

XDP - eXpress Data Path

Dávid Bolvanský

Návrh, správa a bezpečnosť (NSB)

Vysokorychlostné spracovanie paketov - súčasnosť

Dosahujeme milióny prichádzajúcich paketov za sekundu:

- 10 Gbps: 14.8Mpps / 67.5 ns na paket
- 100 Gbps: 148Mpps / 6.75 ns na paket

... avšak spracovanie paketov v operačných systémoch je príliš pomalé

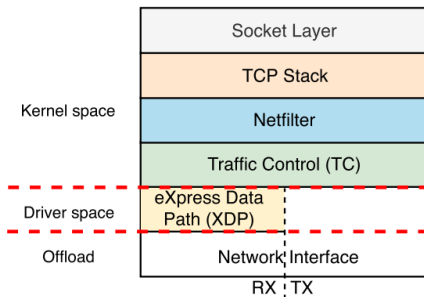
Bežné riešenia:

- Obchádzanie jadra (kernel bypass) - vysoký výkon 😊, zložitá integrácia so systémom ☹
- Fast-path frame-to-userspace riešenia - ovládanie jadrom 😊, nízky výkon ☹
- Vlastné moduly v kerneli – vyhýbanie sa prepínaniu kontextu 😊, náročné na údržbu ☹

Úvod k XDP

Čo je XDP?

- vysokovýkonná, programovateľná sieťová dátová cesta založená na eBPF v Linuxovom jadre, dostupná od verzie 4.8
- spracovanie paketov na najnižšom bode sieťového SW stacku



Obrázok: Sieťový stack v Linuxovom jadre

XDP v Linuxovom jadre

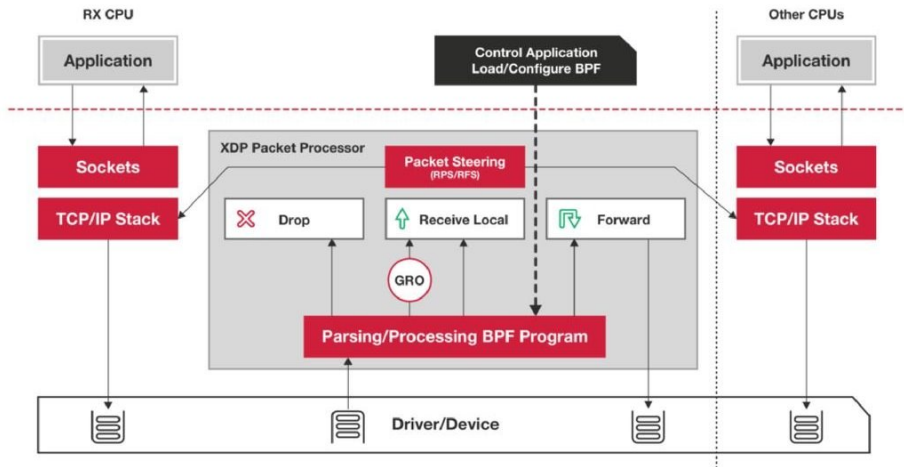
- eXpress Data Path je najnižšia vrstva v sieťovom stacku v Linuxovom jadre. Nachádza sa len na RX ceste, vnútri sieťového ovládača zariadenia, čo umožňuje najskoršie spracovanie paketov v sieťovom stacku, a to ešte pred alokáciou pamäte operačným systémom.
- Poskytuje prípojný bod (hook), na ktorý sa môžu napojiť eBPF programy. V tomto prípojnóm bode sú programy schopné rýchlo sa rozhodovať o prichádzajúcich paketoch a tiež na nich vykonávať ľubovoľné úpravy, čím sa zabraňuje ďalšej rézii spôsobenej spracovaním v jadre.

Kľúčové výhody XDP

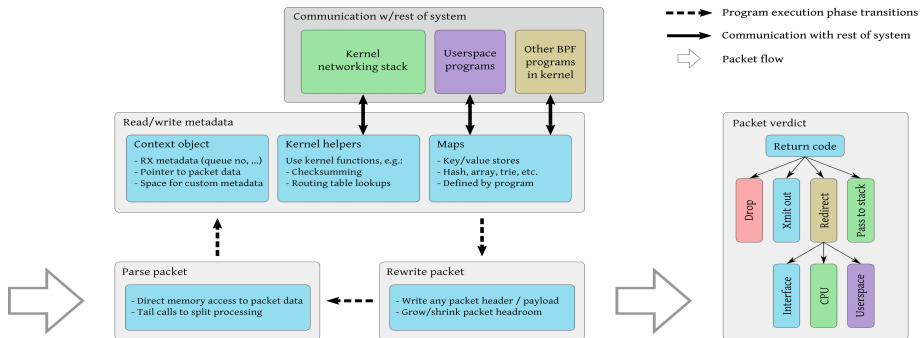
- navrhnuté za účelom poskytovania vysokého výkonu a programovateľnosti (ponúka stabilné API)
- integrované v rámci jadra - ovládač má kontrolu nad hardvérom
- nevyžaduje špecializovaný hardvér
- nevyžaduje obchádzanie jadra (kernel bypass)
- nevyžaduje spätné injektovanie paketov (packet reinjection)
- spolupracuje s TCP/IP stackom, prináša všetky výhody BPF

XDP - priebeh spracovania paketu

XDP: eXpress Data Path



XDP - tok programu



Po príchode paketu sa spúšťa program, ktorý začína analyzovať (parsovať) hlavičky, z ktorých získava informácie, na ktoré bude reagovať. Následne prečíta alebo aktualizuje metadáta z jedného z niekoľkých zdrojov. Nakoniec je možné prepísať paket a rozhodnúť o finálnom verdikte (XDP akcii) nad paketom. Verdikt (XDP akcia) je určený návratovým kódom programu.

- **XDP_ABORTED**

Indikácia chyby eBPF programu, paket je zahodený

- **XDP_DROP**

Zahodenie paketu

- **XDP_PASS**

Predanie paketu ďalej v sieťovom stacku

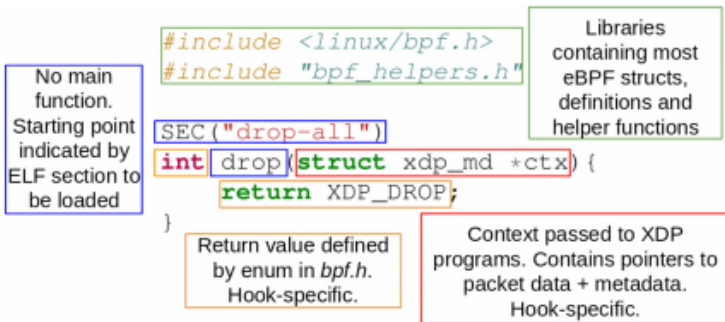
- **XDP_TX**

Prenesenie paketu cez rovnakú sieťovú kartu

- **XDP_REDIRECT**

Presmerovanie paketu na inú sieťovú kartu alebo CPU

XDP - základná štruktúra programu



XDP - ukážka programu

Obsahuje paket IPv4 TCP segment? Ak nie, paket je zahodený.

```
int isTCP( struct xdp_md *ctx ) {
    void *data_end = (void *) (long) ctx->data_end;
    void *data_begin = (void *) (long) ctx->data;
    struct ethhdr *eth = data_begin;

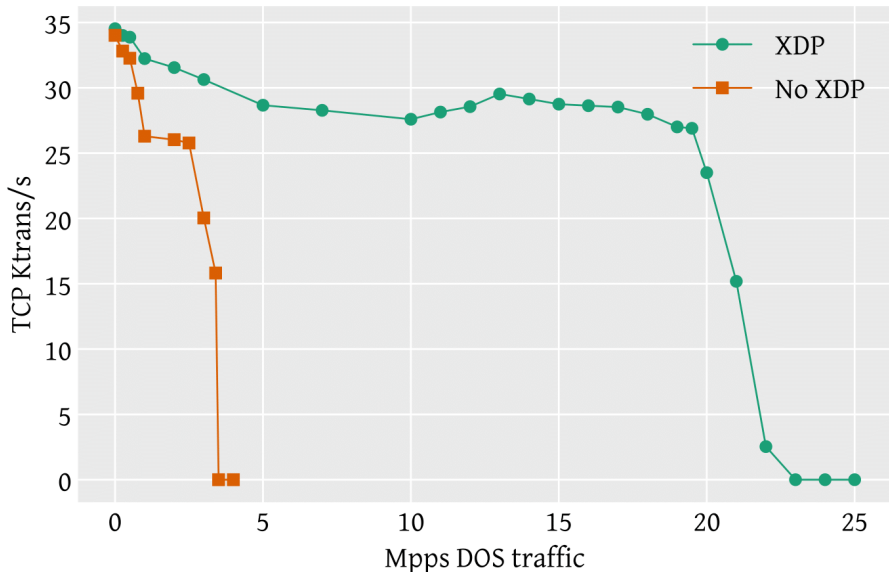
    // Check packet's size
    if(eth + 1 > data_end)
        return XDP_PASS;

    // Check if Ethernet frame has IPv4 packet
    if (eth->h_proto == bpf_htons( ETH_P_IP )) {
        struct iphdr *ipv4 = (struct iphdr *) ( ((void*)eth) + ETH_HLEN );

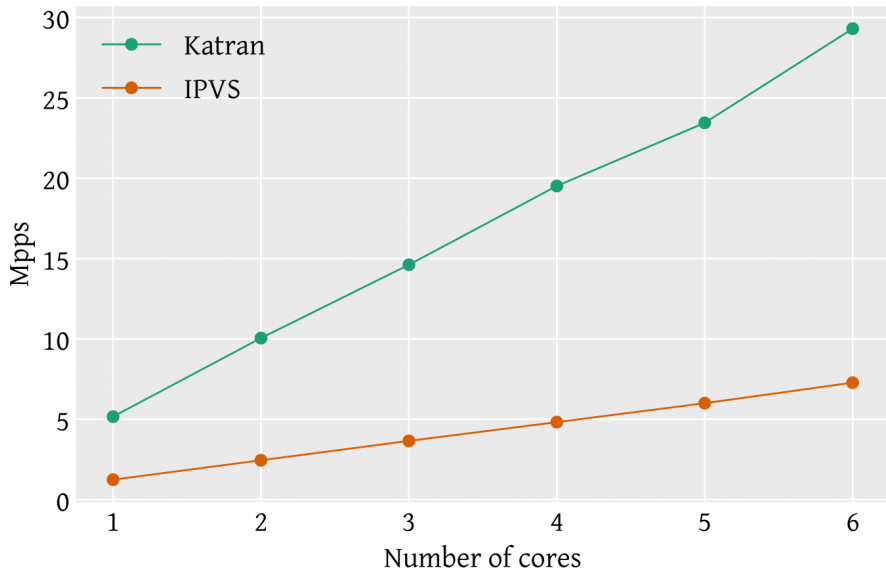
        if(ipv4 + 1 > data_end)
            return XDP_PASS;

        // Check if IPv4 packet contains a TCP segment
        if (ipv4->protocol == IPPROTO_TCP)
            return XDP_PASS;
    }
    return XDP_DROP;
}
```

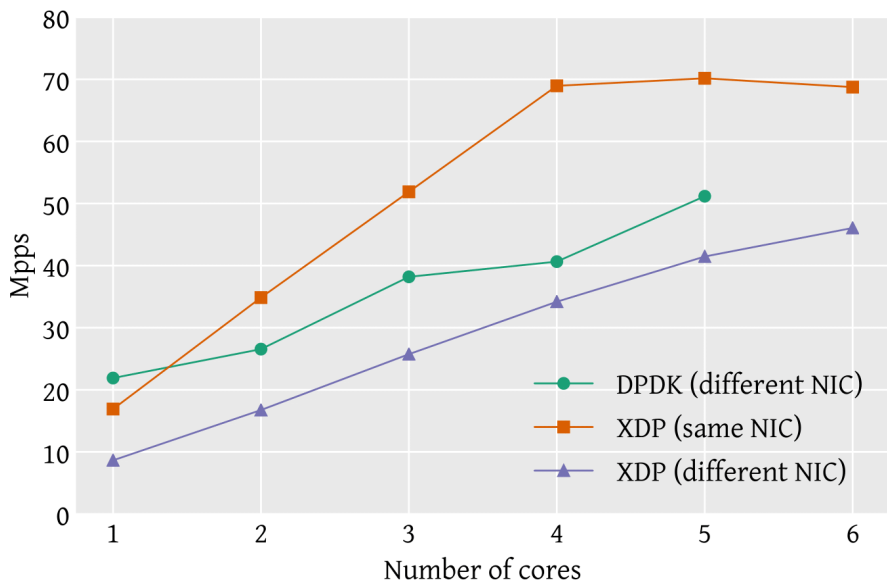
Možnosti použitia XDP - ochrana proti DDoS útokom



Možnosti použitia XDP - vyrovňovanie záťaže



Možnosti použitia XDP - preposielanie paketov



Ďalšie možnosti použitia XDP

- Tunelovanie
- Monitorovacie nástroje
- Hromadné techniky - napr. Generic Receive Offload
- ULP (Upper Layer Protocol) spracovávanie - napr. message delineation

Použité zdroje

- The eXpress Data Path: Fast Programmable Packet Processing in the Operating System Kernel
<https://dl.acm.org/doi/10.1145/3281411.3281443>
- Fast Packet Processing with eBPF and XDP: Concepts, Code, Challenges, and Applications
<https://dl.acm.org/doi/abs/10.1145/3371038>
- XDP - Intro and future use-cases
http://people.netfilter.org/hawk/presentations/xdp2016/xdp_intro_and_use_cases_sep2016.pdf
- XDP - IO Visor Project
www.iovisor.org/technology/xdp
- eBPF / XDP
www.slideshare.net/Netronome/ebpfxdp-sigcomm-2018
- BPF – docs
<https://github.com/iovisor/bpf-docs>