EXCERPT ONLY

LTTng's Trace Filtering and beyond

(with some eBPF goodness, of course!)

Suchakrapani Datt Sharma

TracingSummit (LinuxCon, Seattle), Aug 20, 2015

École Polytechnique de Montréal

Laboratoire DORSAL

eBPF

eBPF

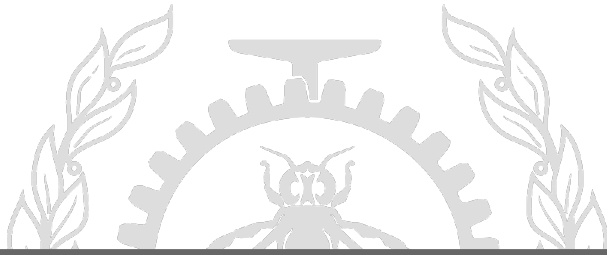
Berkeley Packet Filter (BPF)

- Filter expressions → Bytecode → Interpret
- Fast, small, in-kernel packet & syscall filtering
- Register based, switch-dispatch interpreter

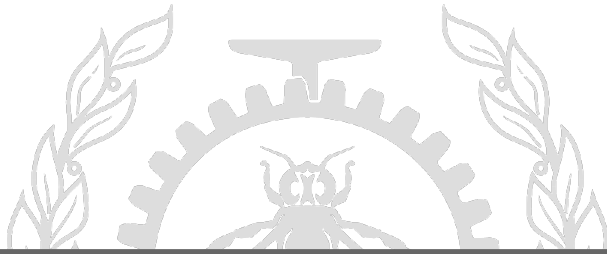
Current Status of BPF

- Extensions for trace filtering (Kprobes!! Kprobes!!)
- More than just filtering. JITed programs - FAST!
- Evolved to *extended* BPF (eBPF)
  - BPF maps, *bpf* syscall - aggregation and userspace access
  - More registers (64 bit), back jumps, tail-calls, verifier

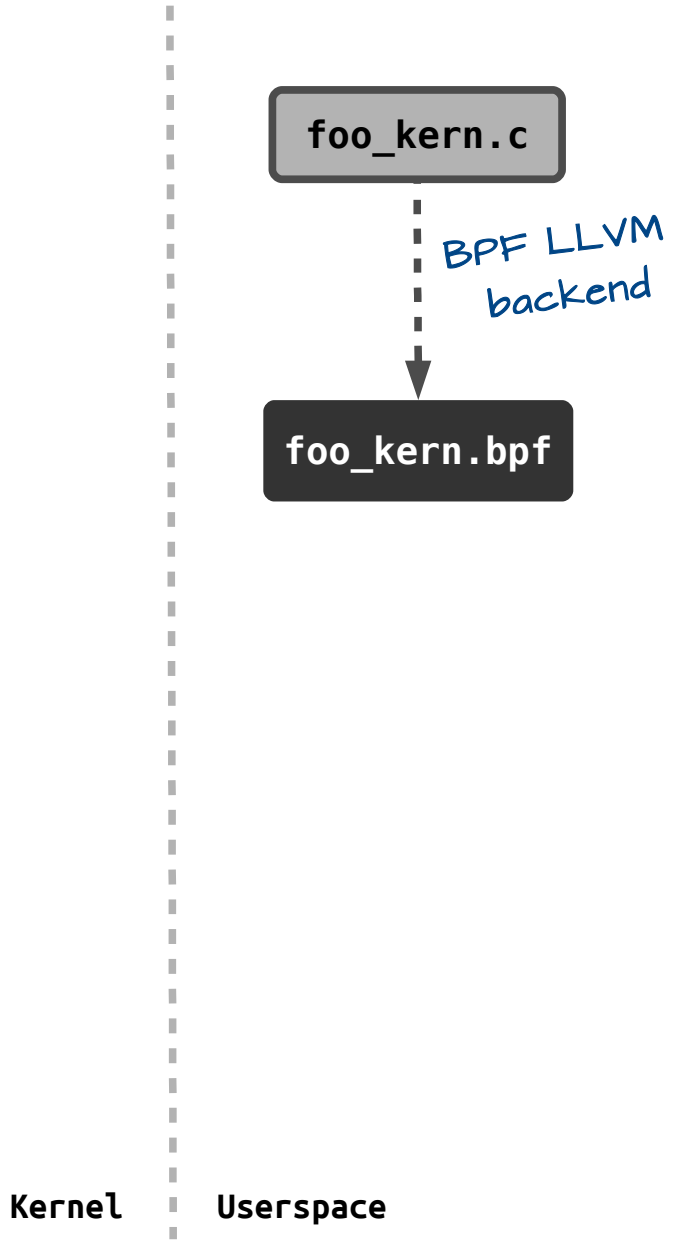
Example BPF Session



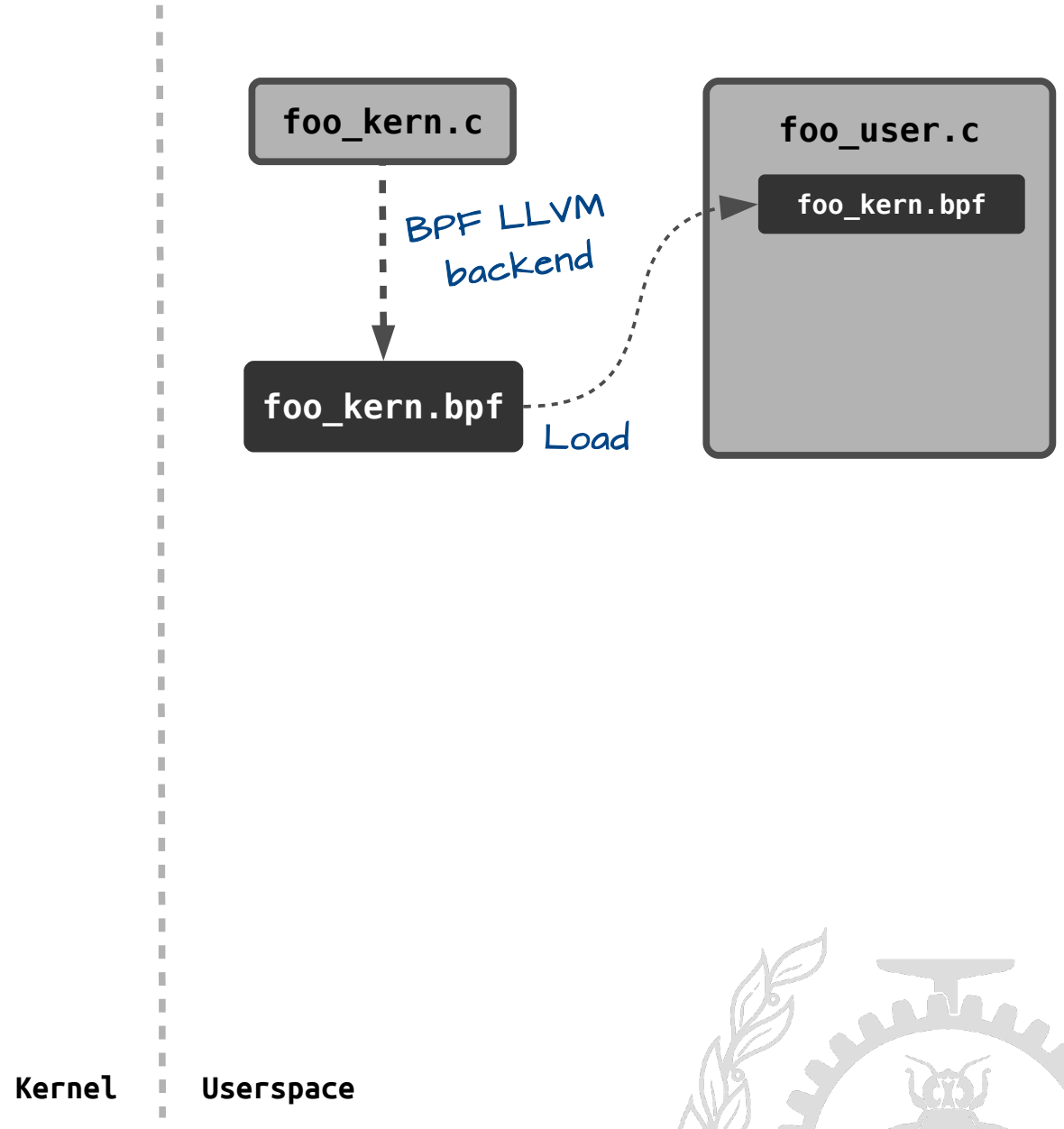
Example eBPF Session



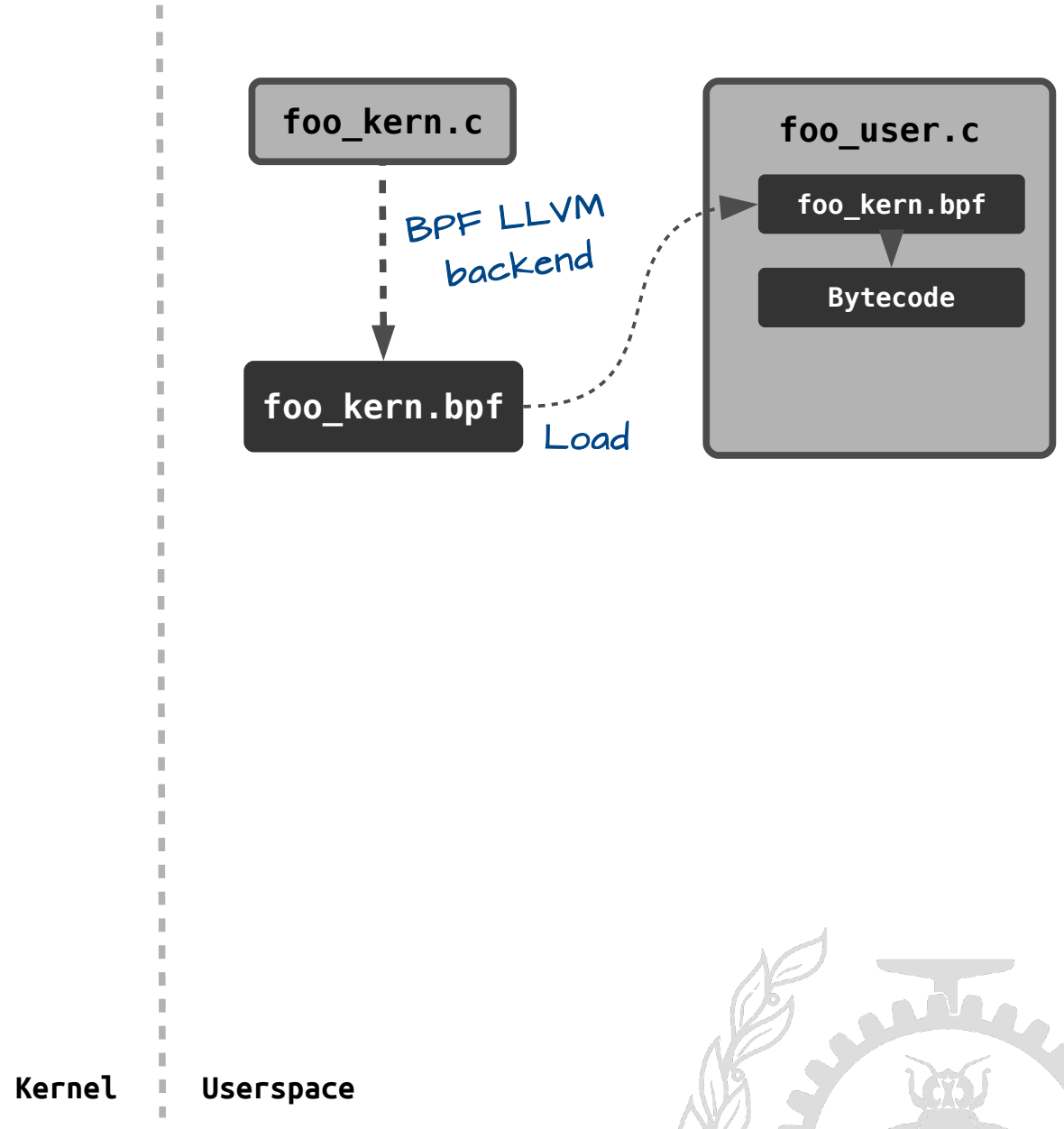
Example eBPF Session



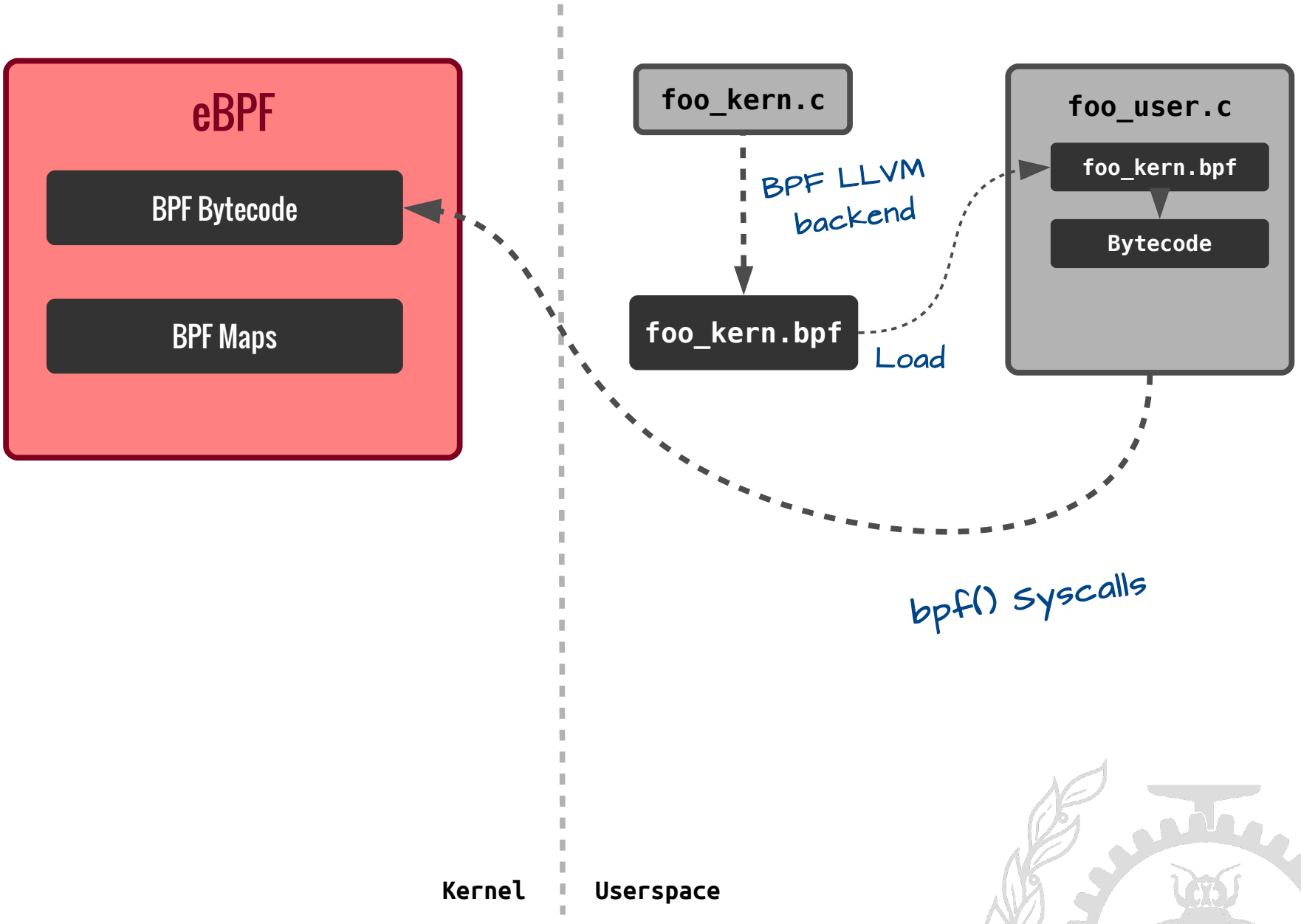
Example eBPF Session



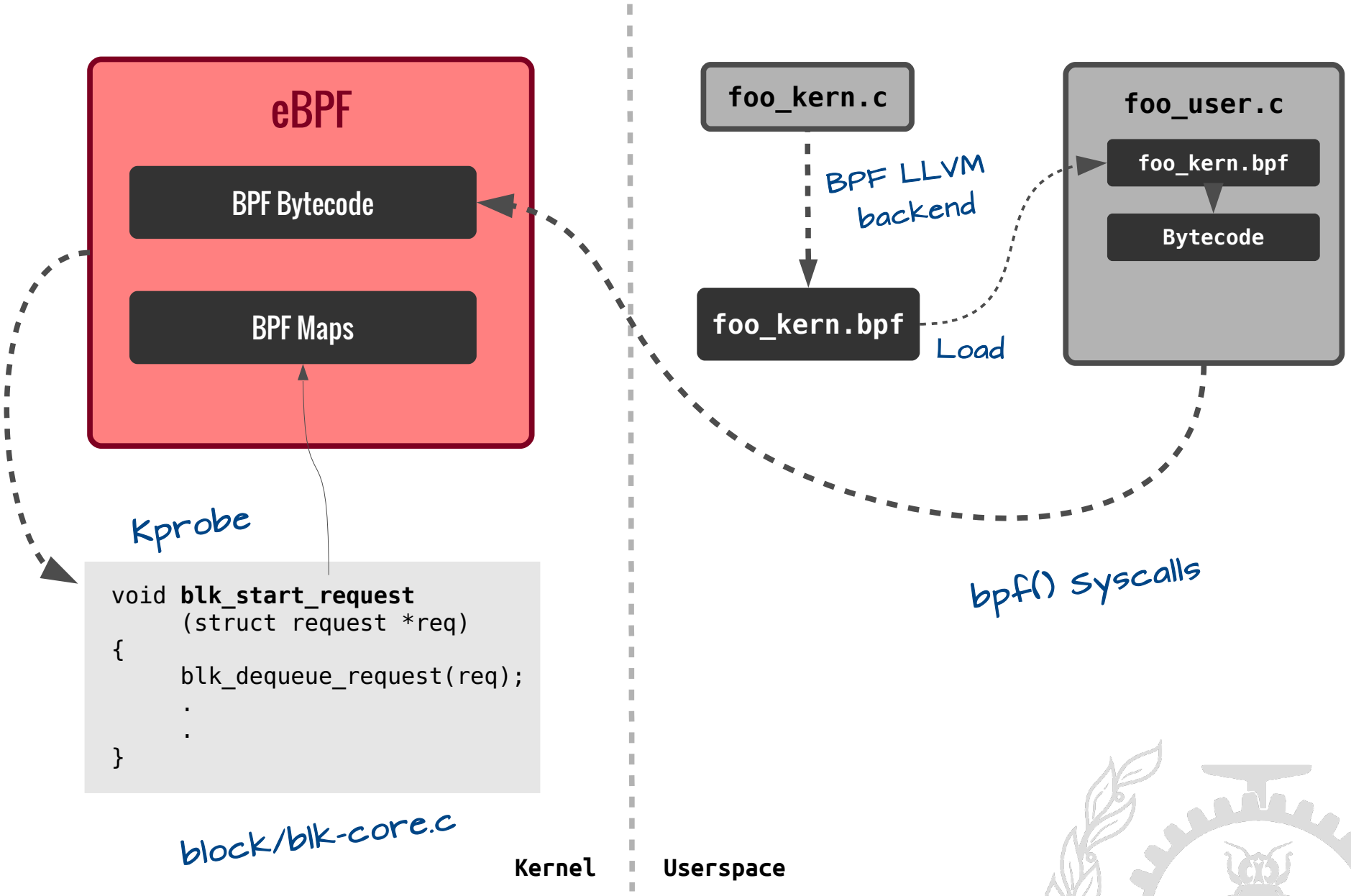
Example eBPF Session



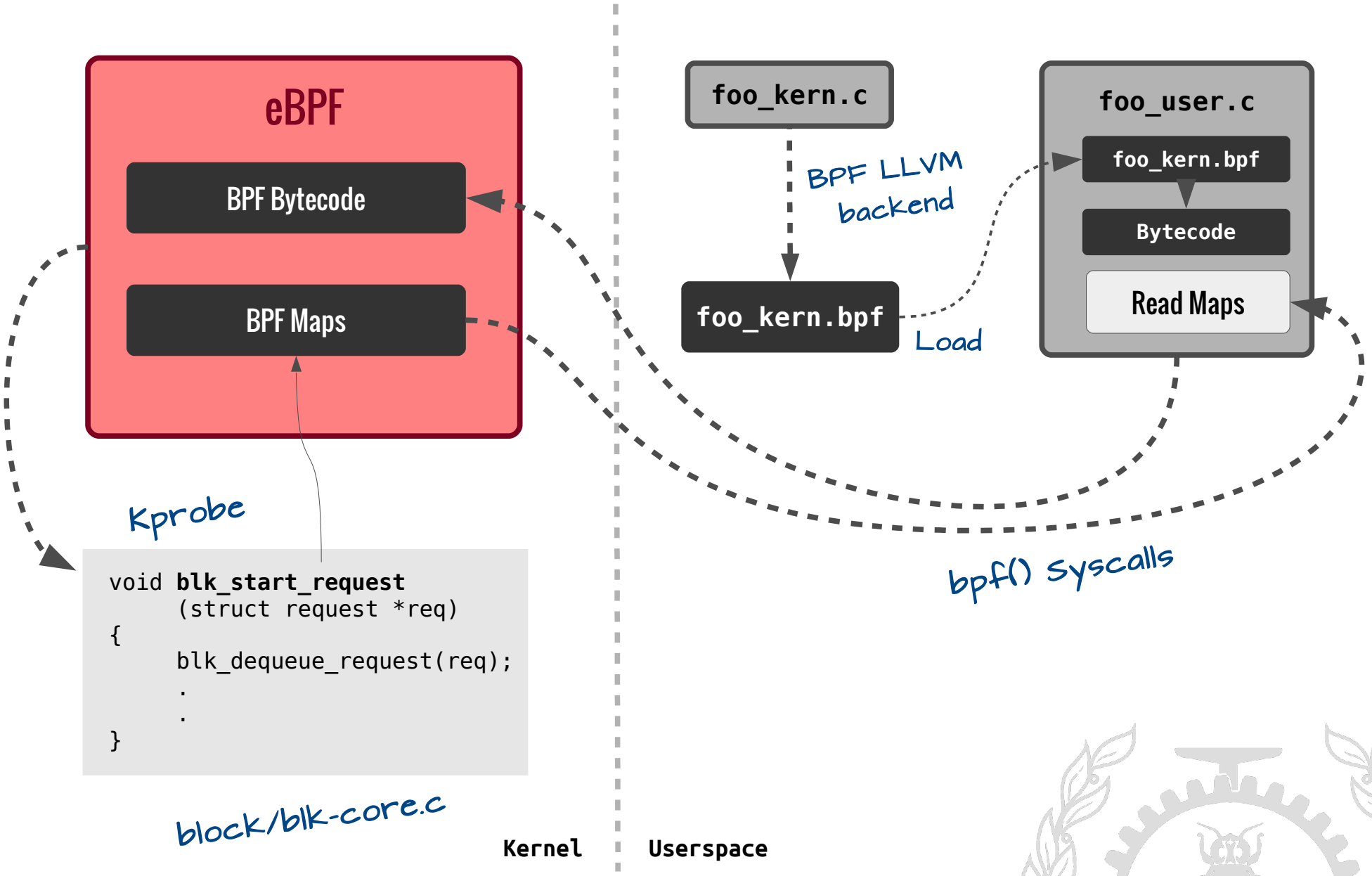
Example eBPF Session



Example eBPF Session

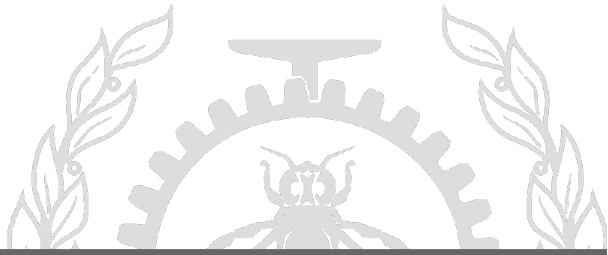


Example eBPF Session

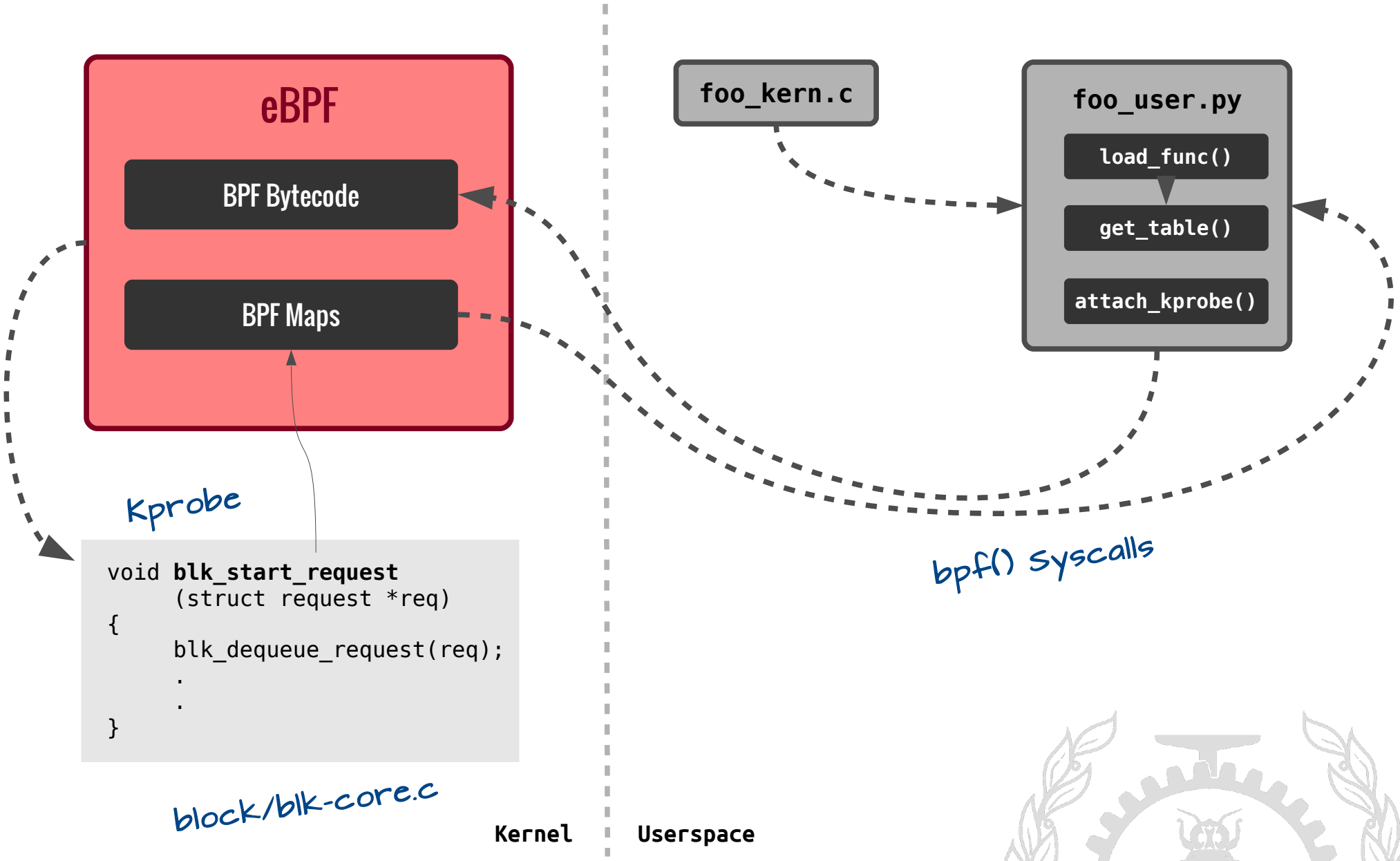


Example BPF session with

https://github.com/iovisor/bcc



Example eBPF Session with bcc



Example eBPF Session with bcc

task_switch.c

```
#include <uapi/linux/ptrace.h>
#include <linux/sched.h>

struct key_t {
  u32 prev_pid;
  u32 curr_pid;
};

BPF_TABLE("hash", struct key_t, u64, stats, 1024);

int count_sched(struct pt_regs *ctx, struct
task_struct *prev) {
  struct key_t key = {};
  u64 zero = 0, *val;

  key.curr_pid = bpf_get_current_pid_tgid();
  key.prev_pid = prev->pid;

  val = stats.lookup_or_init(&key, &zero);
  (*val)++;
  return 0;
}
```

Kernel side BPF program

task_switch.py

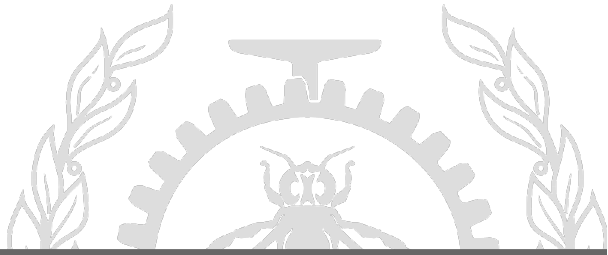
```
from bpf import BPF
from time import sleep

b = BPF(src_file="task_switch.c")
fn = b.load_func("count_sched", BPF.KPROBE)
stats = b.get_table("stats")
BPF.attach_kprobe(fn, "finish_task_switch")

# generate many schedule events
for i in range(0, 100): sleep(0.01)

for k, v in stats.items():
    print("task_switch[%5d->%5d]=%u" %
(k.prev_pid, k.curr_pid, v.value))
```

Userspace side Python program



Questions?

suchakrapani.sharma@polymtl.ca

suchakra on #lttng (irc.oftc.net)

@tuxology

http://suchakra.in