Disaggregated software load balancing
Criteo network

Clos matrix (RFC 7938)

Simple

Scalable
Old edge pod

Network services in edge pod

Big proprietary LB

Only ECMP

Not scalable!
Load balancing disaggregation

Load balancer on servers

Up to 90 servers by DC
ECMP on a Clos matrix

Not easy to do equal cost
ECMP on a Clos matrix

BGP aggregation
L4 LB roles?

- Weighted load balancing
- Keep TCP sessions alive
- Consistent hashing (Maglev)
- Handle L7 LBs maintenance (traffic drain)

We need L4 LBs
L4 LB

- Software
  - IPVS (data plane)
  - Keepalived (control plane)
  - FRR (BGP route injection)
- Direct server return (IPIP encapsulation)
- 21 Gbps by device (25G NIC)
- 4,5M PPS
- Careful placement due to ECMP constraints
VXLAN (RFC 7348)
Software VTEP
How it started

How it's going
Questions ?
Thank you!

Cédric Paillet

c.paillet@criteo.com