eBPF and tc classifier/actions

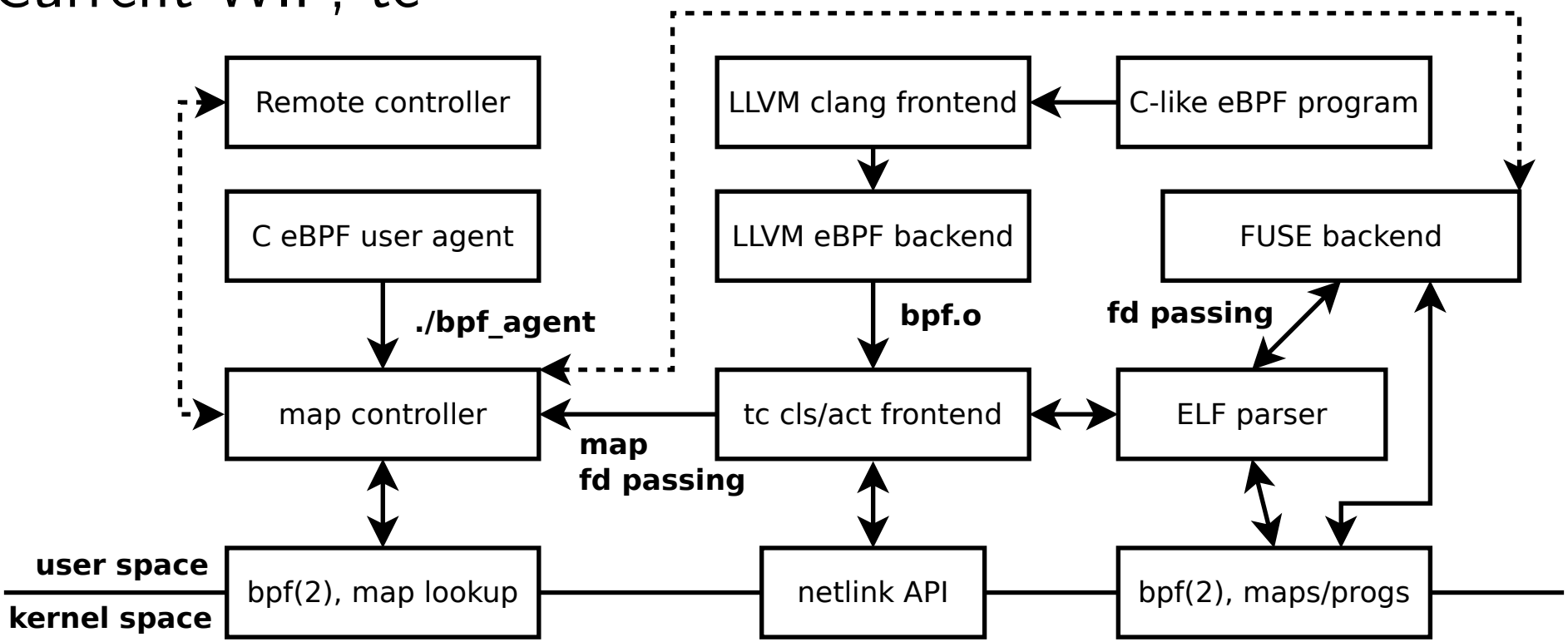
Daniel Borkmann
<daniel@iogearbox.net>

Plumbers, August 21, 2015

(More or less) recent updates

- Kernel
  - Addition of eBPF support in cls_bpf and act_bpf
  - Program invocation under RCU only (act_bpf ready soon, too)
  - Selected skb field readouts/mangling, cgroups support
  - skb data mangling, encap/decap, various new helpers
  - eBPF tail call added, s390 JIT got eBPF support
  - Extensive eBPF test suite updates
  - Partially due to that also lots of fixes
- tc
  - Frontend eBPF support for cls_bpf and act_bpf
  - ELF object parser, map/program loader
  - Map file descriptor handover via UDS
  - BPF exec proxy, BPF agent example
  - Initial tail call support
  - Extensive man pages (`bpf(2)`, `tc-bpf(8)`)

Current WIP, tc



- "Persistent" file descriptors via fuse backend for cls_bpf, act_bpf
- fs structure flexible for use with standard Unix tools, e.g. map access, disasm, ...

```
bcc sharing.c
tc filter add dev foo parent ffff: bpf obj sharing.o sec icls [...]
tc filter add dev foo parent 1:    bpf obj sharing.o sec ecls [...]
```

Open discussion

- New file descriptor “sinks” instead of UDS `SCM_RIGHTS`?
- How can we redirect skbs without cloning (`bpf_clone_redirect()`)?
- BPF “channels” feeding socket data wrt. containers?
- Better test generation against verifier.
- ...

`https://git.breakpoint.cc/cgit/dborkman/iproute2.git/log/?h=ebpf-act2`