Open Infrastructure FOUNDATION

INDONESIA OpenInfra Days

Kekuatan Super eBPF

Giri Kuncoro

Senior Software Engineer at **Gojek**

Jakarta, August 21, 2021

About Me

Giri Kuncoro

Senior Software Engineer at **Gojek**

@girikuncoro

GiriKuncoro

eBPF

eBPF

INDONESIA OpenInfra Days

“Superpowers have finally come to Linux”

- Brendan Gregg, Netflix

eBPF

INDONESIA OpenInfra Days

“Superpowers have finally come to Linux”

“eBPF does to Linux what Javascript does to HTML”

- Brendan Gregg, Netflix

Open Infrastructure FOUNDATION

GiriKuncoro

BRI EasyStack Biznet GioCloud boer technology NVIDIA

eBPF

INDONESIA OpenInfra Days

Run code in the kernel without having to write a kernel module

man bpf

man bpf

The **bpf**() system call performs a **range of operations** related to extended Berkeley Packet Filters. Extended BPF (or eBPF) is similar to the original ("classic") BPF (cBPF) used to filter network packets.

man bpf

The **bpf**() system call performs a **range of operations** related to extended Berkeley Packet Filters. Extended BPF (or eBPF) is similar to the original ("classic") BPF (cBPF) used to filter network packets.

For both cBPF and eBPF **programs**,

man bpf

The **bpf**() system call performs a **range of operations** related to extended Berkeley Packet Filters. Extended BPF (or eBPF) is similar to the original ("classic") BPF (cBPF) used to filter network packets.

For both cBPF and eBPF **programs**, the kernel statically analyzes the programs before loading them, in order to ensure that they **cannot harm the running system**.

man bpf

The **bpf**() system call performs a **range of operations** related to extended Berkeley Packet Filters. Extended BPF (or eBPF) is similar to the original ("classic") BPF (cBPF) used to filter network packets.

For both cBPF and eBPF **programs**, the kernel statically analyzes the programs before loading them, in order to ensure that they **cannot harm the running system**.

eBPF extends cBPF in multiple ways, including the ability to call a fixed set of **in-kernel helper functions** (via the **BPF_CALL** opcode extension provided by eBPF)

man bpf

The **bpf**() system call performs a **range of operations** related to extended Berkeley Packet Filters. Extended BPF (or eBPF) is similar to the original ("classic") BPF (cBPF) used to filter network packets.

For both cBPF and eBPF **programs**, the kernel statically analyzes the programs before loading them, in order to ensure that they **cannot harm the running system**.

eBPF extends cBPF in multiple ways, including the ability to call a fixed set of **in-kernel helper functions** (via the **BPF_CALL** opcode extension provided by eBPF) and **access shared data structures such as eBPF maps**.

man bpf

eBPF programs can be written in a **restricted C** that is compiled (using the **clang** compiler) into **eBPF bytecode.** Various features are omitted from this restricted C, such as loops, global variables, variadic functions, floating-point numbers, and passing structures as function arguments.

clang & LLVM

INDONESIA OpenInfra Days

The LLVM Project is a collection of modular and reusable compiler and toolchain technologies. Despite its name, LLVM has little to do with traditional virtual machines. The name "LLVM" itself is not an acronym; it is the full name of the project.

clang & LLVM

The LLVM Project is a collection of modular and reusable compiler and toolchain technologies. Despite its name, LLVM has little to do with traditional virtual machines. The name "LLVM" itself is not an acronym; it is the full name of the project.

Clang is an "LLVM native" C/C++/Objective-C compiler, which aims to deliver amazingly fast compiles.

man bpf

The kernel contains a just-in-time (JIT) compiler that **translates eBPF bytecode into native machine code** for better performance.

(limited) C → eBPF bytecode → machine code

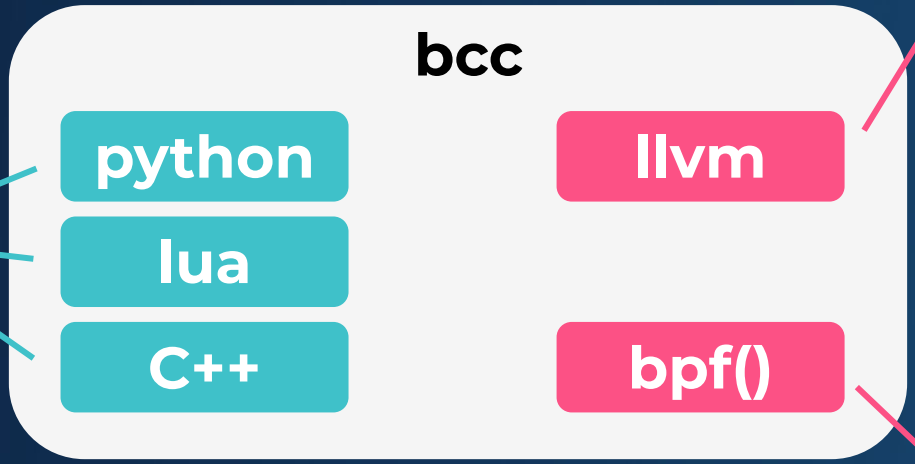
Writing Hello World

bcc

BCC makes BPF programs easier to write, with kernel instrumentation in C (and includes a C wrapper around LLVM), and frontends in Python and Lua.

bcc

Compiles eBPF program



Language support

Wrapper for bpf() syscalls

INDONESIA OpenInfra Days

Attaching eBPF programs to events

Open Infrastructure FOUNDATION

BRI EasyStack Biznet GioCloud boer technology NVIDIA

Triggering eBPF programs

INDONESIA OpenInfra Days

eBPF programs can be attached to different events

- Kprobes
- Uprobes
- Tracepoints
- Network packets
- Perf events

bcc function names

```
b = BPF(text="""
    int kprobe__sys_clone(void *ctx) {
        bpf_trace_printk("Hello, OpenInfra!\\n");
        return 0;
    }
""")

b.trace_print()
```

bpf_trace_printk()

Writing to

`/sys/kernel/debug/tracing/trace_pipe`

eBPF Maps

eBPF maps

Maps are a generic data structure for storage of different types of data. They allow sharing of data between eBPF kernel programs, and also between kernel and user-space applications.

Each map type has attributes:

- type
- max number of elements
- etc.

eBPF maps

INDONESIA OpenInfra Days

Maps are a generic data structure for storage of different types of data. They allow sharing of data between eBPF kernel programs, and also between kernel and user-space applications.

Each map type has attributes:

- type
- max number of elements
- etc.

BPF_MAP_TYPE_UNSPEC
BPF_MAP_TYPE_HASH
BPF_MAP_TYPE_ARRAY
BPF_MAP_TYPE_PROG_ARRAY
BPF_MAP_TYPE_PERF_EVENT_ARRAY
BPF_MAP_TYPE_PERCPU_HASH
BPF_MAP_TYPE_PERCPU_ARRAY
BPF_MAP_TYPE_STACK_TRACE
BPF_MAP_TYPE_CGROUP_ARRAY
BPF_MAP_TYPE_LRU_HASH

```
clang -O2 -emit-llvm -c bpf.c -o - | llc -march=bpf -filetype=obj -o bpf.o
```

(limited) C

ELF object file

- eBPF opcodes
- eBPF maps

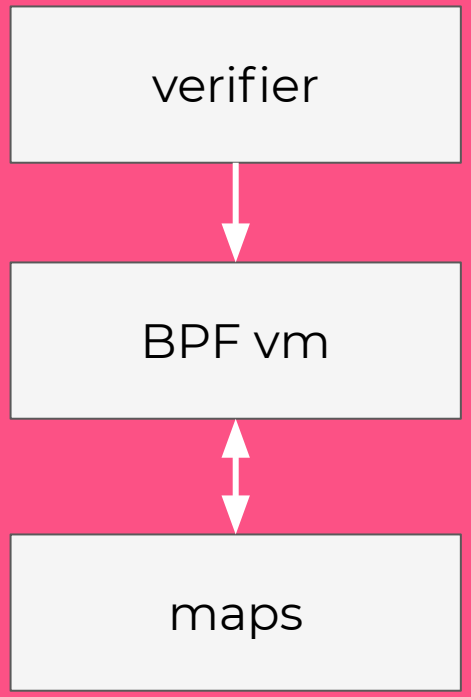
ELF object file

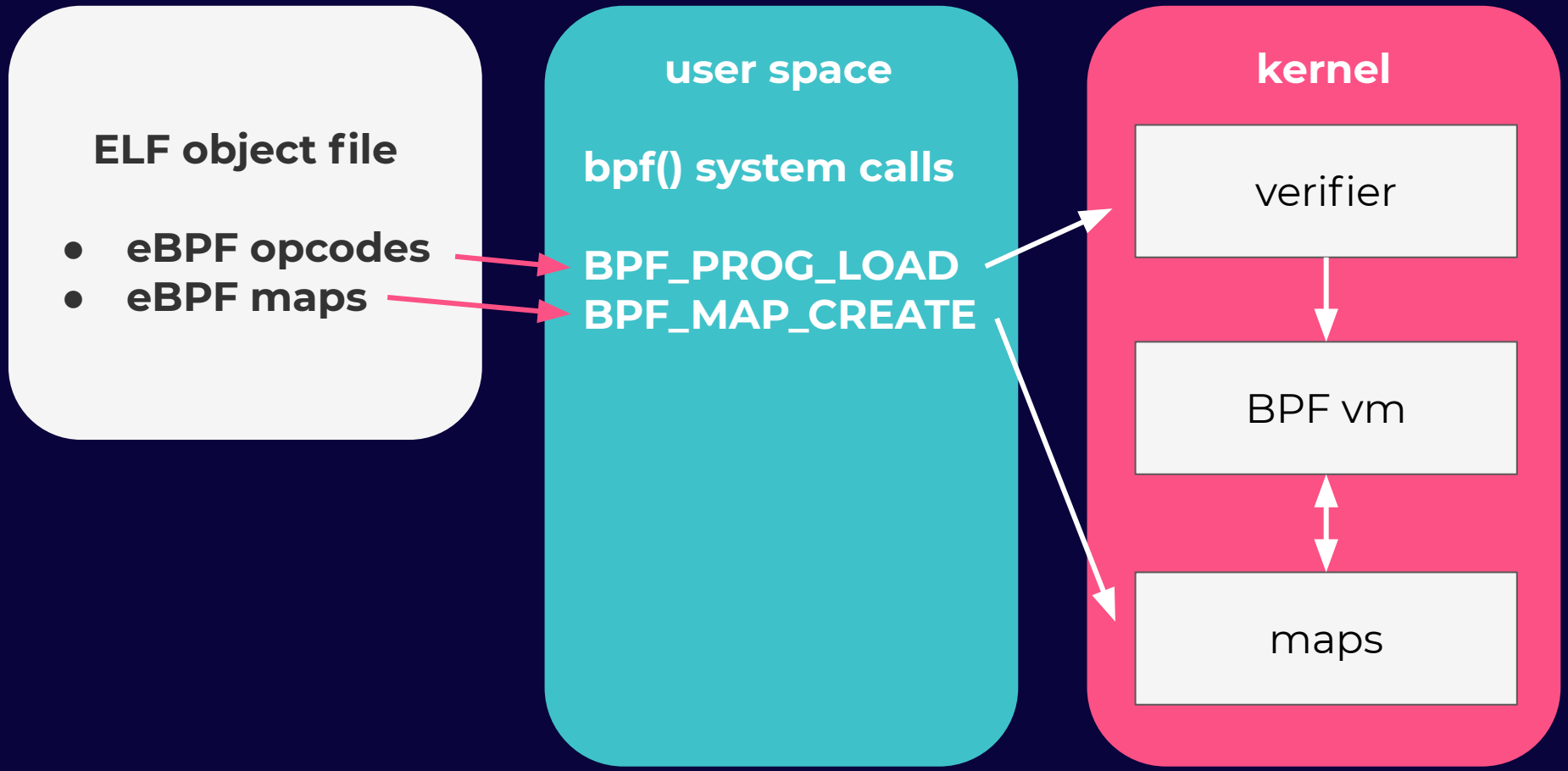
- eBPF opcodes
- eBPF maps

user space

bpf() system calls

kernel




ELF object file
eBPF opcodes
eBPF maps
user space
bpf() system calls
BPF_PROG_LOAD
BPF_MAP_CREATE
kernel
verifier
BPF vm
maps

ELF object file

- eBPF opcodes
- eBPF maps

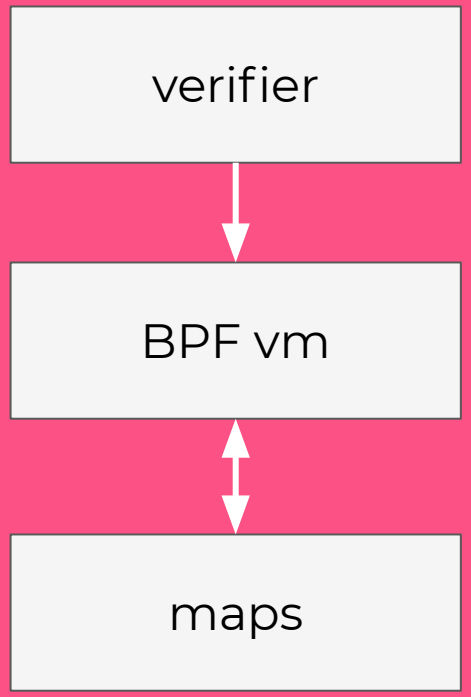
user space

bpf() system calls

BPF_PROG_LOAD
BPF_MAP_CREATE

Attach BPF program to event

kernel



ELF object file

- eBPF opcodes
- eBPF maps

user space

bpf() system calls

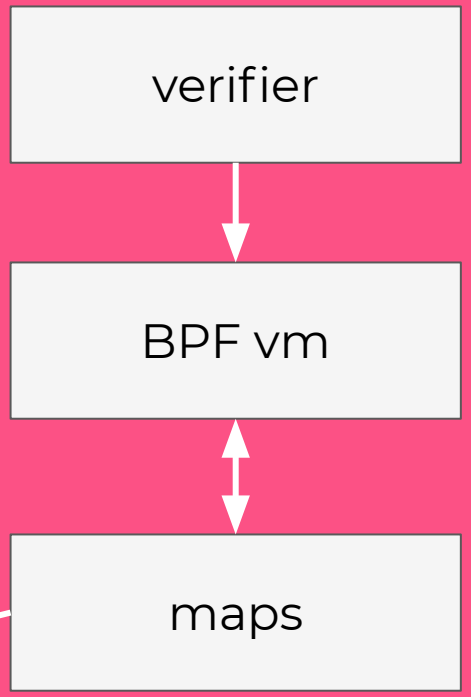
BPF_PROG_LOAD
BPF_MAP_CREATE

Attach BPF program to event

Read / write maps

BPF_MAP_GET_NEXT_KEY
BPF_MAP_LOOKUP_ELEM
BPF_MAP_UPDATE_ELEM
BPF_MAP_DELETE_ELEM

kernel



bcc map

```
b = BPF(text="""
   BPF_HASH(syscalls);
""")

... b["syscalls"].items()...
```

eBPF helper functions

These helpers are used by eBPF programs to interact with the system, or with the context in which they work. For instance, they can be used to print debugging messages, to get the time since the system was booted, to interact with eBPF maps, or to manipulate network packets.

eBPF helper functions

INDONESIA OpenInfra Days

These helpers are used by eBPF programs to interact with the system, or with the context in which they work. For instance, they can be used to print debugging messages, to get the time since the system was booted, to interact with eBPF maps, or to manipulate network packets.

```
bpf_trace_printk()
bpf_map_*_elem()
bpf_get_cuurrent_pid_tgid()
```

Verifier

Each eBPF program is a set of instructions that is safe to run until its completion. An in-kernel verifier statically determines that the eBPF program terminates and is safe to execute.

- No loops
- No bad pointer dereferences
- Restricted program size
- Always exits

INDONESIA
OpenInfra Days

BPF Tools

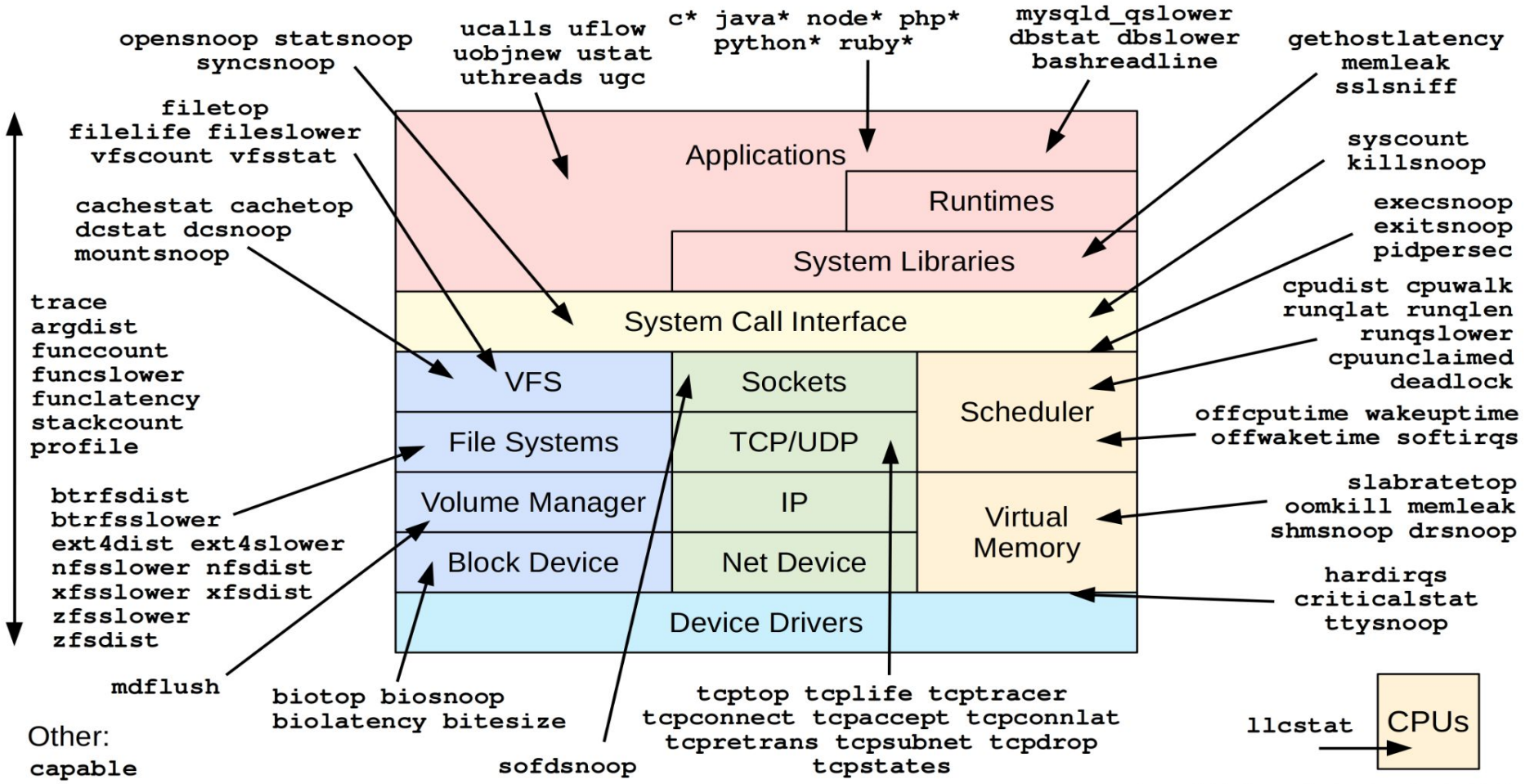
Kekuatan super untuk semua

eBPF

Open Infrastructure
FOUNDATION

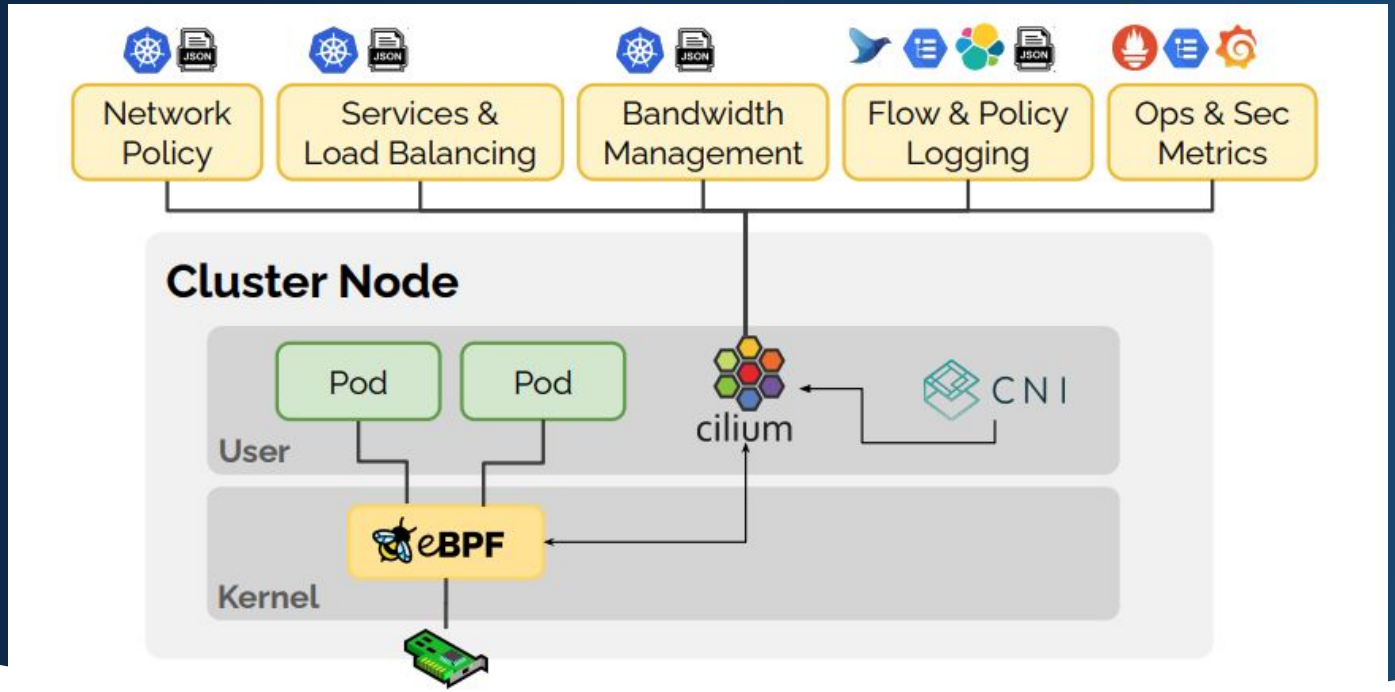
BRI EasyStack Biznet GioCloud boer technology NVIDIA

Linux bcc/BPF Tracing Tools

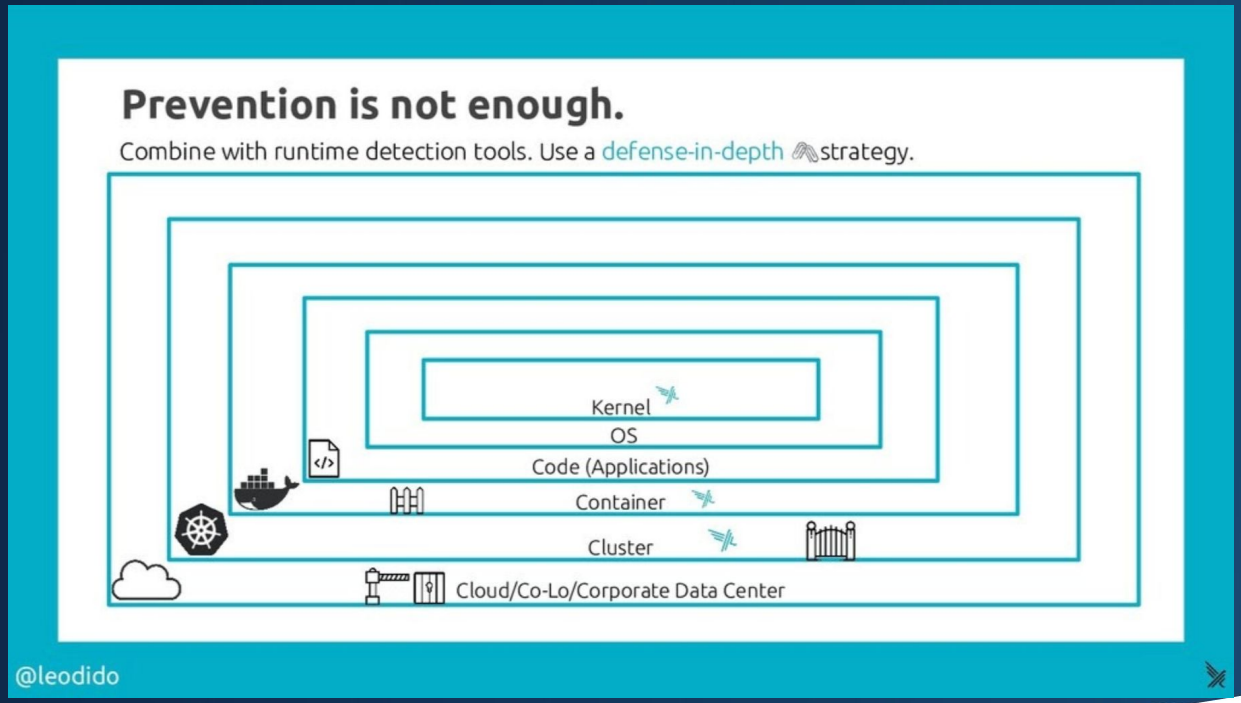


https://github.com/iovisor/bcc#tools 2019

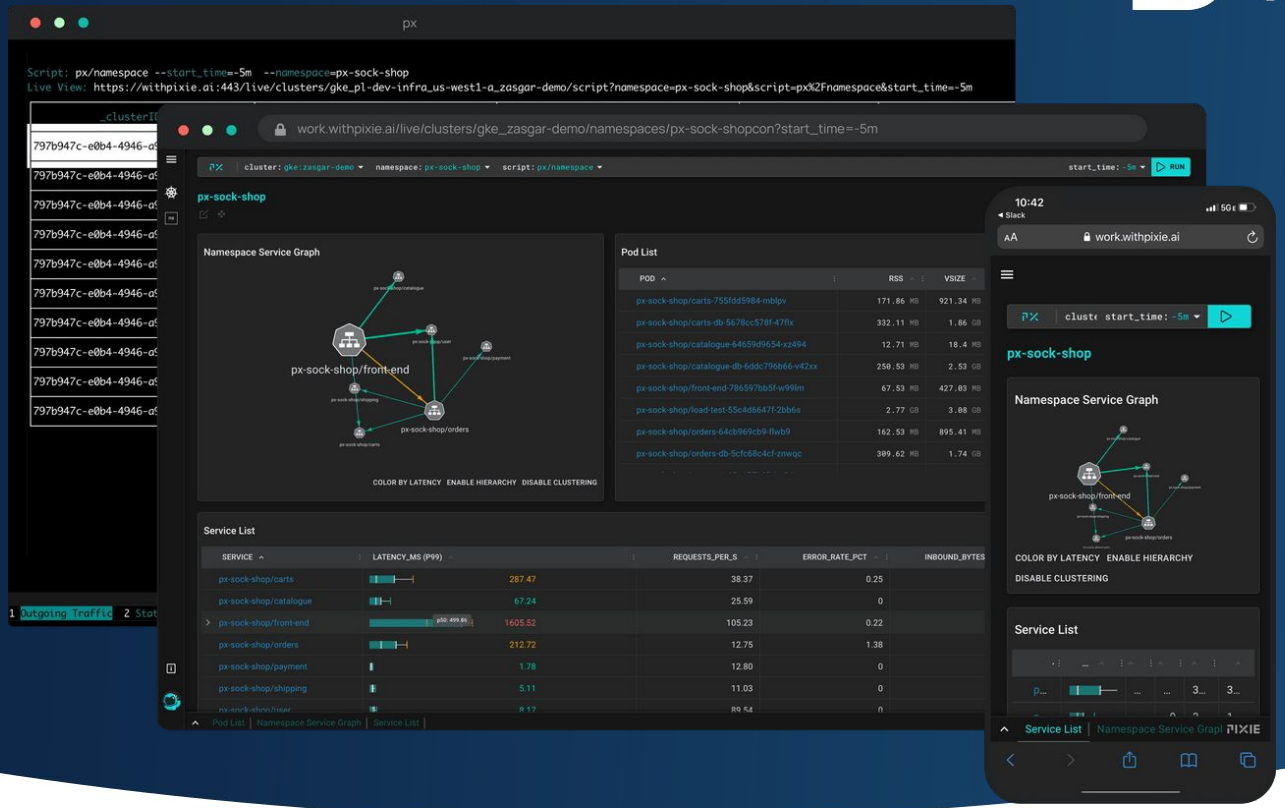
Cilium



Falco


Prevention is not enough.
Combine with runtime detection tools. Use a defense-in-depth strategy.
Kernel
OS
Code (Applications)
Container
Cluster
Cloud/Co-Lo/Corporate Data Center
@leodido

Pixie



INDONESIA OpenInfra Days

Open Infrastructure FOUNDATION

GiriKuncoro

BRI EasyStack Biznet GioCloud boer technology NVIDIA

References

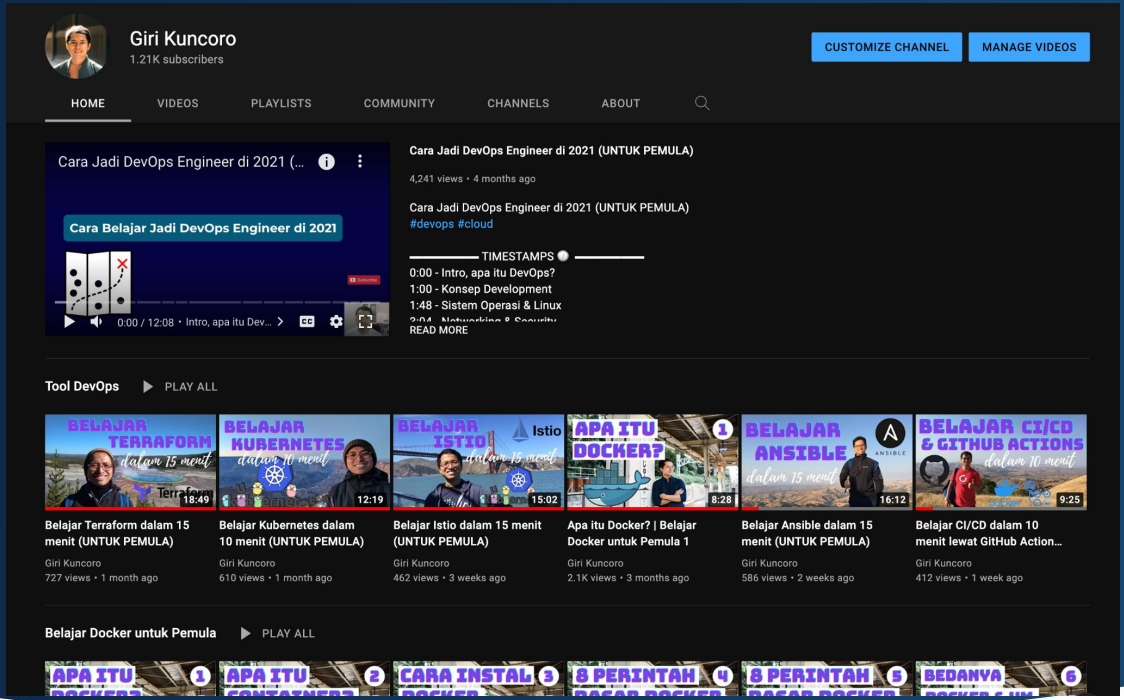
INDONESIA OpenInfra Days

IO Visor Project (iovisor.org)

Brendan Gregg (brendangregg.com)

Youtube Channel

INDONESIA OpenInfra Days



Open Infrastructure FOUNDATION

GiriKuncoro

BRI EasyStack Biznet GioCloud boer technology NVIDIA

Sponsored by:

Open Infrastructure FOUNDATION

NVIDIA®

boer technology

Biznet GioCloud

EasyStack
open cloud computing

BRI

Open Networking Indonesia

imtek

INDOCENTER

Hosted by:

OpenStack Indonesia
Indonesia OpenStack Foundation Community
www.openstack.id

Community Partners:

sdX Indonesia

openSUSE ID

KLAS

gimpscape ID

CLOUD NATIVE INDONESIA

Adaptive Network Laboratory

NETWORK EVOLUTION

Open Infrastructure FOUNDATION

INDONESIA OpenInfra Days

Thanks!

Do you have any questions?

@girikuncoro

GiriKuncoro

Platinum sponsor :

Gold sponsor :

Silver sponsor :

Custom sponsor :