cilium-feeder: CILIUM_FORWARD" -j CILIUM_FORWARD
-A FORWARD -m comment --comment "kubernetes forwarding rules" -j KUBE-FORWARD
-A FORWARD -m conntrack --ctstate NEW -m comment --comment "kubernetes service portals" -j KUBE-SERVICES
-A OUTPUT -m comment --comment "cilium-feeder: CILIUM_OUTPUT" -j CILIUM_OUTPUT
-A OUTPUT -m conntrack --ctstate NEW -m comment --comment "kubernetes service portals" -j KUBE-SERVICES
-A OUTPUT -j KUBE-FIREWALL
-A CILIUM_FORWARD -o cilium_host -m comment --comment "cilium: any->cluster on cilium_host forward accept" -j ACCEPT
-A CILIUM_FORWARD -i cilium_host -m comment --comment "cilium: cluster->any on cilium_host forward accept (nodeport)" -j ACCEPT
-A CILIUM_FORWARD -i lxc+ -m comment --comment "cilium: cluster->any on lxc+ forward accept" -j ACCEPT
-A CILIUM_FORWARD -i cilium_net -m comment --comment "cilium: cluster->any on cilium_net forward accept (nodeport)" -j ACCEPT
-A CILIUM_INPUT ! -d 192.168.213.148/32 -m mark --mark 0x200/0xf00 -m comment --comment "cilium: ACCEPT for proxy traffic" -j ACCEPT
-A CILIUM_OUTPUT ! -s 192.168.213.148/32 -m mark --mark 0xa00/0xfffffeff -m comment --comment "cilium: ACCEPT for proxy return traffic" -j ACCEPT
-A CILIUM_OUTPUT -m mark ! --mark 0xe00/0xf00 -m mark ! --mark 0xd00/0xf00 -m mark ! --mark 0xa00/0xe00 -m comment --comment "cilium: host->any mark as from host" -j MARK --set-xmark 0xc00/0xf00
-A KUBE-FIREWALL -m comment --comment "kubernetes firewall for dropping marked packets" -m mark --mark 0x8000/0x8000 -j DROP
-A KUBE-FIREWALL ! -s 127.0.0.0/8 -d 127.0.0.0/8 -m comment --comment "block incoming localnet connections" -m conntrack ! --ctstate RELATED,ESTABLISHED,DNAT -j DROP
-A KUBE-FORWARD -m conntrack --ctstate INVALID -j DROP
-A KUBE-FORWARD -m comment --comment "kubernetes forwarding rules" -m mark --mark 0x4000/0x4000 -j ACCEPT
-A KUBE-FORWARD -m comment --comment "kubernetes forwarding conntrack pod source rule" -m conntrack --ctstate RELATED,ESTABLISHED -j ACCEPT
-A KUBE-FORWARD -m comment --comment "kubernetes forwarding conntrack pod destination rule" -m conntrack --ctstate RELATED,ESTABLISHED -j ACCEPT
```

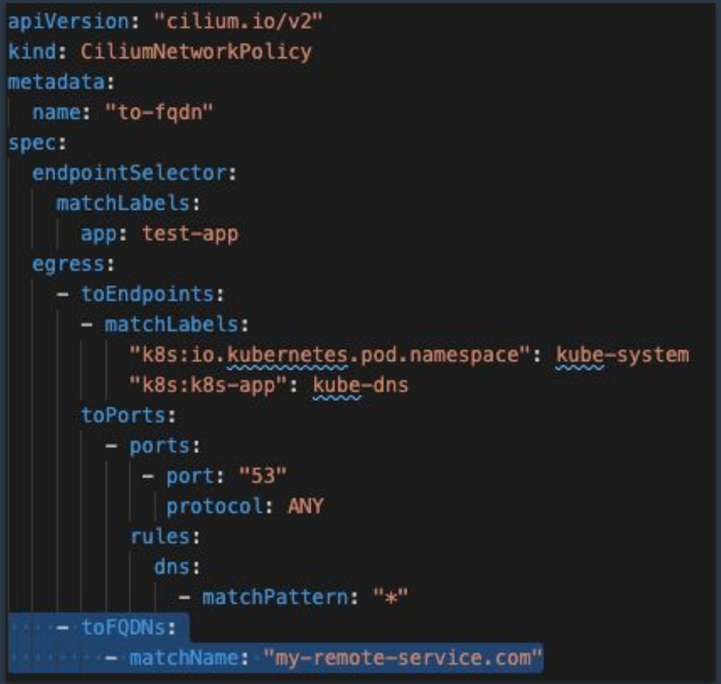
CiliumNetworkPolicies

Layer 7 HTTP Filtering



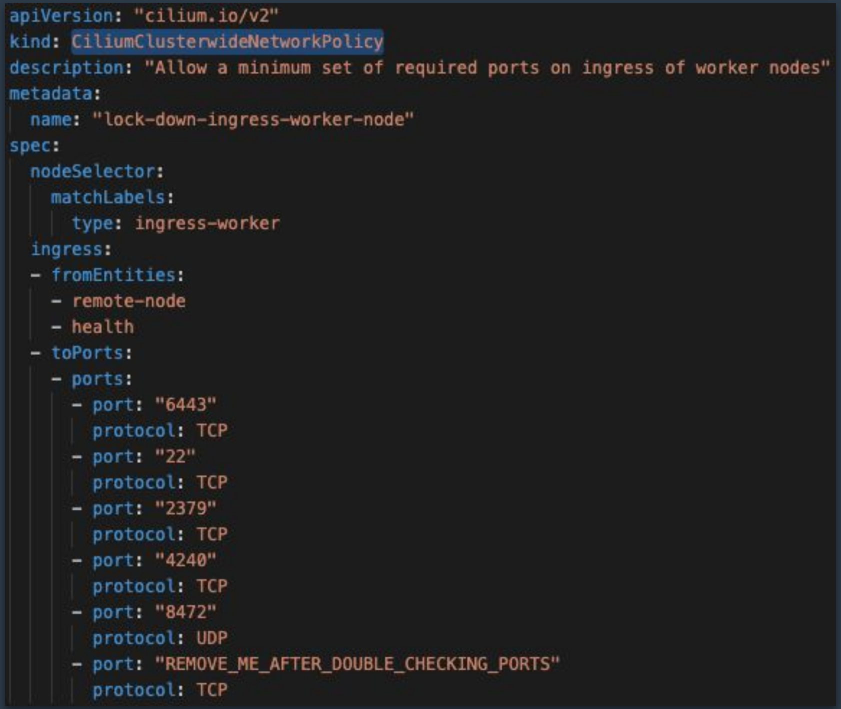
```yaml
apiVersion: "cilium.io/v2"
kind: CiliumNetworkPolicy
description: "Allow HTTP GET /public from env=prod to app=service"
metadata:
  name: "rule1"
spec:
  endpointSelector:
    matchLabels:
      app: service
  ingress:
  - fromEndpoints:
    - matchLabels:
        env: prod
    toPorts:
    - ports:
      - port: "80"
        protocol: TCP
      rules:
        http:
        - method: "GET"
          path: "/public"
```

Outbound to DNS Name



```yaml
apiVersion: "cilium.io/v2"
kind: CiliumNetworkPolicy
metadata:
  name: "to-fqdn"
spec:
  endpointSelector:
    matchLabels:
      app: test-app
  egress:
    - toEndpoints:
      - matchLabels:
          "k8s:io.kubernetes.pod.namespace": kube-system
          "k8s:k8s-app": kube-dns
      toPorts:
        - ports:
           - port: "53"
             protocol: ANY
          rules:
            dns:
              - matchPattern: "*"
    - toFQDNs:
        - matchName: "my-remote-service.com"
```

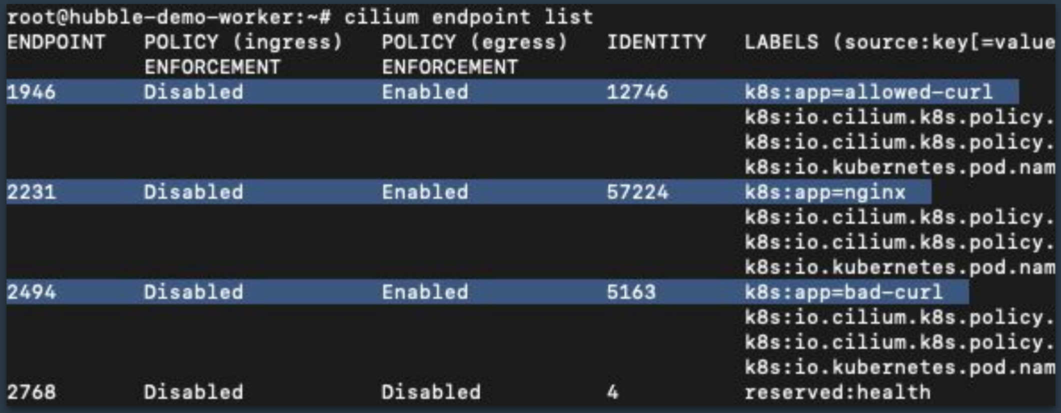
Clusterwide Policy



```yaml
apiVersion: "cilium.io/v2"
kind: CiliumClusterwideNetworkPolicy
description: "Allow a minimum set of required ports on ingress of worker nodes"
metadata:
  name: "lock-down-ingress-worker-node"
spec:
  nodeSelector:
    matchLabels:
      type: ingress-worker
  ingress:
  - fromEntities:
    - remote-node
    - health
  - toPorts:
    - ports:
      - port: "6443"
        protocol: TCP
      - port: "22"
        protocol: TCP
      - port: "2379"
        protocol: TCP
      - port: "4240"
        protocol: TCP
      - port: "8472"
        protocol: UDP
      - port: "REMOVE_ME_AFTER_DOUBLE_CHECKING_PORTS"
        protocol: TCP
```

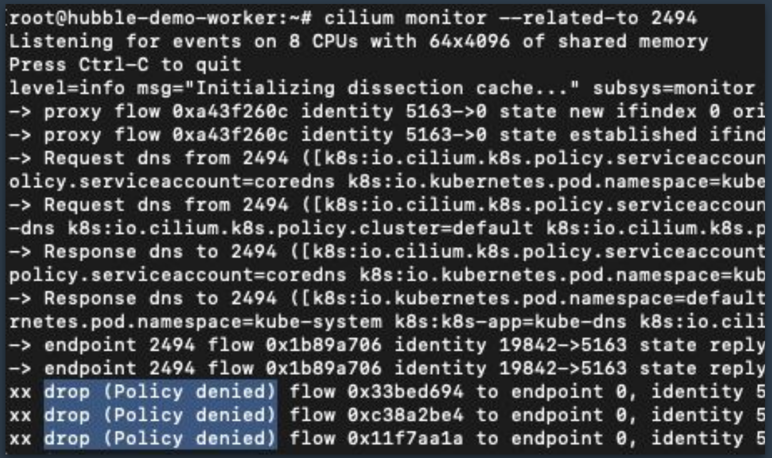
Cilium CLI commands

Listing Endpoints on a Node



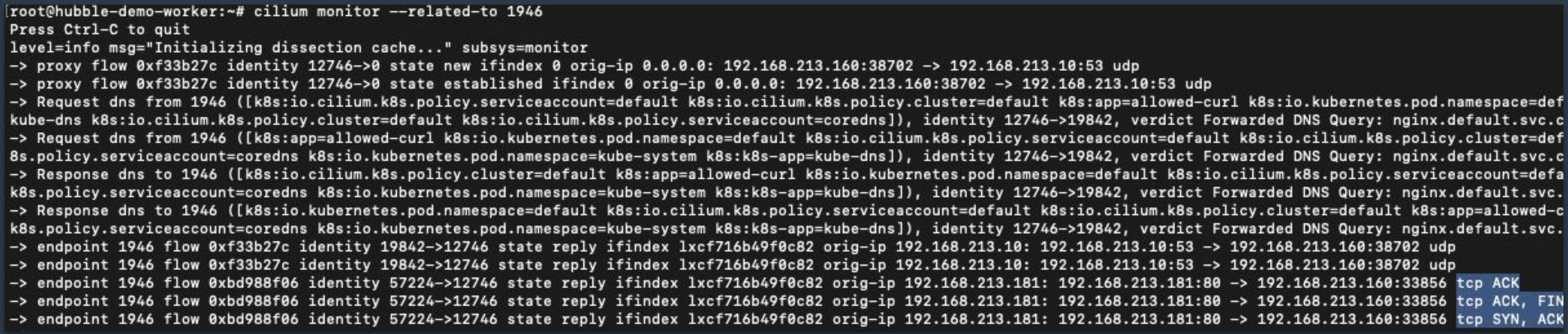
```
root@hubble-demo-worker:~# cilium endpoint list
ENDPOINT   POLICY (ingress)   POLICY (egress)   IDENTITY   LABELS (source:key[=value
           ENFORCEMENT        ENFORCEMENT
1946       Disabled           Enabled           12746      k8s:app=allowed-curl
                                                           k8s:io.cilium.k8s.policy.
                                                           k8s:io.cilium.k8s.policy.
                                                           k8s:io.kubernetes.pod.nam
2231       Disabled           Enabled           57224      k8s:app=nginx
                                                           k8s:io.cilium.k8s.policy.
                                                           k8s:io.cilium.k8s.policy.
                                                           k8s:io.kubernetes.pod.nam
2494       Disabled           Enabled           5163       k8s:app=bad-curl
                                                           k8s:io.cilium.k8s.policy.
                                                           k8s:io.cilium.k8s.policy.
                                                           k8s:io.kubernetes.pod.nam
2768       Disabled           Disabled          4          reserved:health
```

Traffic Denied by Policy



```
root@hubble-demo-worker:~# cilium monitor --related-to 2494
Listening for events on 8 CPUs with 64x4096 of shared memory
Press Ctrl-C to quit
level=info msg="Initializing dissection cache..." subsys=monitor
-> proxy flow 0xa43f260c identity 5163->0 state new ifindex 0 ori
-> proxy flow 0xa43f260c identity 5163->0 state established ifind
-> Request dns from 2494 ([k8s:io.cilium.k8s.policy.serviceaccoun
olicy.serviceaccount=coredns k8s:io.kubernetes.pod.namespace=kube
-> Request dns from 2494 ([k8s:io.cilium.k8s.policy.serviceaccoun
-dns k8s:io.cilium.k8s.policy.cluster=default k8s:io.cilium.k8s.p
-> Response dns to 2494 ([k8s:io.cilium.k8s.policy.serviceaccount
policy.serviceaccount=coredns k8s:io.kubernetes.pod.namespace=kub
-> Response dns to 2494 ([k8s:io.kubernetes.pod.namespace=default
rnetes.pod.namespace=kube-system k8s:k8s-app=kube-dns k8s:io.cili
-> endpoint 2494 flow 0x1b89a706 identity 19842->5163 state reply
-> endpoint 2494 flow 0x1b89a706 identity 19842->5163 state reply
xx drop (Policy denied) flow 0x33bed694 to endpoint 0, identity 5
xx drop (Policy denied) flow 0xc38a2be4 to endpoint 0, identity 5
xx drop (Policy denied) flow 0x11f7aa1a to endpoint 0, identity 5
```

Traffic Allowed by Policy

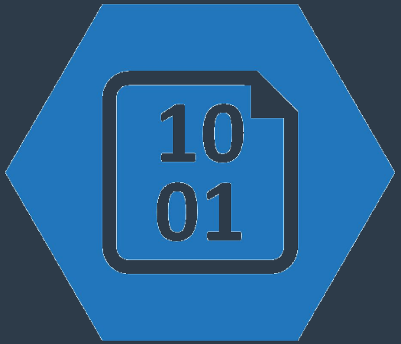


```
[root@hubble-demo-worker:~# cilium monitor --related-to 1946
Press Ctrl-C to quit
level=info msg="Initializing dissection cache..." subsys=monitor
-> proxy flow 0xf33b27c identity 12746->0 state new ifindex 0 orig-ip 0.0.0.0: 192.168.213.160:38702 -> 192.168.213.10:53 udp
-> proxy flow 0xf33b27c identity 12746->0 state established ifindex 0 orig-ip 0.0.0.0: 192.168.213.160:38702 -> 192.168.213.10:53 udp
-> Request dns from 1946 ([k8s:io.cilium.k8s.policy.serviceaccount=default k8s:io.cilium.k8s.policy.cluster=default k8s:app=allowed-curl k8s:io.kubernetes.pod.namespace=def
kube-dns k8s:io.cilium.k8s.policy.cluster=default k8s:io.cilium.k8s.policy.serviceaccount=coredns]), identity 12746->19842, verdict Forwarded DNS Query: nginx.default.svc.c
-> Request dns from 1946 ([k8s:app=allowed-curl k8s:io.kubernetes.pod.namespace=default k8s:io.cilium.k8s.policy.serviceaccount=default k8s:io.cilium.k8s.policy.cluster=def
8s.policy.serviceaccount=coredns k8s:io.kubernetes.pod.namespace=kube-system k8s:k8s-app=kube-dns]), identity 12746->19842, verdict Forwarded DNS Query: nginx.default.svc.c
-> Response dns to 1946 ([k8s:io.cilium.k8s.policy.cluster=default k8s:app=allowed-curl k8s:io.kubernetes.pod.namespace=default k8s:io.cilium.k8s.policy.serviceaccount=defa
k8s.policy.serviceaccount=coredns k8s:io.kubernetes.pod.namespace=kube-system k8s:k8s-app=kube-dns]), identity 12746->19842, verdict Forwarded DNS Query: nginx.default.svc.
-> Response dns to 1946 ([k8s:io.kubernetes.pod.namespace=default k8s:io.cilium.k8s.policy.serviceaccount=default k8s:io.cilium.k8s.policy.cluster=default k8s:app=allowed-c
k8s.policy.serviceaccount=coredns k8s:io.kubernetes.pod.namespace=kube-system k8s:k8s-app=kube-dns]), identity 12746->19842, verdict Forwarded DNS Query: nginx.default.svc.
-> endpoint 1946 flow 0xf33b27c identity 19842->12746 state reply ifindex lxcf716b49f0c82 orig-ip 192.168.213.10: 192.168.213.10:53 -> 192.168.213.160:38702 udp
-> endpoint 1946 flow 0xf33b27c identity 19842->12746 state reply ifindex lxcf716b49f0c82 orig-ip 192.168.213.10: 192.168.213.10:53 -> 192.168.213.160:38702 udp
-> endpoint 1946 flow 0xbd988f06 identity 57224->12746 state reply ifindex lxcf716b49f0c82 orig-ip 192.168.213.181: 192.168.213.181:80 -> 192.168.213.160:33856 tcp ACK
-> endpoint 1946 flow 0xbd988f06 identity 57224->12746 state reply ifindex lxcf716b49f0c82 orig-ip 192.168.213.181: 192.168.213.181:80 -> 192.168.213.160:33856 tcp ACK, FIN
-> endpoint 1946 flow 0xbd988f06 identity 57224->12746 state reply ifindex lxcf716b49f0c82 orig-ip 192.168.213.181: 192.168.213.181:80 -> 192.168.213.160:33856 tcp SYN, ACK
```

Hubble Benefits

- Durable log storage and enterprise Security Information and Event Management (SIEM) integration
- hubble observe command to help with troubleshooting
- Features to expose network traffic flows to teams
  - Hubble UI
  - Network flow logs exported to logging stack
- Tracking network traffic to specific binaries

Durable Audit Log Storage

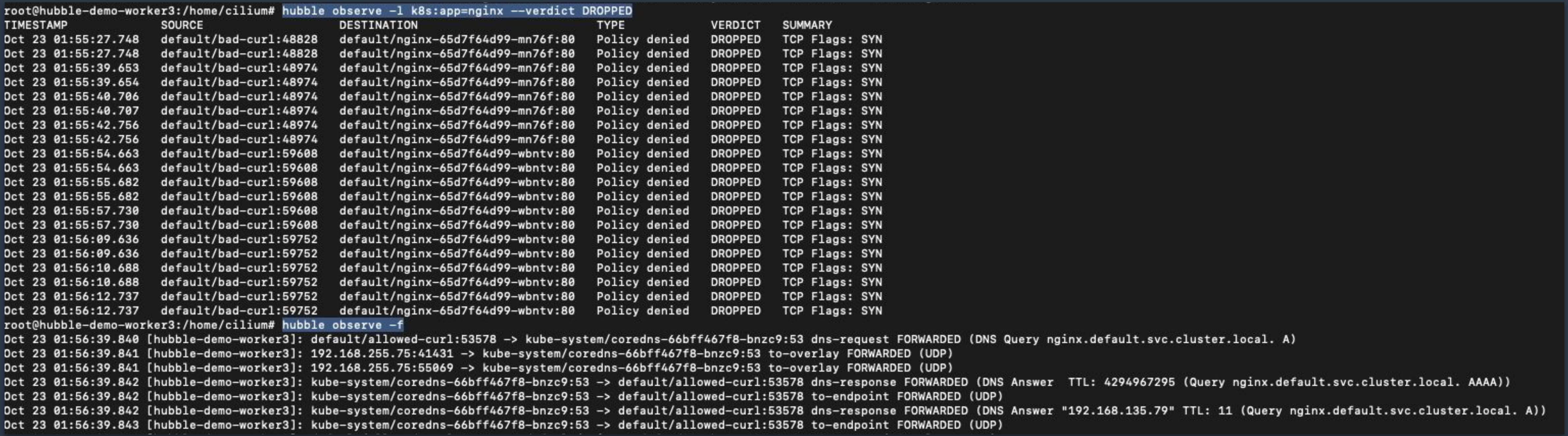


splunk>®

LogRhythm®

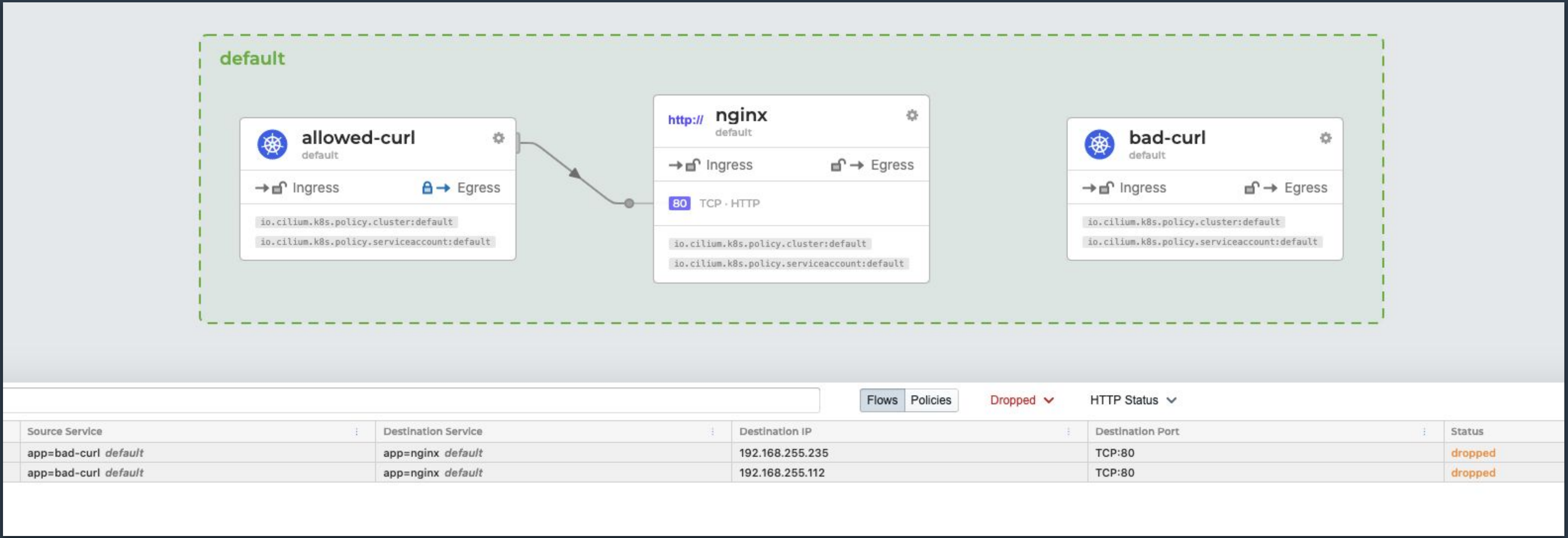
sumologic®

Hubble Observe Command



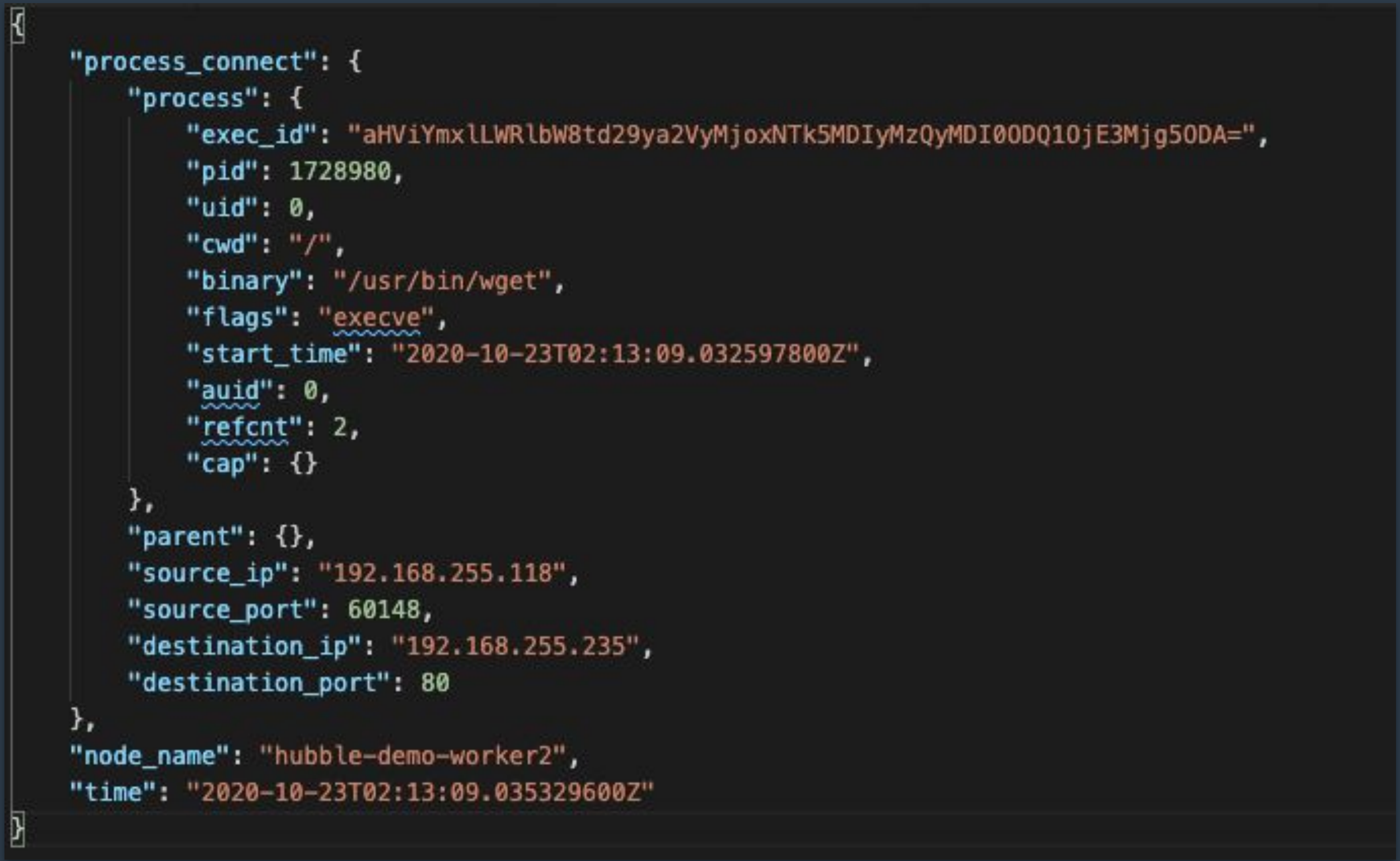
```
root@hubble-demo-worker3:/home/cilium# hubble observe -l k8s:app=nginx --verdict DROPPED
TIMESTAMP           SOURCE                    DESTINATION                     TYPE           VERDICT    SUMMARY
Oct 23 01:55:27.748   default/bad-curl:48828   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:27.748   default/bad-curl:48828   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:39.653   default/bad-curl:48974   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:39.654   default/bad-curl:48974   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:40.706   default/bad-curl:48974   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:40.707   default/bad-curl:48974   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:42.756   default/bad-curl:48974   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:42.756   default/bad-curl:48974   default/nginx-65d7f64d99-mn76f:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:54.663   default/bad-curl:59608   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:54.663   default/bad-curl:59608   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:55.682   default/bad-curl:59608   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:55.682   default/bad-curl:59608   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:57.730   default/bad-curl:59608   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:55:57.730   default/bad-curl:59608   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:56:09.636   default/bad-curl:59752   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:56:09.636   default/bad-curl:59752   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:56:10.688   default/bad-curl:59752   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:56:10.688   default/bad-curl:59752   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:56:12.737   default/bad-curl:59752   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
Oct 23 01:56:12.737   default/bad-curl:59752   default/nginx-65d7f64d99-wbntv:80    Policy denied   DROPPED   TCP Flags: SYN
root@hubble-demo-worker3:/home/cilium# hubble observe -f
Oct 23 01:56:39.840 [hubble-demo-worker3]: default/allowed-curl:53578 -> kube-system/coredns-66bff467f8-bnzc9:53 dns-request FORWARDED (DNS Query nginx.default.svc.cluster.local. A)
Oct 23 01:56:39.841 [hubble-demo-worker3]: 192.168.255.75:41431 -> kube-system/coredns-66bff467f8-bnzc9:53 to-overlay FORWARDED (UDP)
Oct 23 01:56:39.841 [hubble-demo-worker3]: 192.168.255.75:55069 -> kube-system/coredns-66bff467f8-bnzc9:53 to-overlay FORWARDED (UDP)
Oct 23 01:56:39.842 [hubble-demo-worker3]: kube-system/coredns-66bff467f8-bnzc9:53 -> default/allowed-curl:53578 dns-response FORWARDED (DNS Answer  TTL: 4294967295 (Query nginx.default.svc.cluster.local. AAAA))
Oct 23 01:56:39.842 [hubble-demo-worker3]: kube-system/coredns-66bff467f8-bnzc9:53 -> default/allowed-curl:53578 to-endpoint FORWARDED (UDP)
Oct 23 01:56:39.842 [hubble-demo-worker3]: kube-system/coredns-66bff467f8-bnzc9:53 -> default/allowed-curl:53578 dns-response FORWARDED (DNS Answer "192.168.135.79" TTL: 11 (Query nginx.default.svc.cluster.local. A))
Oct 23 01:56:39.843 [hubble-demo-worker3]: kube-system/coredns-66bff467f8-bnzc9:53 -> default/allowed-curl:53578 to-endpoint FORWARDED (UDP)
```

Network Visibility for Teams


default
allowed-curl
default
Ingress
Egress
io.cilium.k8s.policy.cluster:default
io.cilium.k8s.policy.serviceaccount:default
http://
nginx
default
Ingress
Egress
80
TCP · HTTP
io.cilium.k8s.policy.cluster:default
io.cilium.k8s.policy.serviceaccount:default
bad-curl
default
Ingress
Egress
io.cilium.k8s.policy.cluster:default
io.cilium.k8s.policy.serviceaccount:default
Flows
Policies
Dropped
HTTP Status

Source Service	Destination Service	Destination IP	Destination Port	Status
app=bad-curl *default*	app=nginx *default*	192.168.255.235	TCP:80	dropped
app=bad-curl *default*	app=nginx *default*	192.168.255.112	TCP:80	dropped

Searchable Logs and Which Binary



```
{
    "process_connect": {
        "process": {
            "exec_id": "aHViYmxlLWRlbW8td29ya2VyMjoxNTk5MDIyMzQyMDI0ODQ1OjE3Mjg5ODA=",
            "pid": 1728980,
            "uid": 0,
            "cwd": "/",
            "binary": "/usr/bin/wget",
            "flags": "execve",
            "start_time": "2020-10-23T02:13:09.032597800Z",
            "auid": 0,
            "refcnt": 2,
            "cap": {}
        },
        "parent": {},
        "source_ip": "192.168.255.118",
        "source_port": 60148,
        "destination_ip": "192.168.255.235",
        "destination_port": 80
    },
    "node_name": "hubble-demo-worker2",
    "time": "2020-10-23T02:13:09.035329600Z"
}
```

eBPF Powers All of This