Linux Piter 2019

Comparison of eBPF, XDP and DPDK for packet inspection

Marian Marinov <mm@yuhu.biz>
Chief System Architect of SiteGround

SiteGround

Who am I?

- **Who am I?**

- Chief System Architect of SiteGround.com
- Sysadmin since 1996
- Organizer of OpenFest, BG Perl Workshops, LUG-BG, RailsGirls and others
- Teaching Network Security and Linux System Administration courses in Sofia University and SoftUni

SiteGround

Why do we need this?

Frequency of DoS/DDoS attacks to our infrastructure

- 4-10 Gbps 6-8 times a month
- 10-40 Gbps maybe 2-3 times a month
- 100+ Gbps around 2 times a month

SiteGround

More stats

Attacks resulting in service degradation:

- for the past 276 days we had 31 DDoS attacks
- some of the months, no attacks
- but some months, up to 9

- 2019 – 31 attacks
- 2018 – 75 attacks
- 2017 – 69 attacks
- 2016 – 84 attacks

Note: I have manually counted the e-mails. The numbers can be slightly inaccurate.

SiteGround

Most attacks are basic

- 20k pps toward ISC Bind can consume **up to 30 CPU cores**
- a child can generate that on its laptop, at home

SiteGround

General solutions

- Buy additional bandwidth
- Buy a very expensive scrubbing device

OR

- Offload this task to other companies, like CloudFlare

SiteGround

Hosted solution issues

- Not every DataCenter is wiling to invest in these devices
- Shared devices
- Attacks can be larger then the capacity of the device
- Larger attacks almost always result in null route
- Attacks saturating the uplinks can affect other machines in the rack and/or row

SiteGround

Cloud solution issues

- You have to point your DNS to the service provider
- Controlling your DNS is now only API based
- Large DNS updates become an issue
- Not suitable for hosting companies

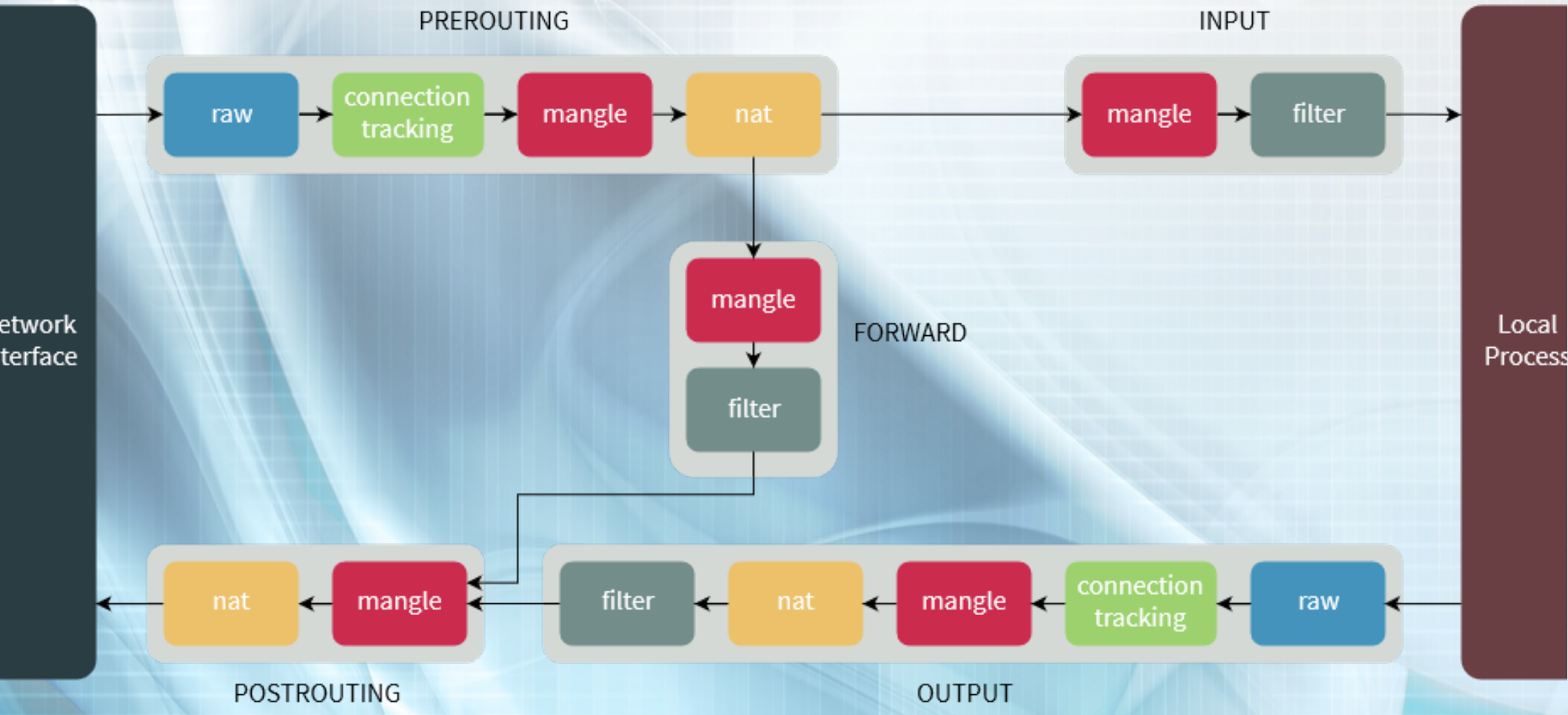
SiteGround

Requirements?

- Build a VM that can handle 10Gbps with ~8Mpps
  - Why a VM?
- scrub UDP DNS and NTP traffic
- scrub TCP traffic by implementing SYN cookies
- scrub all unrelated traffic
- cache HTTP responses(wishful thinking) :)

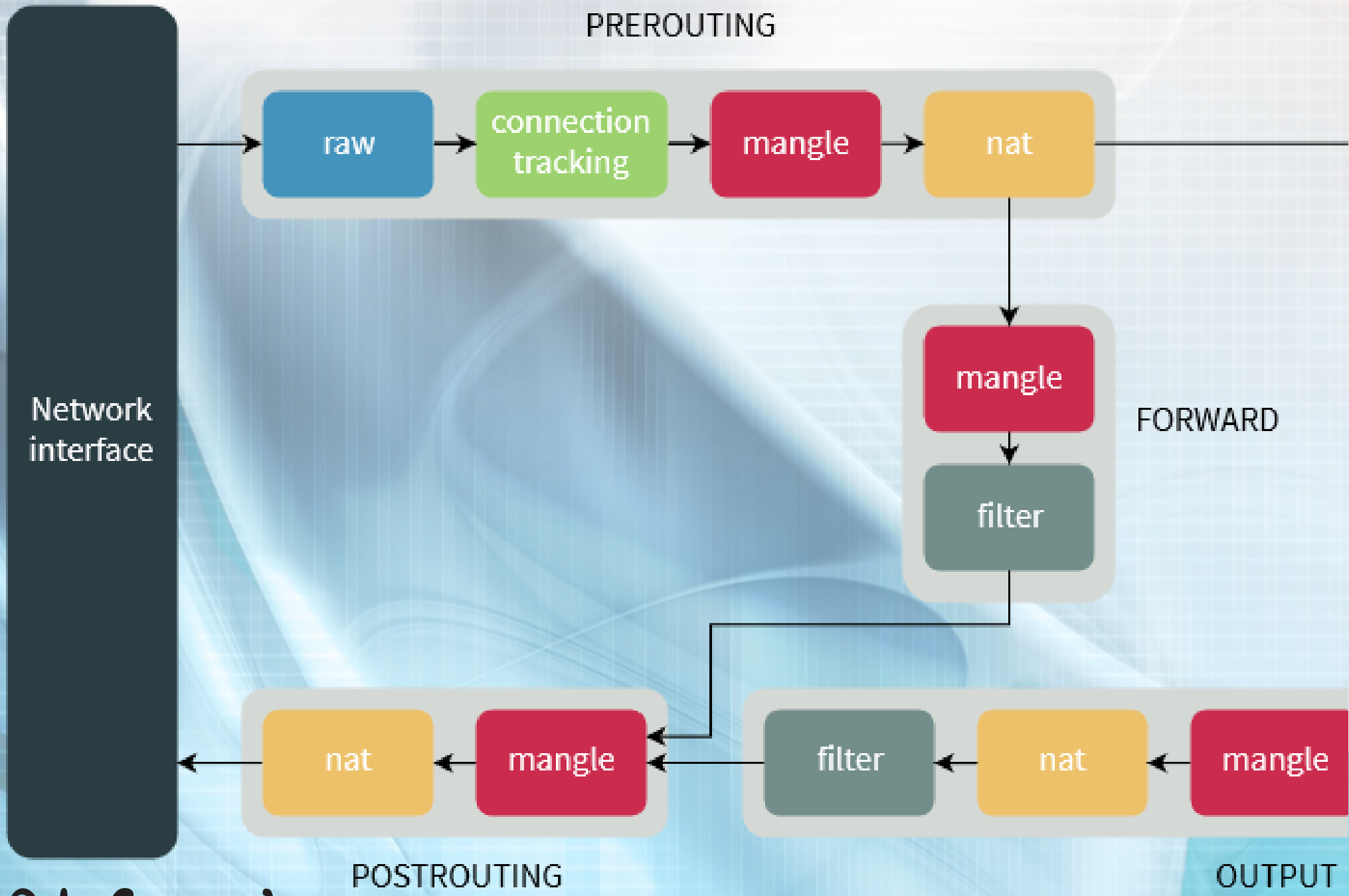
SiteGround

Linux Network Flow



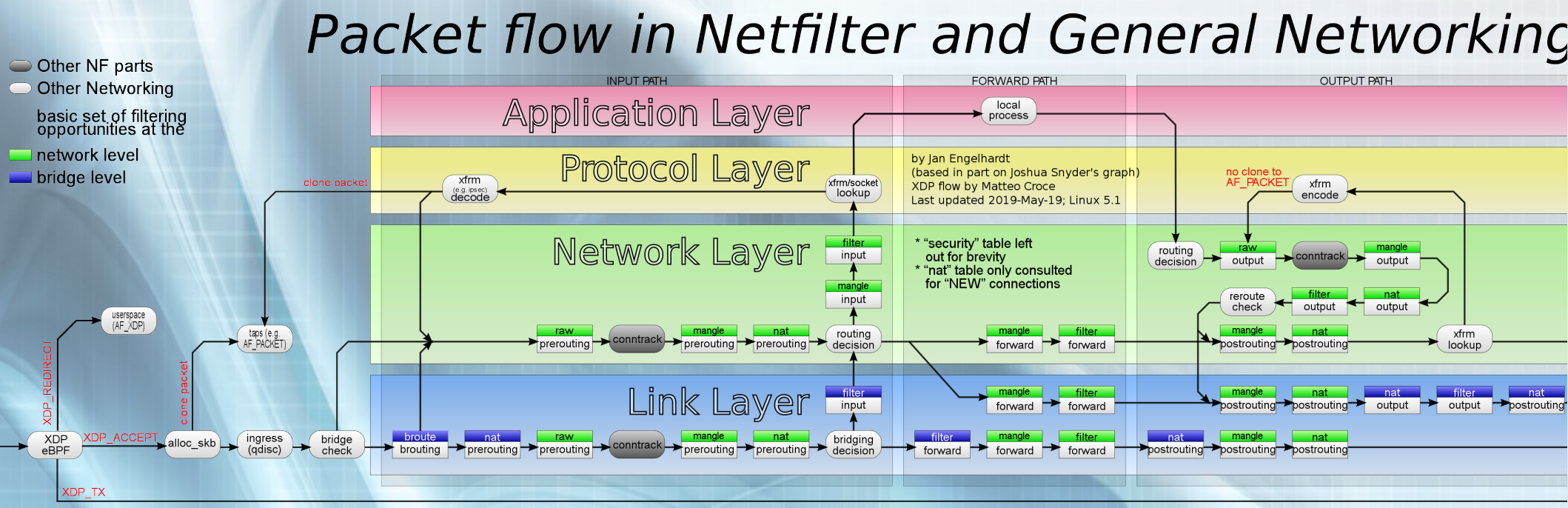
SiteGround

Linux Network Flow



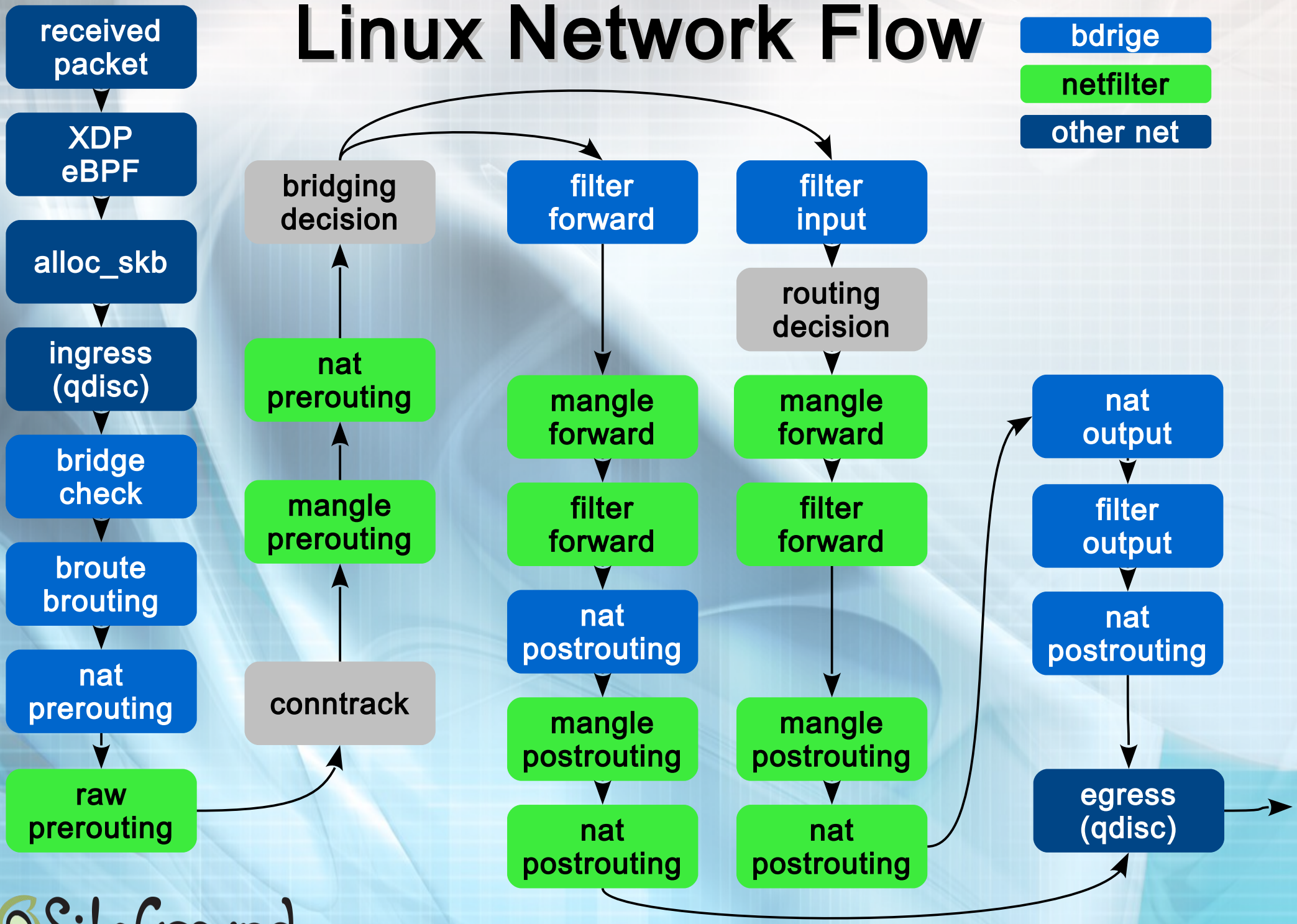
SiteGround

Linux Network Flow



SiteGround

Linux Network Flow



SiteGround

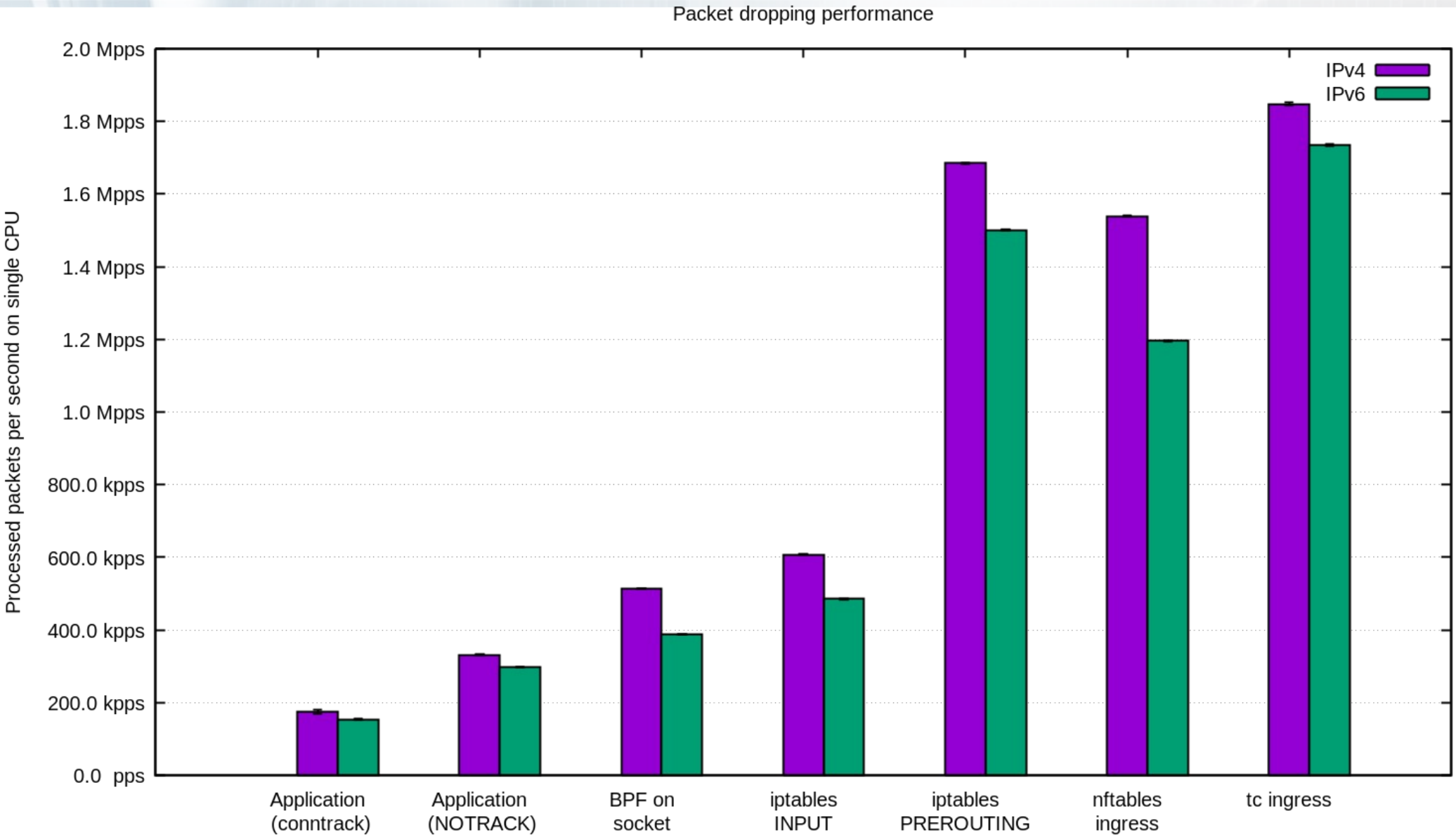
10M packet drop

- in 2018 CloudFlare published the article: **How to drop 10m packets**
- I confirm their results with a few additions:

  - iptables can drop at best 2m pps
    Note: with only one entry in the PREROUTING chain of the mangle table
  - heaving multiple entries in that chain easily becomes a problem
  - even if you use ipset with that, you have a big problem when updating that information

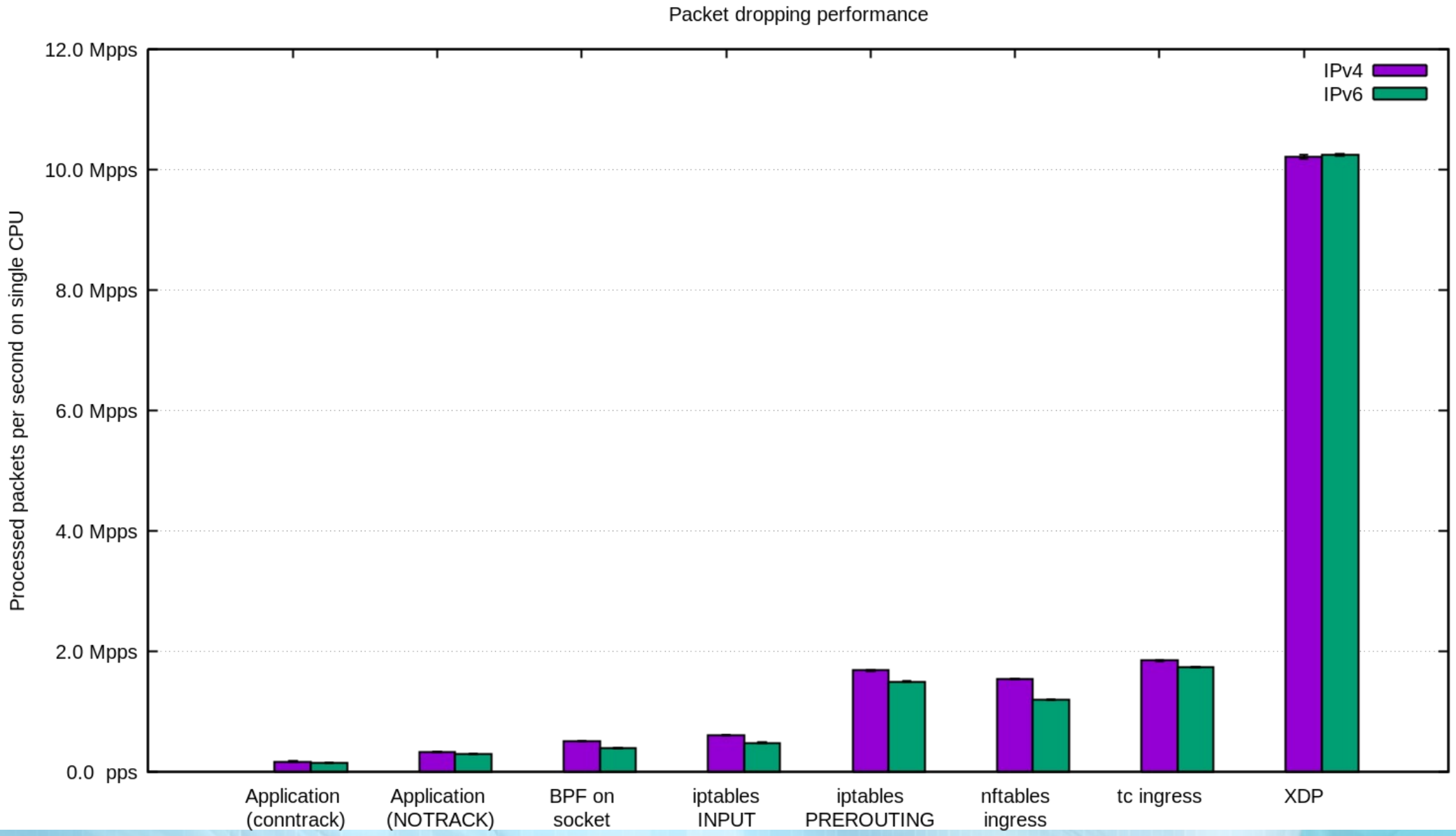
SiteGround

CloudFlare results



SiteGround

CloudFlare results XDP



SiteGround

10M packet drop

- CloudFlare demo code can be found on **GitHub**

SiteGround

So, how I started?

- I already knew about XDP
- But I decided to be “smart ass” and wrote an iptables module…

- It could handle between 260k and 280k pps

SiteGround

Not good enough... eBPF

- I also knew I can use eBPF for that...
  - from the talk of Daniel Borkmann from FOSDEM 2016

- It was better, but not enough...
  - 320-350k pps drop rate
  - with 2000 domains and UDP packet checking
  - no checksums thou

SiteGround

Data
Plane
Development
Kit

DPDK

- I had previous experience with DPDK
- So I ordered one Intel and one SolarFlare NICs
- With both I managed to drop anything that was below the 10G limit of the cards
- With SolarFlare I even tested uploading code into the NIC it self

SiteGround

Data
Plane
Development
Kit

Complex DPDK

- Nobody, except me, was interested in supporting DPDK code
- Writing and updating DPDK is not trivial
- DPDK required specific HW that may not be available in the DataCenter

SiteGround

Data
Plane
Development
Kit

DPDK and P4

- A friend(Boyan Krosnov) told me about P4
- P4 made updating the logic and content of the filter program a lot simpler for me...

SiteGround

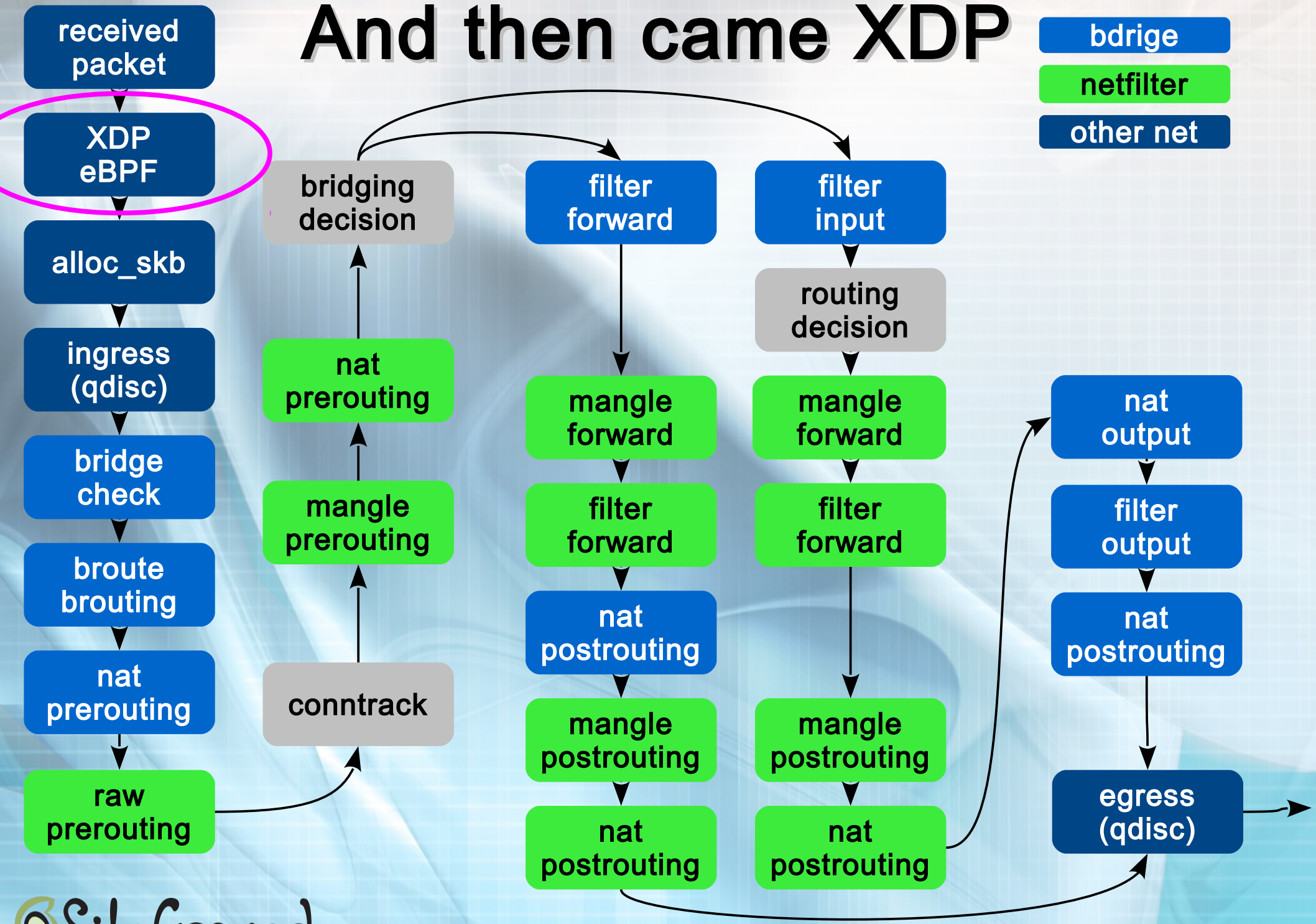
Data
Plane
Development
Kit

P4 and people

- P what?
- If we were to use DPDK with P4, everyone had to learn the language :(

SiteGround

And then came XDP



SiteGround

eXpress
Data
Path

And then came XDP

- Extremely fast and closest to the NIC, same as DPDK
- Supported by many drivers
- Extendable with eBPF functions
- Developed by Jasper Brouer

SiteGround

What I ended up, with?

- A filter similar to what CF did with their DROP example
- instead of a comparing a single prefix, I'm extracting the UDP data if the packet is UDP
- then the extracted data is compared with a BPF map
- I wrote a simple user space tool, that updates the map in the kernel
- voila I had a fast scrubber

the UDP scrubber

- if the DNS request is not for a domain that is within the list in the map I drop the packet
- ToDo: add caching of responses with TTL

the TCP scrubber

This is where I had to stop :(

- compare the packet's dst port and allow it only if it is:
  - SYN to a port that is allowed
    - send and receive SYN cookies here
  - part of already existing connection by examining its own db of tuples and the supplied by the user space(other VMs)

the TCP scrubber

This is where I had to stop :(

- It should handle the SYN cookie for the servers behind and replay the initial SYN if correct SYN,ACK is received

Testing the bastard

I knew I was able to drop packets fast...
But I needed a proof ;)

- I had a talk with Jasper at Linux Plumbers 2019
- He pointed me to
his patched version of pktgen on GitHub :)

SiteGround

Now...

How to get from 10Gbps
to 200Gbps?

SiteGround

Now...

How to get from 10Gbps to 200Gbps?

- Combining multiple VMs with ECMP
  - I did that directly on the switch :)

SiteGround

Links

How to drop 10 million packets per second
https://blog.cloudflare.com/how-to-drop-10-million-packets/
https://github.com/cloudflare/cloudflare-blog/tree/master/2018-07-dropping-packets

XDP tutorial
https://github.com/xdp-project/xdp-tutorial
More XDP materials:
https://www.iovisor.org/technology/xdp

Enhanced pktgen by Jasper
https://github.com/netoptimizer/network-testing

SiteGround

Links

Linux tc and eBPF
https://archive.fosdem.org/2016/schedule/event/ebpf/attachments/slides/1159/export/events/attachments/ebpf/slides/1159/ebpf.pdf
man pages
  http://man7.org/linux/man-pages/man8/tc-bpf.8.html
  http://man7.org/linux/man-pages/man2/bpf.2.html

SolarFlare AOR firmware development kit
https://www.colfaxdirect.com/store/pc/viewPrd.asp?idproduct=1585
Data Plane Development Kit
  https://www.dpdk.org/
P4 Language Specification
  https://p4.org/p4-spec/docs/P4-16-v1.0.0-spec.pdf
P4 meets DPDK
https://www.dpdk.org/wp-content/uploads/sites/35/2017/09/DPDK-Userspace2017-Day2-12-SANDOR_LAKI-T4P4S.pdf

SiteGround

SiteGround

Thank you!

SiteGround