ISOVALENT

Tetragon: Runtime Observability and Security

Tetragon

Security Observability & Runtime Enforcement

Metrics

Events

SIEM

fluentd

Logs

Traces

JSON

Tetragon Agent

Pod

app.py

Func Calls

Pod

app.go

Code Exec

Kernel Runtime

Smart Collector

Stack Traces

Ring Buffer

Metrics

Hash Maps

Linux Kernel

Process Execution

Syscall Activity

System Calls

File Access

VFS

Seq Attack

TCP/IP

NS Escapes

Priv Escalations

Namespaces

Data Access

Storage

HTTP, DNS, TLS

eth0

Network

ISOVALENT

Agenda

- eBPF-based Runtime Security - Tetragon
- Demo and Dashboard share
- Next Steps

ISOVALENT

BPF
The Cloud Native Observability Platform

BPF The Cloud Native Observability Platform: Why is this Interesting

With an observability platform we can answer...

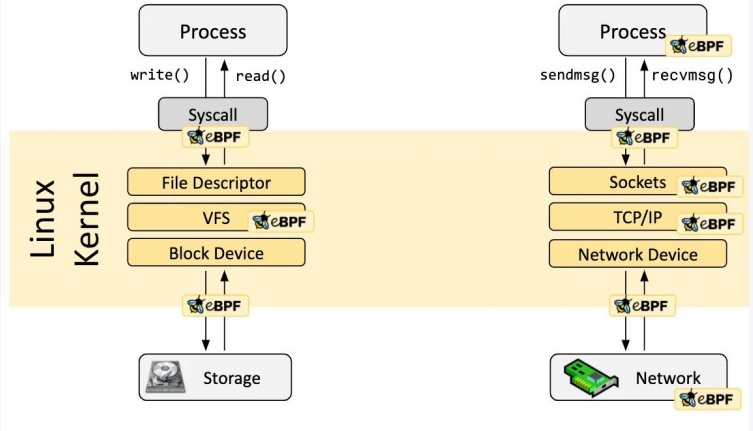
- What binaries have run in the past and are running now?
- Are we running latest version of libraries?
- What network connections are open: both listening, connecting for UDP/TCP/RawSockets/...
- Is my network healthy: are there drops, is the latency within SLAs, detecting bursts and dips, ...
- What files are being accessed, written, executed, mmapped, ...
- Are my connections encrypted? And with what TLS, IPSec, Wireguard.
- Are my TLS connections meeting compliance requirements.
- What syscalls are my applications using today, do they suddenly use new syscalls, args, binaries?

Competing Ideas

- Real time (time scales of us, ms, ...)
- Minimal CPU and memory constraints (<1vcpu, <500MB)
- Minimal Application impacts (<10%)
- Offline and Online modeles, respect the pipeline limitations

BPF The Cloud Native Observability Platform: Tetragon:

- Platform to deploy BPF hooks in clusters, aggregate data, alert on events and push data into SIEM, Security Pipeline, Grafana.
- Extend Linux observability and security model to be Kubernetes aware
  - Pods, Labels, Containers, etc
- Can be dynamically done at scale (10k,20k,...+ nodes)
- Minimally invasive when done with care



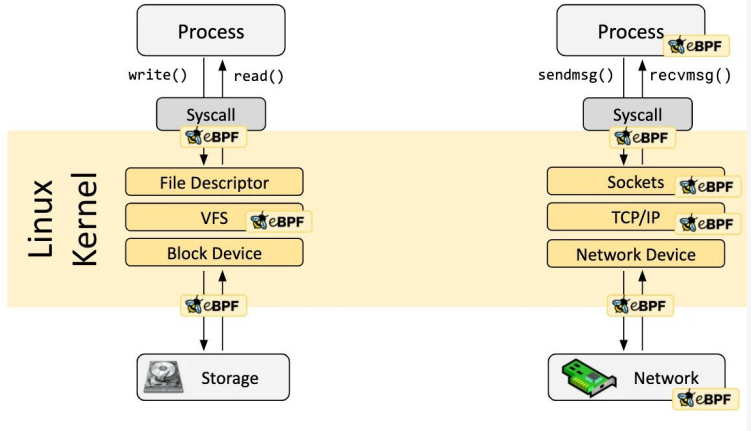
BPF The Cloud Native Observability Platform: Tetragon Philosophy:

- Most common Threat Model: Users can not be trusted
  - User memory is untrusted (TOCTOU)
  - Uprobes are untrusted
  - Missing data is a serious bug (SEV-*)

- Designed to scale (10, 50, 100, 1k, 10k,20k,...+ nodes)
  - Will not add features that don't scale
  - Filters and aggregation in BPF 1st
  - Stop gap Filter and aggregate in user space

- Fail closed

- Be kind to the security/audit stack
  - Events cost money $$$
  - Rate limits
  - Filter and aggregate in BPF and user space



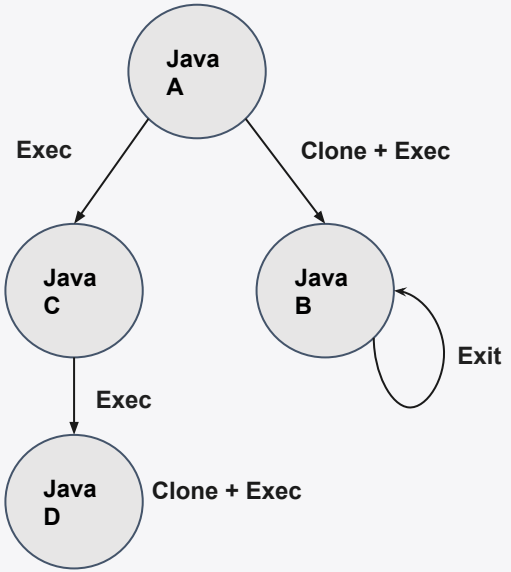
ISOVALENT

BPF The Cloud Native Observability Platform
Tetragon: Core Feature: Execution Traces

Every Executed Binary In the System Is Recorded

- System is K8s Cluster, Servers, and VMs
- For historical DB time series database


```
"process_exec":
    "process":{
      "exec_id":"bWluaWt1YmU6MT...",
      "sha-256":"...",
      "pid":17978,
      "cwd":"/",
      "binary":"/docker-entrypoint.sh",
      "arguments":"/docker-entrypoint.sh nginx -g,
      "start_time":"2021-10-13T12:58:31.794Z",
      "pod":{...},
      "parent_exec_id":"...",
      "cap":{...}
    },
    "parent":{...},
  "node_name":"minikube",
  "time":"2021-10-13T12:58:31.794Z"
```

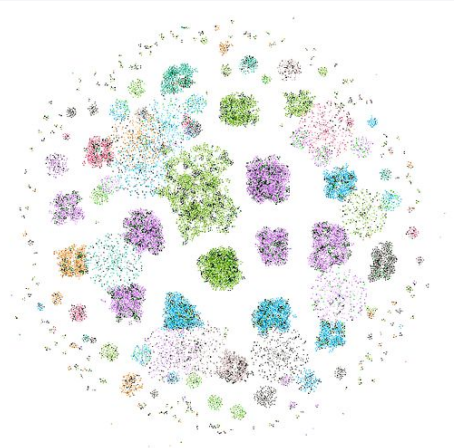



BPF The Cloud Native Observability Platform
Tetragon: Core Feature: Identity and Location

Identity = { binary, pid, libs, args, buildID, digest }

Location = { cluster, node, namespace, pod, container, time }

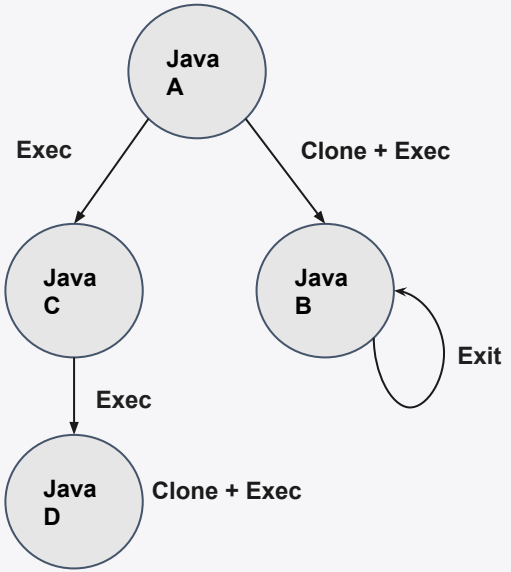
F(Identity, Location) → Unique ID



BPF The Cloud Native Observability Platform
Tetragon: Core Feature: Execution Traces

Customer Dashboard Grafana


```
"process_exec":
    "process":{
      "exec_id":"bWluaWt1YmU6MT...",
      "sha-256":"...",
      "pid":17978,
      "cwd":"/",
      "binary":"/docker-entrypoint.sh",
      "arguments":"/docker-entrypoint.sh nginx -g,
      "start_time":"2021-10-13T12:58:31.794Z",
      "pod":{...},
      "parent_exec_id":"...",
      "cap":{...}
    },
    "parent":{...},
  "node_name":"minikube",
  "time":"2021-10-13T12:58:31.794Z"
```

BPF The Cloud Native Observability Platform
Tetragon: Kprobe, anything

```
kind: TracingPolicy
Metadata:
  name: "sys-write-follow-fd-etc-pwd"
Spec:
  Kprobes:
  - call: "fd_install"
    syscall: false
    Args:
    - index: 0
      type: int
    - index: 1
      type: "file"
```

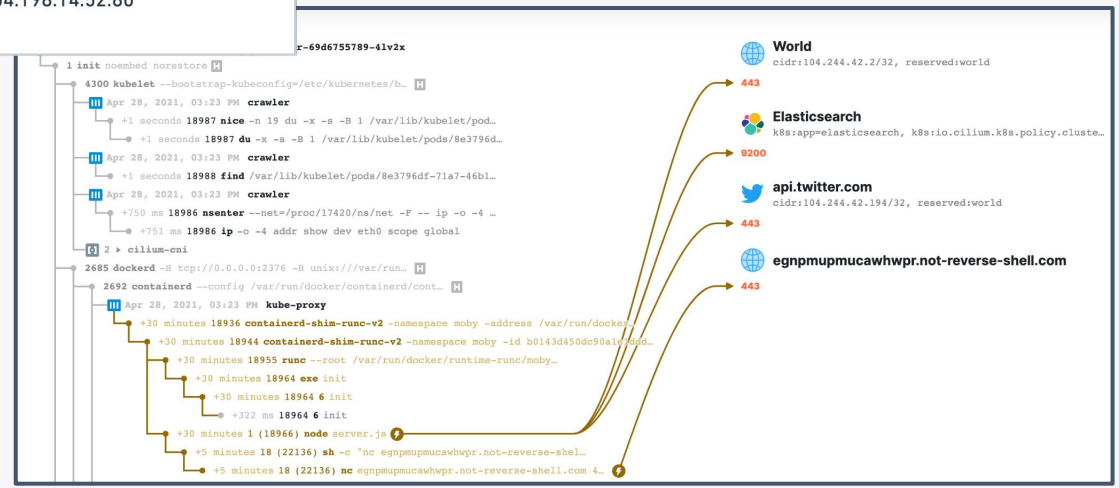
```
selectors:
    - matchPIDs:
      - operator: NotIn
        followForks: true
        isNamespacePID: true
        Values:
        - 0
        - 1
      matchArgs:
      - index: 1
        operator: "Equal"
        Values:
        - "etc/passwd"
      matchActions:
      - action: FollowFD
```

BPF The Cloud Native Observability Platform
Tetragon: TCP/UDP Network Connectivity

Every connect/listen/accept in the system is known.

System: K8s nodes, VMs, Servers

```
🚀 process default/xwing /usr/bin/curl http://cilium.io
🔌 connect default/xwing /usr/bin/curl tcp 10.244.0.6:34965 -> 104.198.14.52:80
📫 sendmsg default/xwing /usr/bin/curl tcp 10.244.0.6:34965 -> 104.198.14.52:80 bytes 73
🧹 close   default/xwing /usr/bin/curl tcp 10.244.0.6:34965 -> 104.198.14.52:80
💥 exit    default/xwing /usr/bin/curl http://cilium.io 0
```



BPF The Cloud Native Observability Platform

Tetragon: File Integrity Monitoring

```
kprobes:
- call: "fd_install"
  syscall: false
  args:
    - index: 0
      type: int
    -  index: 1
      type: "file"
  selectors:
  - matchPIDs:
    - operator: NotIn
      followForks: true
      isNamespacePID: true
      Values:
      - 0
      - 1
  matchArgs:
    - index: 1
      operator: "Equal"
      Values:
      - "etc/passwd"
  matchActions:
  - action: FollowFD
    argFd: 0
    argName: 1
```

```
🚀 process default/xwing /usr/bin/vi /etc/passwd
📬 open    default/xwing /usr/bin/vi /etc/passwd
📚 read    default/xwing /usr/bin/vi /etc/passwd 1269 bytes
📪 close   default/xwing /usr/bin/vi /etc/passwd
📬 open    default/xwing /usr/bin/vi /etc/passwd
📝 write   default/xwing /usr/bin/vi /etc/passwd 1277 bytes
💥 exit    default/xwing /usr/bin/vi /etc/passwd 0
```


```
  "function_name":"__x64_sys_write",
  "args":[
    {
      "file_arg":{
        "path":"etc/passwd"
        "bytes_arg":
"ZGFlbW9uOng6MjoyOmRhZW1vbjovc2Jpbjovc2Jpbi9ub2xvZ2luCmFkbTp4OjM6NDphZG06L3Zhci9hZG06L3NiaW4vbm9sb2dpbgp
      "size_arg":"627"
    }
  ],
  "action":"KPROBE_ACTION_POST"
},
"node_name":"gke-kprobe-validation-default-pool-b5e9dab6-k0cj",
"time":"2021-10-18T20:08:55.567Z"
```

BPF The Cloud Native Observability Platform Tetragon: File Integrity Monitoring

Customer Grafana Dashboard

BPF The Cloud Native Observability Platform

Tetragon: Enforcement

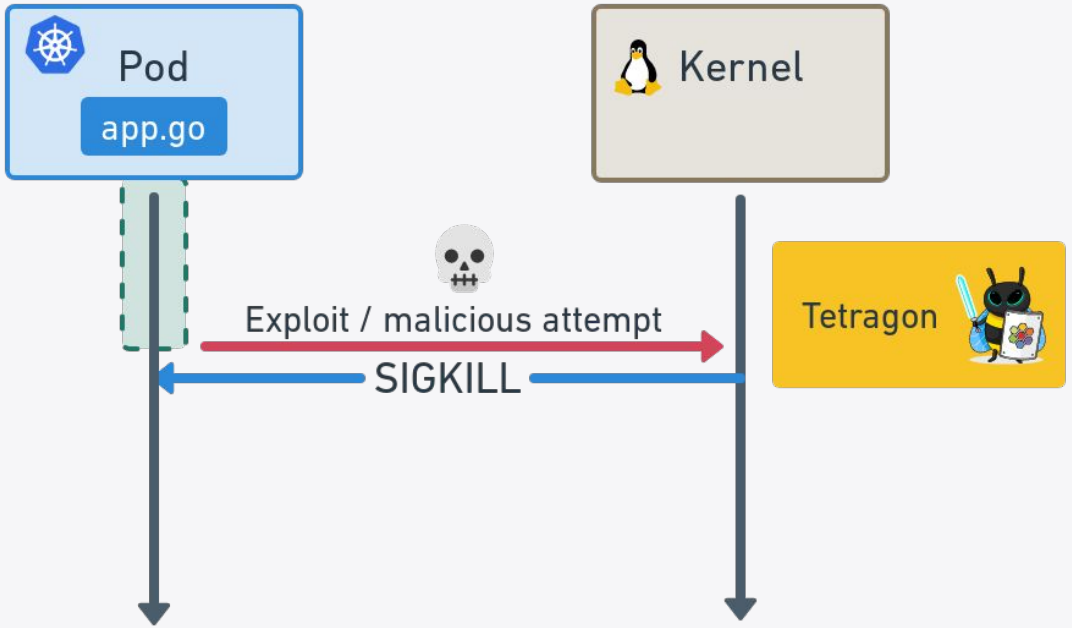
```
kprobes:
  - call: "__x64_sys_mount"
  syscall: true
  args:
  - index: 0
    type:  "string"
  - index: 1
    type:  "string"
  selectors:
  - matchPIDs:
    - operator: NotIn
      followForks: false
      isNamespacePID: true
      values:
      - 1
    matchActions:
    - action: Override
      action: Sigkill
```

```
"process_kprobe":{
    "process":{
  "exec_id":"Z2tlLWtwcm9i...",
      "pid":193983,
       "pod":{}
      },
    "parent_exec_id":"Z2tlLWtwcm9i...",},
    "parent":{}},
    "function_name":"__x64_sys_mount",
    "args":[{
        "string_arg":"/dev/sda1"},{"string_arg":"/tmp"}],
    "action":"KPROBE_ACTION_SIGKILL","KPROBE_ACTION_OVERRIDE"   },
"node_name":"gke-kprobe-validating-default-pool-ad5a40e9-xttn",
"time":"2021-10-18T16:09:18.882Z"
```

BPF The Cloud Native Observability Platform

Tetragon: Enforcement

```
kprobes:
  - call: "__x64_sys_mount"
  syscall: true
  args:
  - index: 0
    type:  "string"
  - index: 1
    type:  "string"
  selectors:
  - matchPIDs:
    - operator: NotIn
      followForks: false
      isNamespacePID: true
      values:
      - 1
    matchActions:
    - action: Override
      action: Sigkill
```



BPF The Cloud Native Observability Platform
Tetragon: Performance Benchmarks

perf stat -e cycles,cycles:u,cycles:k,instructions:u,instructions:k -a -r 3 -- bash -c 'make -j16 && make clean'

	Time	Diff	Cycles	Time	Events	Lost
loader sensor, rb 16M	564.571	+2.79%	23,254,118,945,251	+3.45%	3,719,746	0
syscalls (190) filtered, rb 1M, exported	806.350	+6.42%	30,884,295,872,860	+5.26%	1,446,198	0
syscalls (190) filtered, open/close, rb 16M, exported*	931.960	+23.00%	37,412,599,649,778	+27.50%	27,460,143	48,789,763

* *Worse Case: Monitoring every open/close and exporting it to JSON/GRPC/Metrics

BPF The Cloud Native Observability Platform
Tetragon: Demo

Tetragon

BPF The Cloud Native Observability Platform
Tetragon: Future Work

- T-digest and Q-digest overall theme in kernel aggregation and summary
- Multi-attach kprobe, uprobe
- SR-IOV metrics (customers using SR-IOV for latency)
- Application signatures
- BPF signatures
- Kernel/User stack traces
- Iterate net_device list, net_ns list, sockets, inodes
- TX/RX NIC descriptors for low-level NIC health metrics
- Sockmap/Sockhash fixes and improvements
- KTLS coming soon time to get it working golang, java, openssl, ...

How to contribute?

- **github.com/cilium/tetragon**
  - Use the tool: report bugs, create feature request, tell your user experience
  - Improve the documentation (open issues)
  - Add your use cases “./crds/examples”, “./contrib”
  - Tell us about how it doesn’t work for some use cases
  - Feedback on UI, CRDs, etc
  - Fix a bug, Implement a feature

- **Lots of work across all layers of the stack**
  - Documentation, K8s, Golang, Systems Programming, BPF, Linux Kernel, Packaging

Thank you! Q&A

cilium/tetragon

@ciliumproject

cilium.io

@jrfastab

@sharlns