

XDP: Past, Present and Future

Toke Høiland-Jørgensen
Red Hat

XDP: Recently introduced features

- Multibuf (Lozenzo, Maciej, Eelco)
- XDP hints on RX (Jesper, Stanislav, Toke)
- Feature flags (Lorenzo)
- More stack helpers (conntrack, synproxy) (Kartikeya, Lorenzo, Maxim)
- XDP_REDIRECT improvements
 - Multicast (Hangbin)
 - Map lookup, hashmap type, bpf_redirect() performance (Toke)
 - Programs in devmap/cpumap (David Ahern, Lorenzo)
- Eliminating indirect calls (Björn)
- Live mode BPF_PROG_RUN (Toke)
- AF_XDP need_wakeup mode (Magnus and Maxim)
- Atomic replace and bpf_link attachment (Toke, Andrii)

XDP driver support

	Basic	Redirect	ndo_xmit	Multibuf RX	Multibuf TX	XSK ZC	HW offload		Basic	Redirect	ndo_xmit	Multibuf RX	Multibuf TX	XSK ZC	HW offload
atlantic	✓	✓	✓	✓	✓			mlx4	✓	✓					
bnxt	✓	✓	(✓)	✓	(✓)			mlx5	(✓)	(✓)	(✓)	(✓)	(✓)	(✓)	
cpsw	✓	✓	✓					mtk_eth	(✓)	(✓)	(✓)		(✓)		
cpsw_new	✓	✓	✓					mvneta	✓	✓	✓	✓	✓		
dpaa	✓	✓	✓					mvpp2	(✓)	(✓)	(✓)				
dpaa2	✓	✓	✓			✓		netsec	✓	✓	✓				
ena	(✓)	(✓)						nfp	✓					(✓)	(✓)
enetc	✓	✓	✓	✓	✓			octeontx2	✓	✓	(✓)				
fec	(✓)	(✓)						qede	✓	✓	✓				
funeth	✓	✓	(✓)		(✓)			sfc	✓	✓	✓				
gve	(✓)	(✓)	(✓)		(✓)			sfc-siena	✓	✓	✓				
hv_netvsc	✓	✓	✓					stmmac	✓	✓	(✓)			✓	
i40e	✓	✓	(✓)	✓	(✓)	✓		thunder	(✓)						
ice	✓	✓	(✓)	✓	(✓)	✓		tsnep	✓	✓	✓		✓	✓	
igb	✓	✓	(✓)		(✓)			tun	(✓)	(✓)	(✓)				
igc	✓	✓	(✓)		(✓)	✓		veth	(✓)	(✓)	(✓)	(✓)	(✓)		
ixgbe	✓	✓	(✓)		(✓)	✓		virtio_net	✓	✓	(✓)	(✓)	(✓)		
ixgbevf	✓							vmxnet3	✓	✓	✓				
lan966x	(✓)	(✓)	(✓)					xen-	✓	✓	✓				
mana	✓	✓	✓					netfront							

✓: Always enabled. (✓): Configuration-dependent.

XDP: Ongoing work

- XDP hints on TX (Stanislav)
- Veth optimisations (Jesper)
- Multiprog attachment (Daniel)
- Queueing (Toke)

Generic XDP

- Bulking on redirect
- GSO frames through multibuf API
- Veth redirect conversion

AF_XDP xmit path

- Getting rid of socket-allocated SKBs?
- Why allocate SKBs at all?
- TX hints (ongoing, Stanislav)

XDP Hints

- More metadata fields
- Saving for redirect and skb creation (in xdp_frame)
- Naming: “XDP metadata” is not very googlable!

Datapath helpers

What do we need to build a transparent fast path using XDP?

- Bridge lookup(?)
- Netfilter / flowtables lookup (or acceleration?)

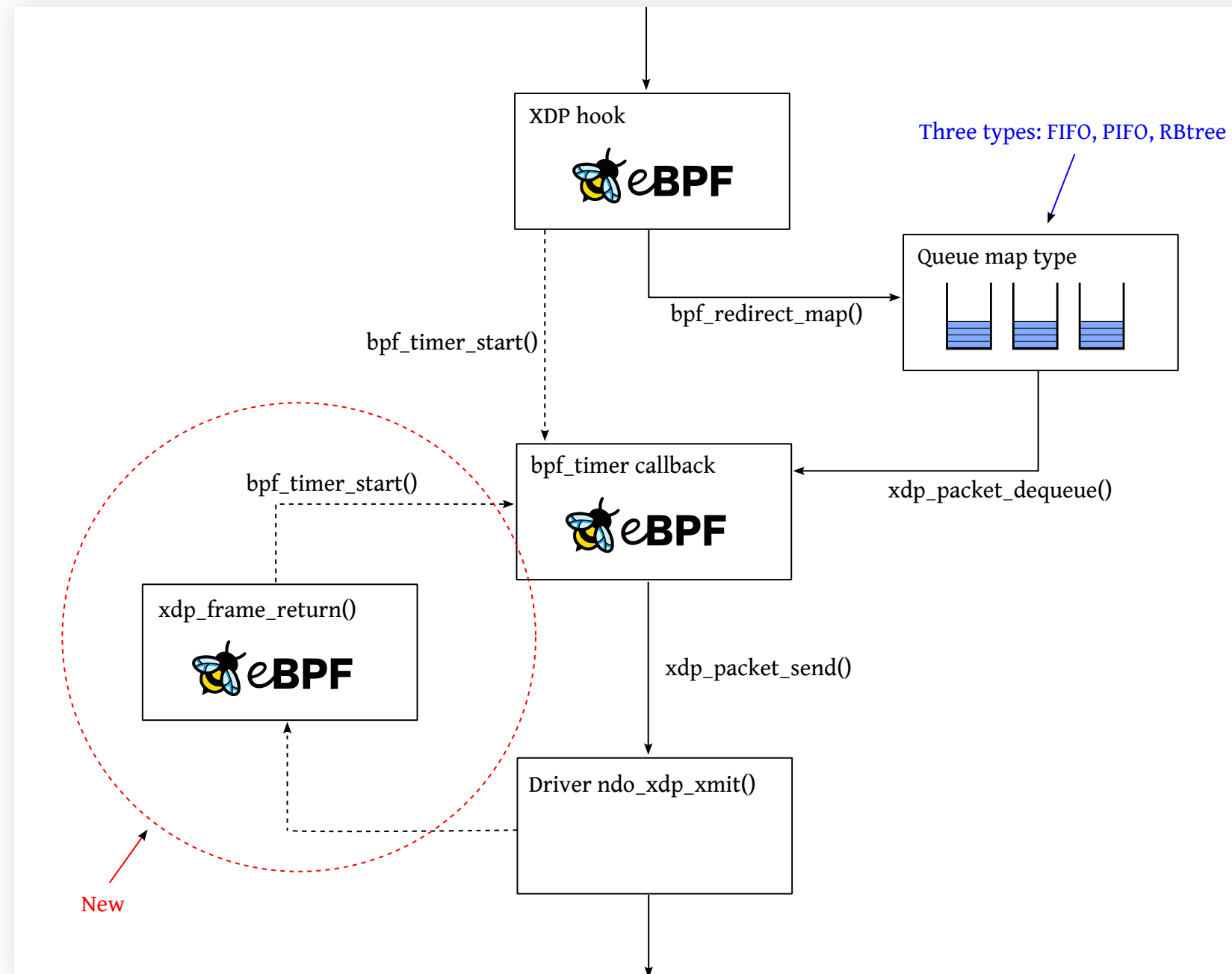
XDP queueing

Last presented at LPC 2022: <https://lpc.events/event/16/contributions/1351/>

Progress stalled a bit since then, but currently working on:

- Rebase on top of bpf-next in progress (convert to kfuncs, dynptr)
- Finding a way to control HW queue depth (like BQL)

XDP queueing - high-level diagram



XDP queueing - outstanding issues

- Context of the callback function
 - Softirq (which?) or move to kthread (like cpumap)?
 - Minimising overhead of callback
- Validating HWQ depth limiting approach
- Handling competition with netstack?
- Which type(s) of queue map are really useful?
- Finish rebase, lots more validation and testing