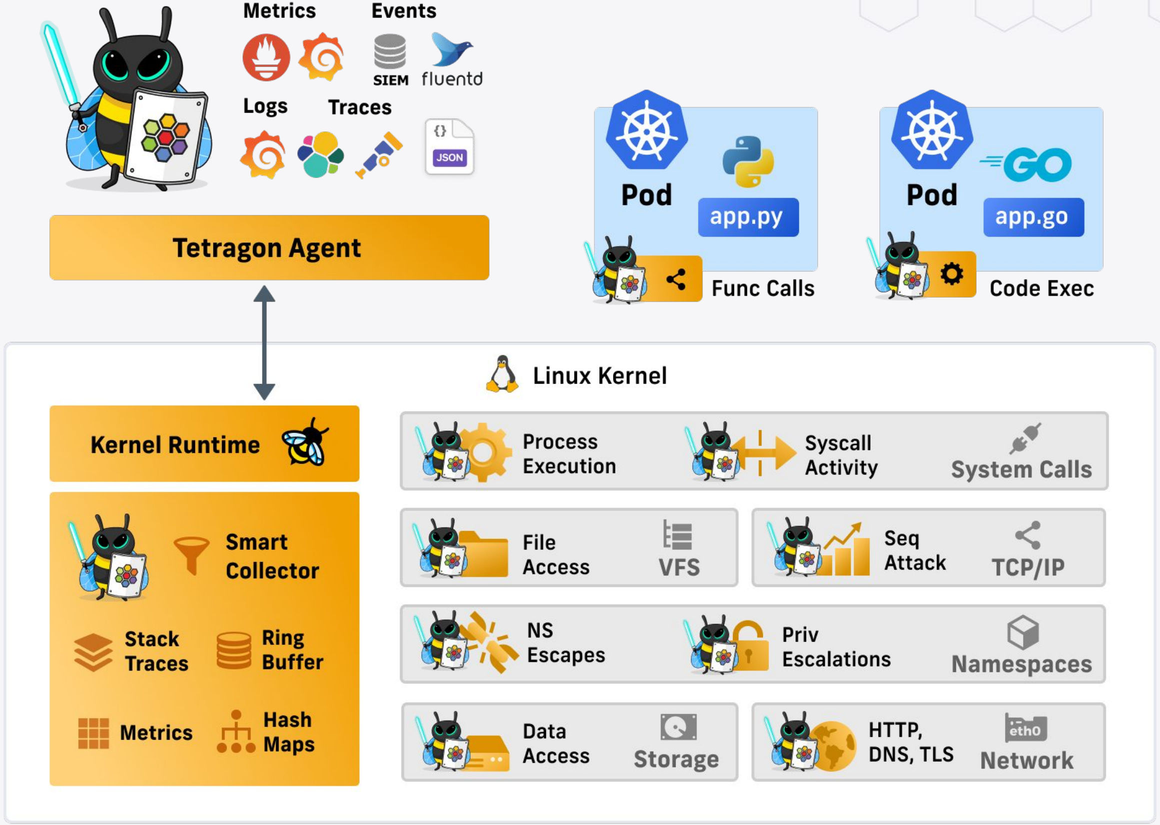
Tetragon:
Auditing and Enforcement

John Fastabend

ISOVALENT
now part of CISCO

Tetragon

Security Observability & Runtime Enforcement

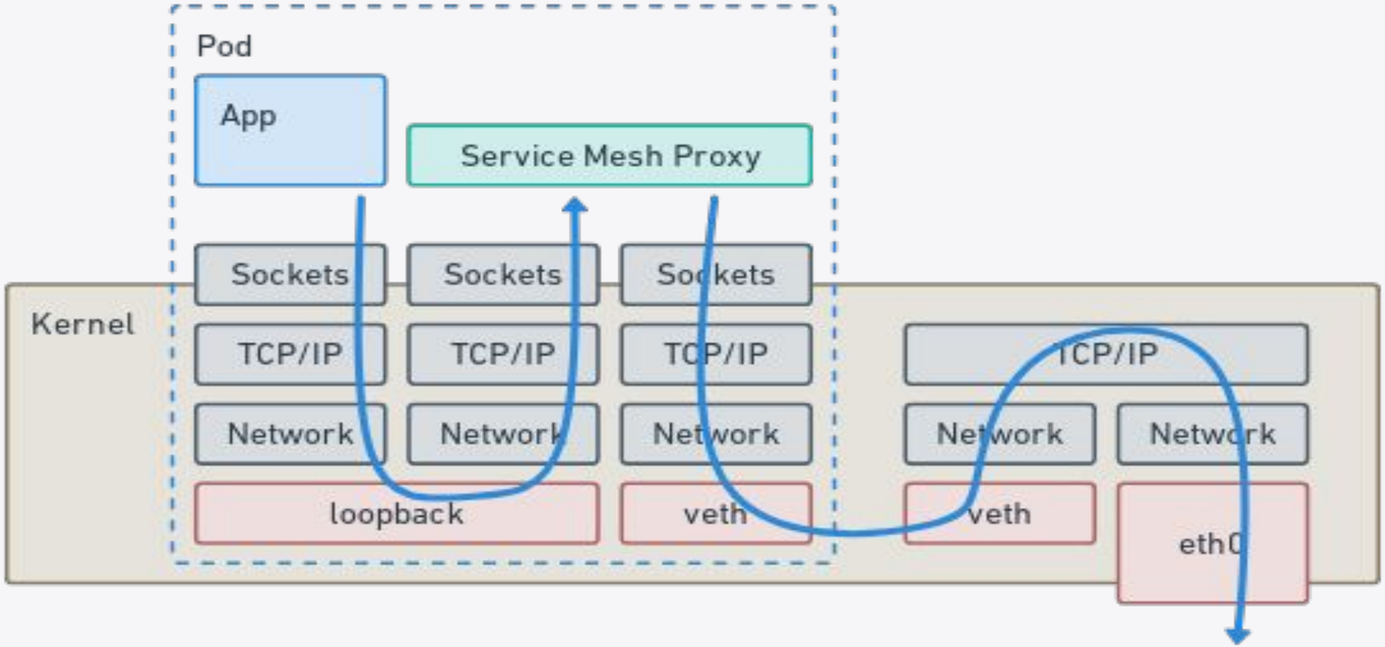


ISOVALENT

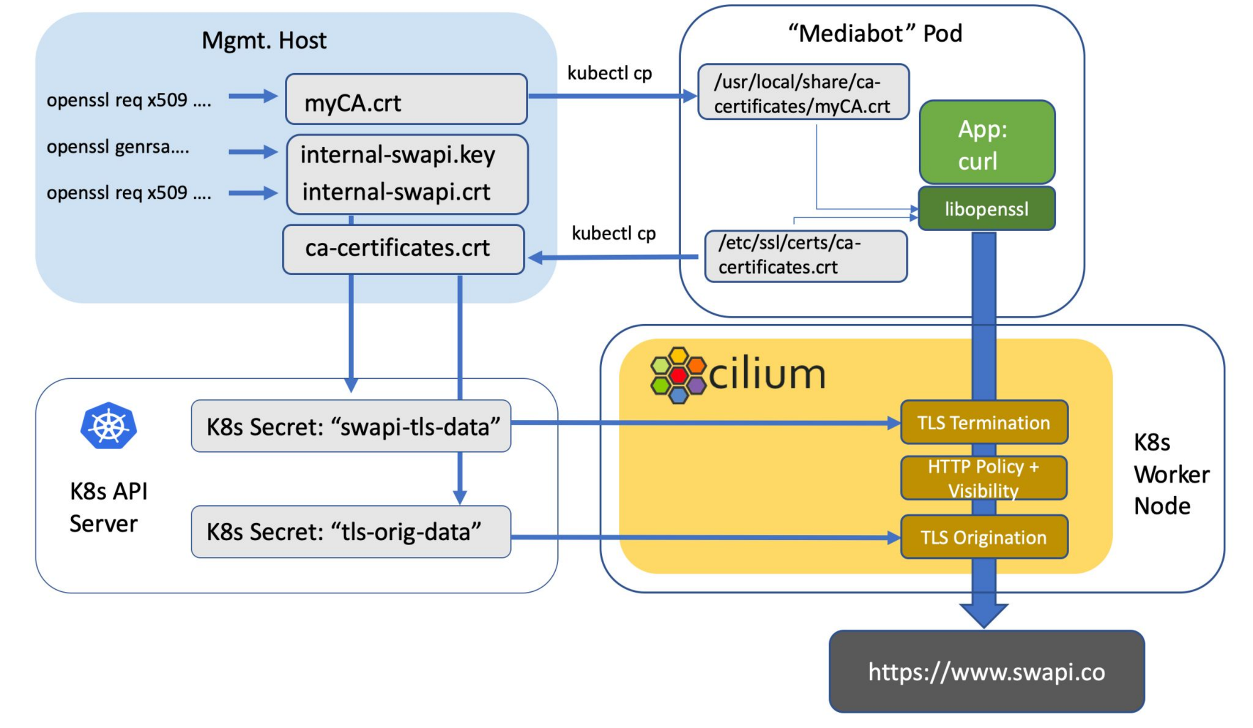
Agenda: Walkthrough a BPF Networking Stack

- L7: State of parsers
- L3/L4: Process Aware Network Enforcement
- L2: What is going on down here?
- Push vs Pull
- Next Steps

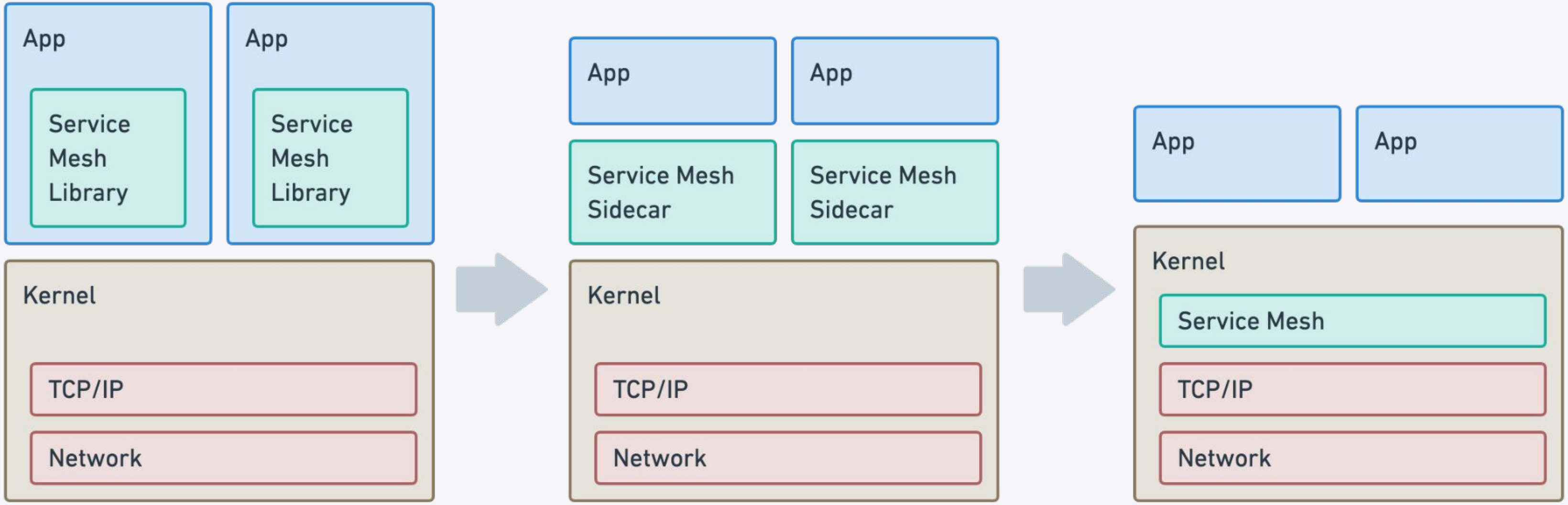
L7: Parsers Why?



kTLS + L7: Why?

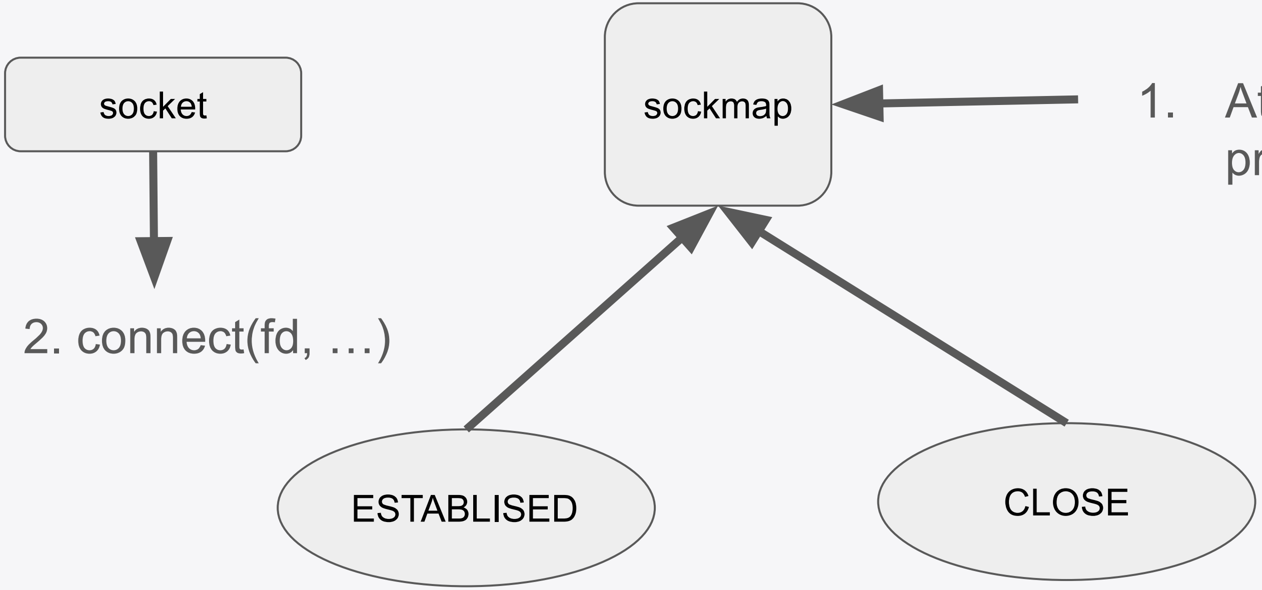

Mgmt. Host
openssl req x509 ....
myCA.crt
openssl genrsa....
openssl req x509 ....
internal-swapi.key
internal-swapi.crt
ca-certificates.crt
kubectl cp
kubectl cp
"Mediabot" Pod
/usr/local/share/ca-certificates/myCA.crt
App: curl
libopenssl
/etc/ssl/certs/ca-certificates.crt
K8s API Server
K8s Secret: "swapi-tls-data"
K8s Secret: "tls-orig-data"
cilium
TLS Termination
HTTP Policy + Visibility
TLS Origination
K8s Worker Node
https://www.swapi.co

L7: Parsers as (security) kernel primitive



L7: Parsers as (security) kernel primitive

Life cycle



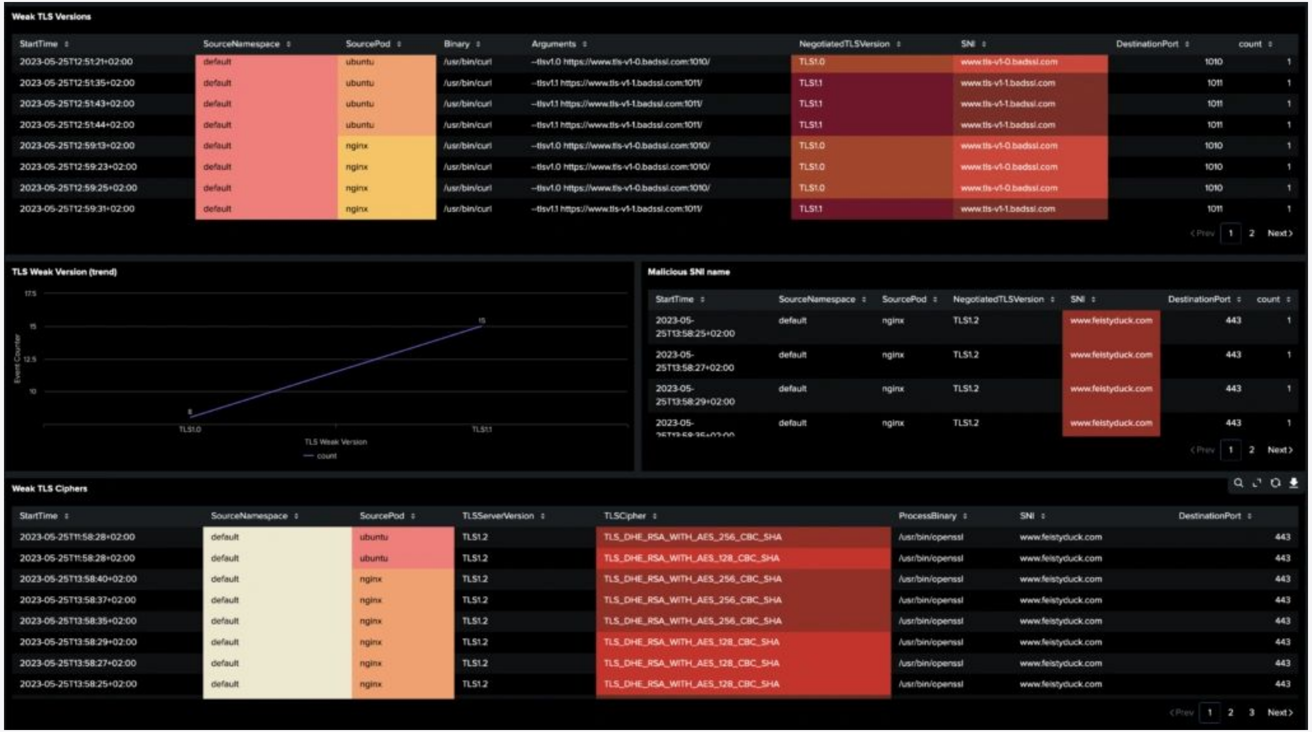
1. Attach BPF programs

2. connect(fd, …)

3. update_elem(key, sock(fd))

4. delete_elem(...); kfree_rcu()

L7: Parsers as (security) kernel primitive



Weak TLS Versions

StartTime	SourceNamespace	SourcePod	Binary	Arguments	NegotiatedTLSVersion	SNI	DestinationPort	count
2023-05-25T12:51:21+02:00	default	ubuntu	/usr/bin/curl	--tlsv1.0 https://www.tls-v1-0.badssl.com:1010/	TLS1.0	www.tls-v1-0.badssl.com	1010	1
2023-05-25T12:51:35+02:00	default	ubuntu	/usr/bin/curl	--tlsv1.1 https://www.tls-v1-1.badssl.com:1011/	TLS1.1	www.tls-v1-1.badssl.com	1011	1
2023-05-25T12:51:43+02:00	default	ubuntu	/usr/bin/curl	--tlsv1.1 https://www.tls-v1-1.badssl.com:1011/	TLS1.1	www.tls-v1-1.badssl.com	1011	1
2023-05-25T12:51:44+02:00	default	ubuntu	/usr/bin/curl	--tlsv1.1 https://www.tls-v1-1.badssl.com:1011/	TLS1.1	www.tls-v1-1.badssl.com	1011	1
2023-05-25T12:59:13+02:00	default	nginx	/usr/bin/curl	--tlsv1.0 https://www.tls-v1-0.badssl.com:1010/	TLS1.0	www.tls-v1-0.badssl.com	1010	1
2023-05-25T12:59:23+02:00	default	nginx	/usr/bin/curl	--tlsv1.0 https://www.tls-v1-0.badssl.com:1010/	TLS1.0	www.tls-v1-0.badssl.com	1010	1
2023-05-25T12:59:25+02:00	default	nginx	/usr/bin/curl	--tlsv1.0 https://www.tls-v1-0.badssl.com:1010/	TLS1.0	www.tls-v1-0.badssl.com	1010	1
2023-05-25T12:59:31+02:00	default	nginx	/usr/bin/curl	--tlsv1.1 https://www.tls-v1-1.badssl.com:1011/	TLS1.1	www.tls-v1-1.badssl.com	1011	1

TLS Weak Version (trend)

Malicious SNI name

StartTime	SourceNamespace	SourcePod	NegotiatedTLSVersion	SNI	DestinationPort	count
2023-05-25T13:58:25+02:00	default	nginx	TLS1.2	www.feistyduck.com	443	1
2023-05-25T13:58:27+02:00	default	nginx	TLS1.2	www.feistyduck.com	443	1
2023-05-25T13:58:29+02:00	default	nginx	TLS1.2	www.feistyduck.com	443	1
2023-05-25T13:58:35+02:00	default	nginx	TLS1.2	www.feistyduck.com	443	1

Weak TLS Ciphers

StartTime	SourceNamespace	SourcePod	TLSServerVersion	TLSCipher	ProcessBinary	SNI	DestinationPort
2023-05-25T11:58:28+02:00	default	ubuntu	TLS1.2	TLS_DHE_RSA_WITH_AES_256_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443
2023-05-25T11:58:28+02:00	default	ubuntu	TLS1.2	TLS_DHE_RSA_WITH_AES_128_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443
2023-05-25T13:58:40+02:00	default	nginx	TLS1.2	TLS_DHE_RSA_WITH_AES_256_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443
2023-05-25T13:58:37+02:00	default	nginx	TLS1.2	TLS_DHE_RSA_WITH_AES_256_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443
2023-05-25T13:58:35+02:00	default	nginx	TLS1.2	TLS_DHE_RSA_WITH_AES_256_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443
2023-05-25T13:58:29+02:00	default	nginx	TLS1.2	TLS_DHE_RSA_WITH_AES_128_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443
2023-05-25T13:58:27+02:00	default	nginx	TLS1.2	TLS_DHE_RSA_WITH_AES_128_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443
2023-05-25T13:58:25+02:00	default	nginx	TLS1.2	TLS_DHE_RSA_WITH_AES_128_CBC_SHA	/usr/bin/openssl	www.feistyduck.com	443

L7: Parsers as (security) kernel primitive

- **Streaming Parsers**: 5.15*, 6.1*, 6.5, 6.8, bpf-next
- **Distributions**: AL2022, AL2023, Ubuntu 22.04/24.04, GKE rapid
- **Architecture:** ARM, X86
- **CI:**
  - Nginx compliance test
  - Tetragon CI tests
  - ./selftests/bpf/sockmap

L7: Parsers as (security) kernel primitive

- **Verdict/StrParser**:

  Open issue:

  Updates tp->**copied_seq** as data is aggregate. But, copied_seq is used to wakeup tcp_poll().

  Result:

  Application may wake up before data is copied to socke receive_queue. Fix is to delay copied_seq update until data is enqueued in receive_queue after BPF program runs. Care is needed because copied_seq has implications on acks.

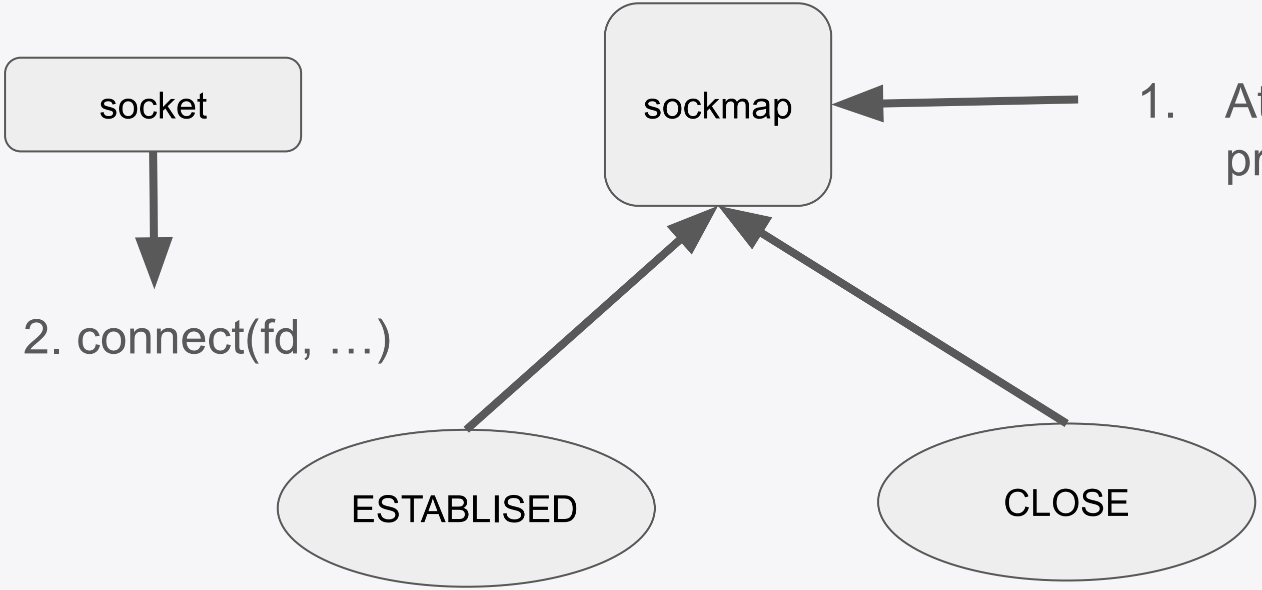
L7: Parsers as (security) kernel primitive

- **Zerocopy:**
  If we allow this it is problematic for security. Zero copy and L7 security tooling do not seem to compatible.

  syzbot reported an issue that needs to be addressed.

  Just block zerocopy on BPF sockets? But it is still useful for !security and best effort.

L7: Parsers as (security) kernel primitive
Future



socket

sockmap

1. Attach BPF programs

2. connect(fd, …)

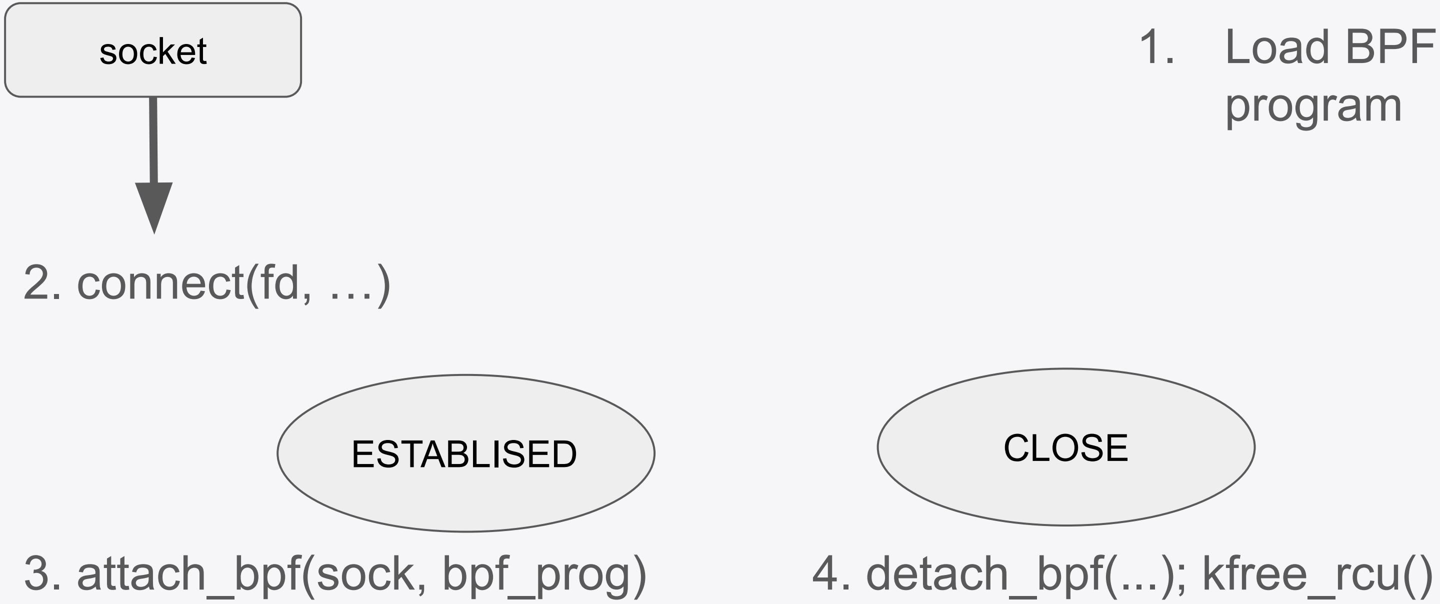
ESTABLISED

CLOSE

3. update_elem(key, sock(fd))

4. delete_elem(...); kfree_rcu()

L7: Parsers as (security) kernel primitive
Future



1. Load BPF program

2. connect(fd, …)

3. attach_bpf(sock, bpf_prog)

4. detach_bpf(...); kfree_rcu()

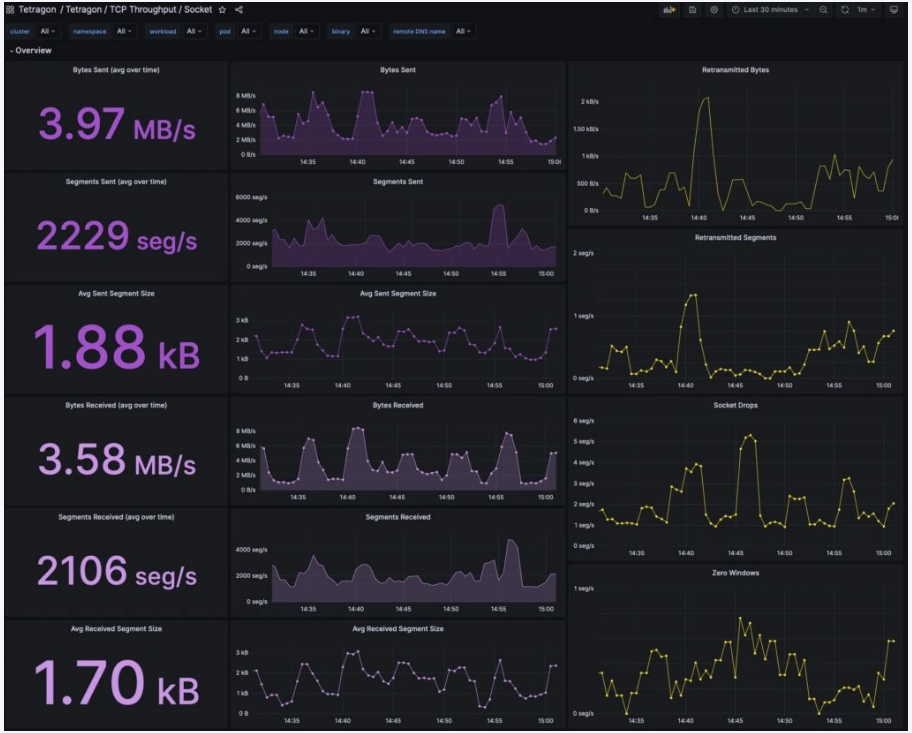
L7: Parsers as (security) kernel primitive
Future

KTLS:

- Library supported: Openssl 3.0
- Library in use:
    - Go crypto/tls
    - Java TLS
    - *

DTLS: ?
Quic: ?

L3: Audit and Enforcement



Tetragon / Tetragon / TCP Throughput / Socket

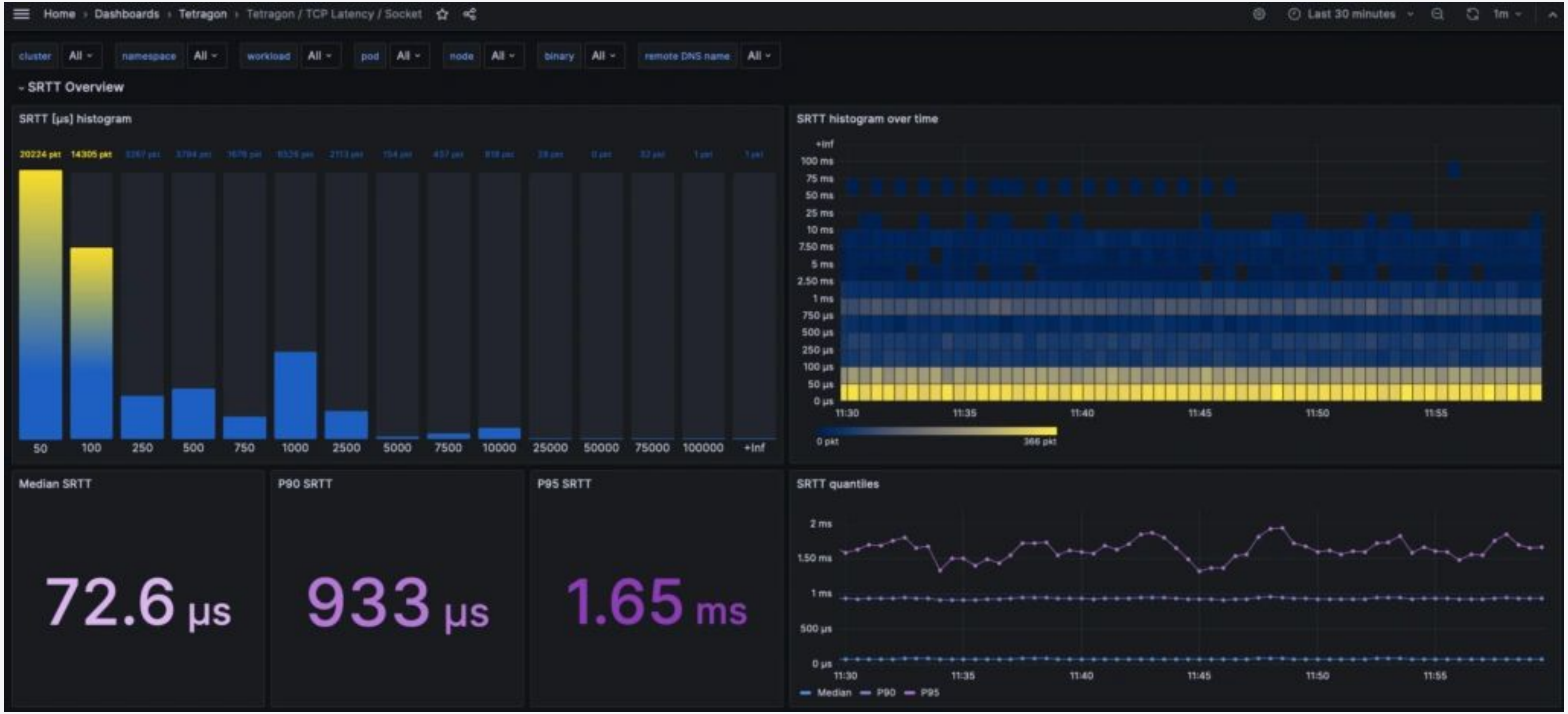
Last 30 minutes · 1m

cluster All · namespace All · workload All · pod All · node All · binary All · remote DNS name All

Overview

Panel	Value
Bytes Sent (avg over time)	3.97 MB/s
Segments Sent (avg over time)	2229 seg/s
Avg Sent Segment Size	1.88 kB
Bytes Received (avg over time)	3.58 MB/s
Segments Received (avg over time)	2106 seg/s
Avg Received Segment Size	1.70 kB

L3: Audit and Enforcement



Home › Dashboards › Tetragon › Tetragon / TCP Latency / Socket

cluster: All · namespace: All · workload: All · pod: All · node: All · binary: All · remote DNS name: All

Last 30 minutes · 1m

SRTT Overview

Median SRTT: 72.6 µs

P90 SRTT: 933 µs

P95 SRTT: 1.65 ms

L3: Audit and Enforcement:
If I had a TCAM ...

- Policy enforcement requires wildcard lookups

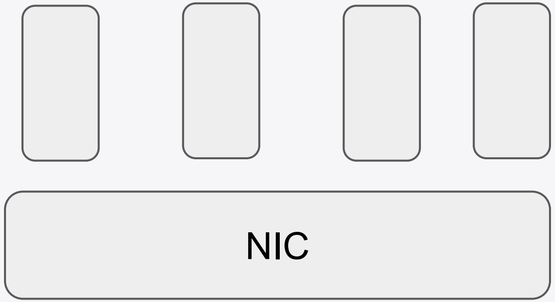
  web-client : * : ebpf.io
  web-client : 443 : *
  * : 80 : *

- Without TCAM we end up with multiple hash lookups.

- Todo understand algorithm trade-offs and performance testing

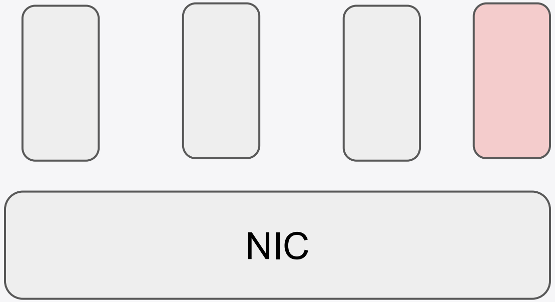
L2: NIC Stats

- NIC:
  - TX / RX bytes
  - Drops, Errors



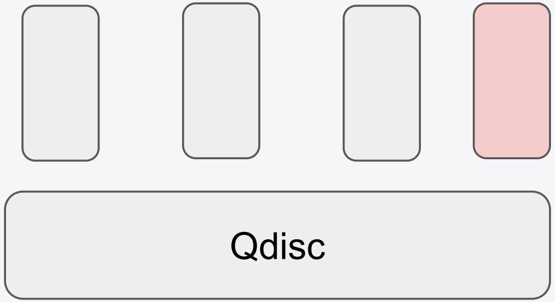
L2: NIC Stats

- NIC:
  - Lack generic mechanism to understand details
  - netns(), dev() iterators missing



L2: Qdisc Occupancy and time on Qdisc

- Qdisc: Occupancy histogram



Tetragon Interesting Comment: Pull not Push

- Current Model:
  - BPF
    - Observe interesting event
    - Apply Filters
    - Push Events through Ring Buffer
  - Userspace
    - Reads Ring Buffer
    - Logic to aggregate, summarize, …
    - Push to Pipeline/DB

Thank you!

- cilium/tetragon
- @ciliumproject
- cilium.io

@jrfastab