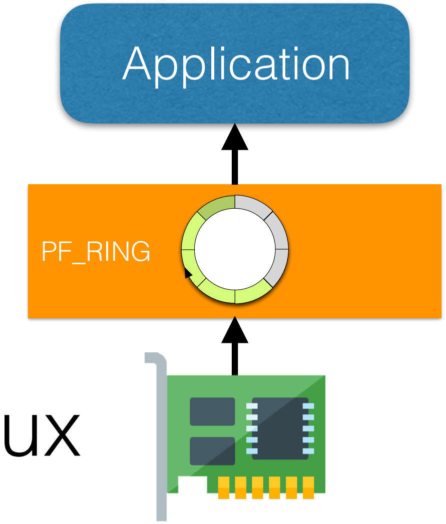
ntop MiniConf 2021
What's new in PF_RING 8.x

Alfredo Cardigliano
cardigliano@ntop.org

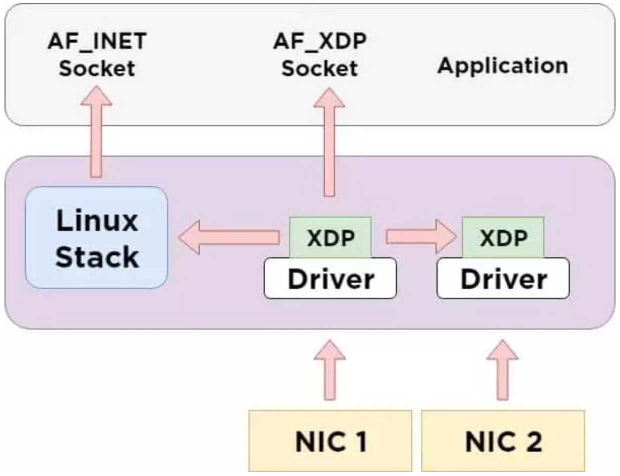
Introduction



- PF_RING packet capture SDK
  - Any commodity adapter supported (linux performance)
  - Accelerated Zero Copy drivers (PF_RING ZC) for Intel commodity adapters
  - Support for specialized FPGA adapters (Napatech, Silicom/Fiberblaze, Accolade, and many others)

XDP

- eXpress Data Path
- Programmable (eBPF), high-performance packet processing in the Linux kernel
- Actions: drop, send back, modify, pass to the kernel, deliver to an application
- AF_XDP socket for packet capture



AF_XDP Performance

- Copy mode for legacy drivers
- Zero Copy mode supported by many Linux drivers today
- Slower than full kernel bypass technologies (kernel is still involved), but much faster than vanilla drivers
- In our tests (Xeon E3):
  - Single queue: 7 Mpps
  - 4 RSS queues: 15 Mpps (10 Gbit)

AF_XDP Integration

- PF_RING 8.0 includes enhanced AF_XDP support:
  - Full Zero Copy buffers management
  - Batch capture (introduced also a new PF_RING API)
  - Performance improvements
- Not as fast as PF_RING ZC drivers (capable of 15-20 Mpps on a single core), but a good option for adapters which are not supported by PF_RING ZC

Let's Recap

- Linux drivers (any adapter)
  - Up to 2-3 Gbps
- XDP drivers (any adapter with Zero Copy drivers)
  - Up to 10 Gbps, big average packet size
- Intel adapters with PF_RING ZC drivers
  - 10+ Gbps any packet size
  - Up to 100 Gbps with real-life traffic and RSS (Intel E810 introduced last year)
- FPGA adapters
  - 100 Gbps any packet size

Mellanox/NVIDIA Adapters

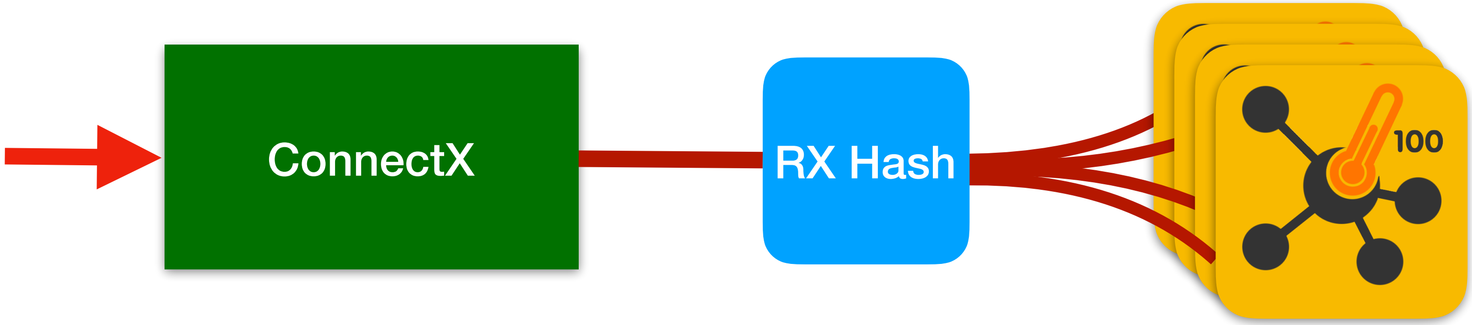
- Low cost commodity adapters (same price range as Intel)
- 1/10/25/40/50/100/200 Gbit
- Hardware offloads:
  - Load-balancing (RSS)
  - Traffic duplication
  - Packet filtering
  - Nanosecond timestamps

PF_RING ZC for Mellanox

- New Zero Copy driver for Mellanox adapters
- Introduced in PF_RING 8.1
- Supported adapters: ConnectX 4/5/6
- Native driver:
  - Mellanox was already supported via AF_XDP, but this delivers way better performance
  - Direct access to all hardware offload capabilities

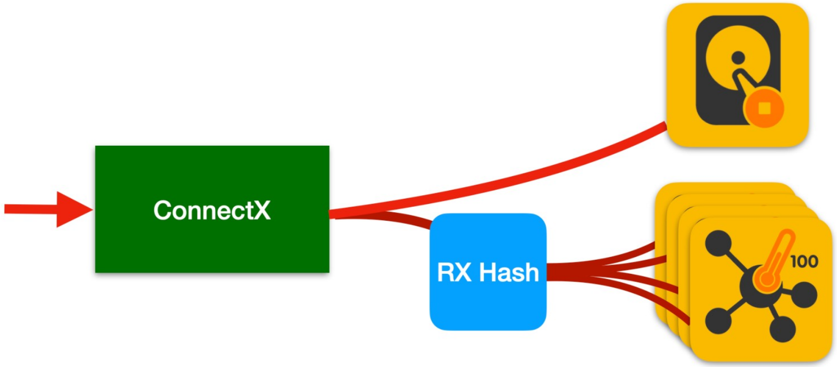
Load-Balancing (RSS)

- Load balance traffic to multiple queues/cores
- Similar to RSS on Intel
- Constraint: multithreaded applications only
- Example: suitable for nProbe Cento to scale the performance up to 100 Gbps



Traffic Duplication

- Native in-hardware packet duplication (open the same interface multiple times), not available on Intel
- Start nProbe Cento and n2disk on the same interface (they both receive the same packets), with different load-balancing configurations:
    - Load-balance to 8 RSS queues for nProbe Cento
    - Single queue for n2disk (to avoid shuffling packets)



Packet Filtering

- Flexible in-hardware packet filtering (combination of all common header fields, rule priority, ...)
- Up to 64k rules
- Rules are per application: nProbe Cento can instruct the adapter to receive all traffic, while n2disk discards in hardware all traffic which is not relevant
- Automatically generate hardware rules from **BPF** filters (e.g. "*dst host 10.0.0.1 and port 80*")

Performance

- Single core capture on Xeon Gold: 32 Mpps
  - 20 Gbps with worst-case 60-byte packets
  - 40 Gbps with an average packet size of 128 bytes
- Multiple cores (RSS): 100 Gbps line-rate
- Real application performance (**nProbe Cento**)
  - **100 Gbps** with 16 cores
  - 40 Gbps with 4 cores

What's next?

- Packets captured with PF_RING do not carry metadata like user and application that produced the traffic (relevant when doing security analysis)
- Adding support for process and user information in nProbe (SRC_PROC_PID, SRC_PROC_NAME, SRC_PROC_USER_NAME, SRC_PROC_PACKAGE_NAME, ..)
- Use PF_RING as SDK for capturing system events for connections, sockets and related information like process and user

n2disk (Continuous Recording)

- In the last year..
  - Improved integration with ntopng
    - Ability to drill down and extract traffic (PCAP) recorded by n2disk
    - Ability to export flows to ntopng to provide visibility on recorded traces (PF_RING FT and nDPI support)
  - Traffic indexing and extraction by source Device and Port ID (provided by Arista switches)
  - Improved PCAP management and automation with external scripts
- What's next
  - Ability to export flows to ClickHouse (compatible with ntopng)
  - PCAP data encryption at-rest

nScrub (DDoS Mitigation)

- In the last year..
  - Improved attackers and (huge) white/black lists management
  - Support non Intel/ZC interfaces (XDP, Mellanox, FPGAs)
  - Support for AMD systems (cost-effective boxes with AMD and Mellanox)
  - Extended policies (e.g. IPSEC support)
- What's next
  - Improve the integration with ntopng and other applications
  - Encrypted, authenticated, fast channel for rules injection
  - Smart mitigation engaging: mitigate traffic towards the actual victim only, when configuring a huge subnet (e.g. ISPs)

Thank You