CRITEO

Disaggregated software load balancing

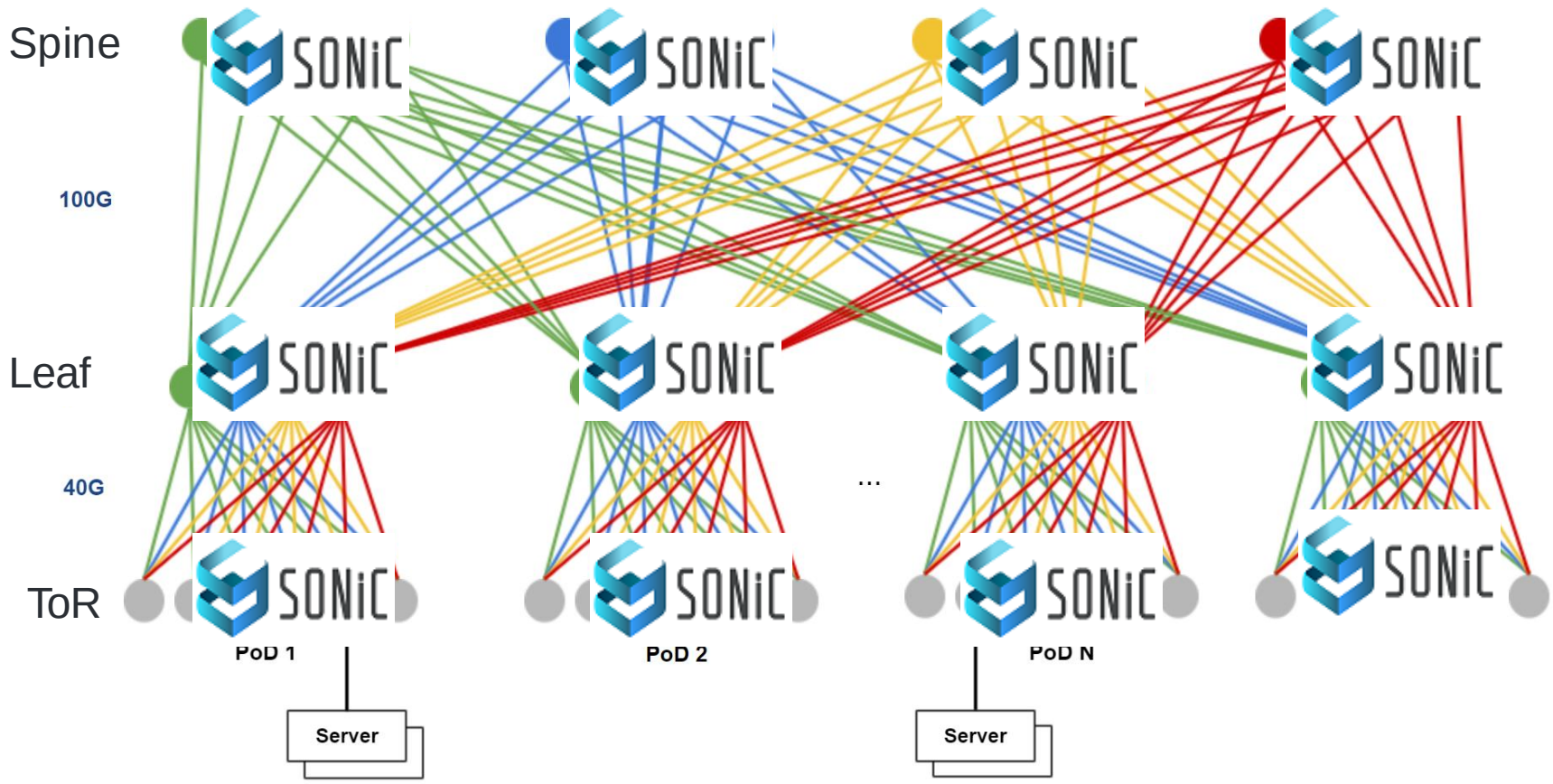
Cédric Paillet

Criteo network

Clos matrix (RFC 7938)

Simple

Scalable



CRITEO

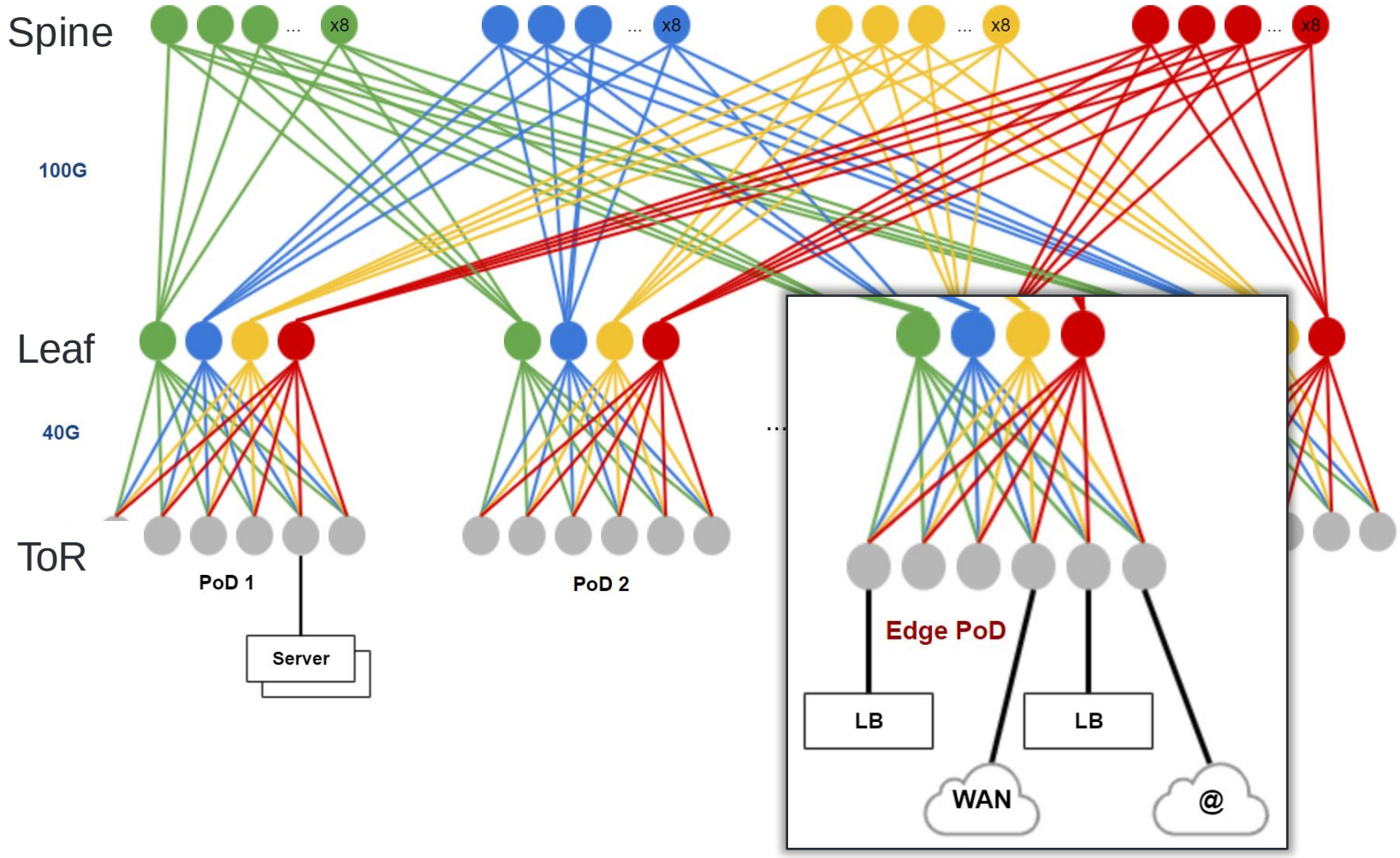
Old edge pod

Network services in edge pod

Big proprietary LB

Only ECMP

Not scalable!

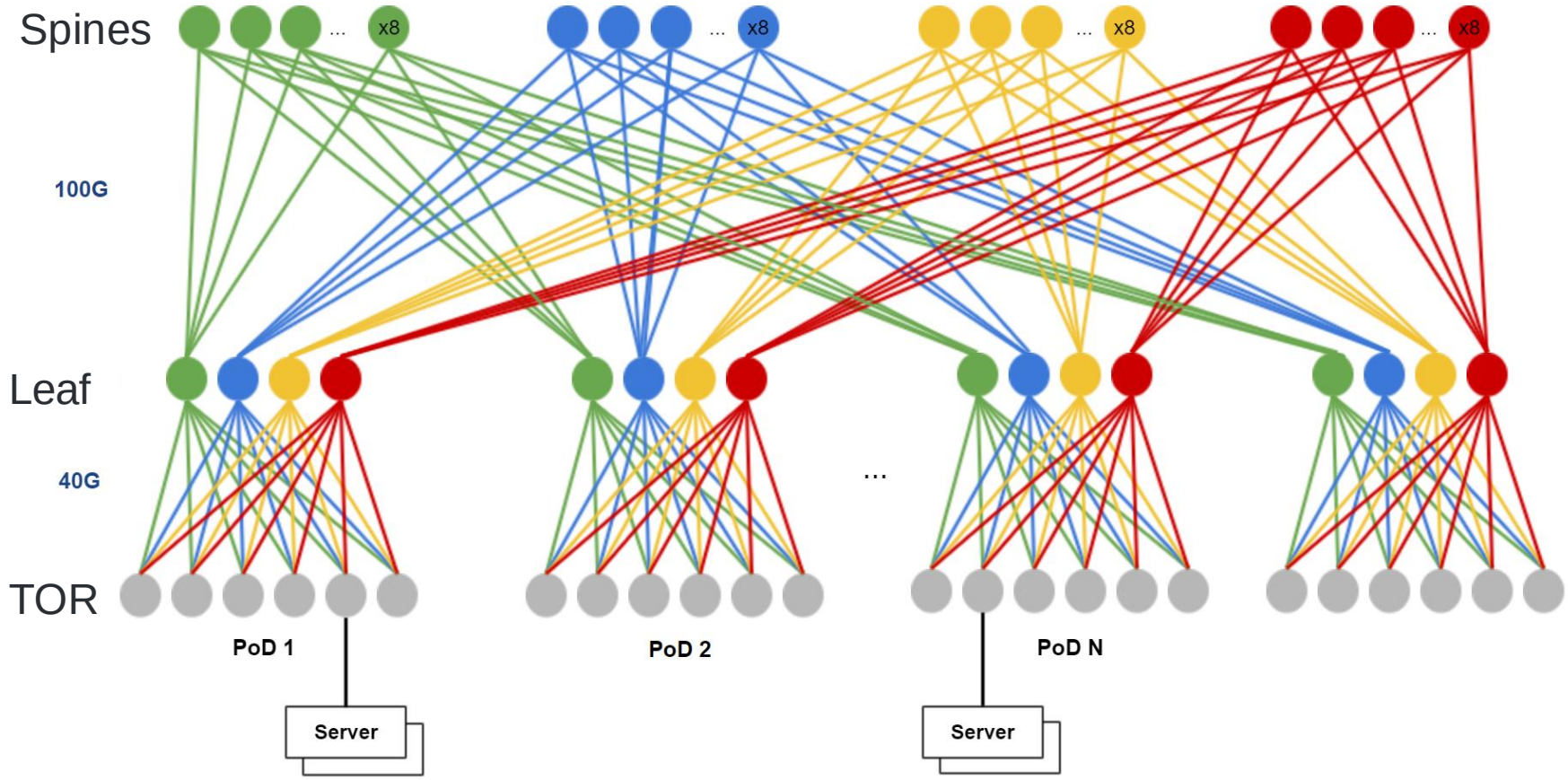


Load balancing disaggregation

Load balancer on servers

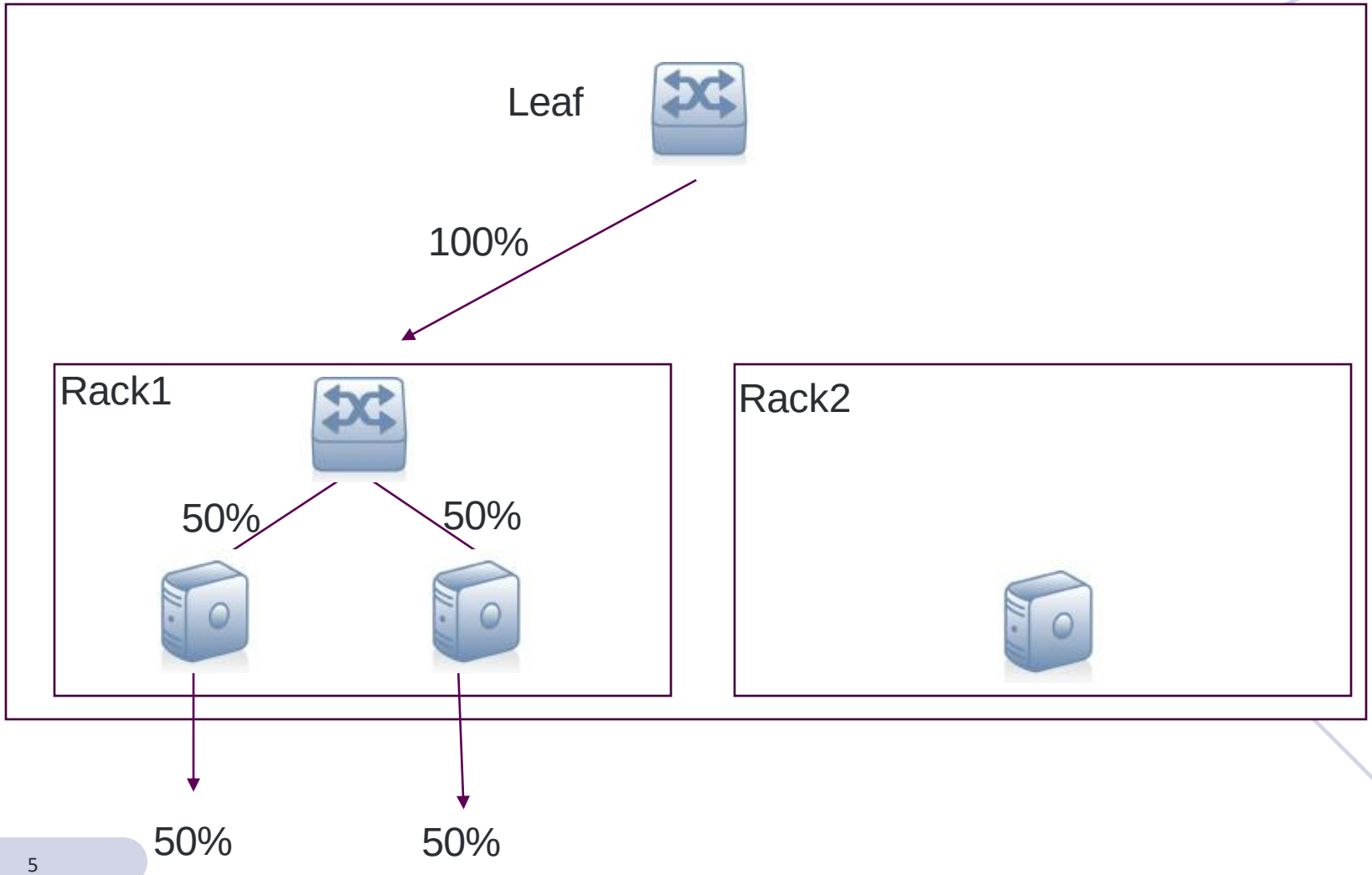
HAPROXY

Up to 90 servers by DC



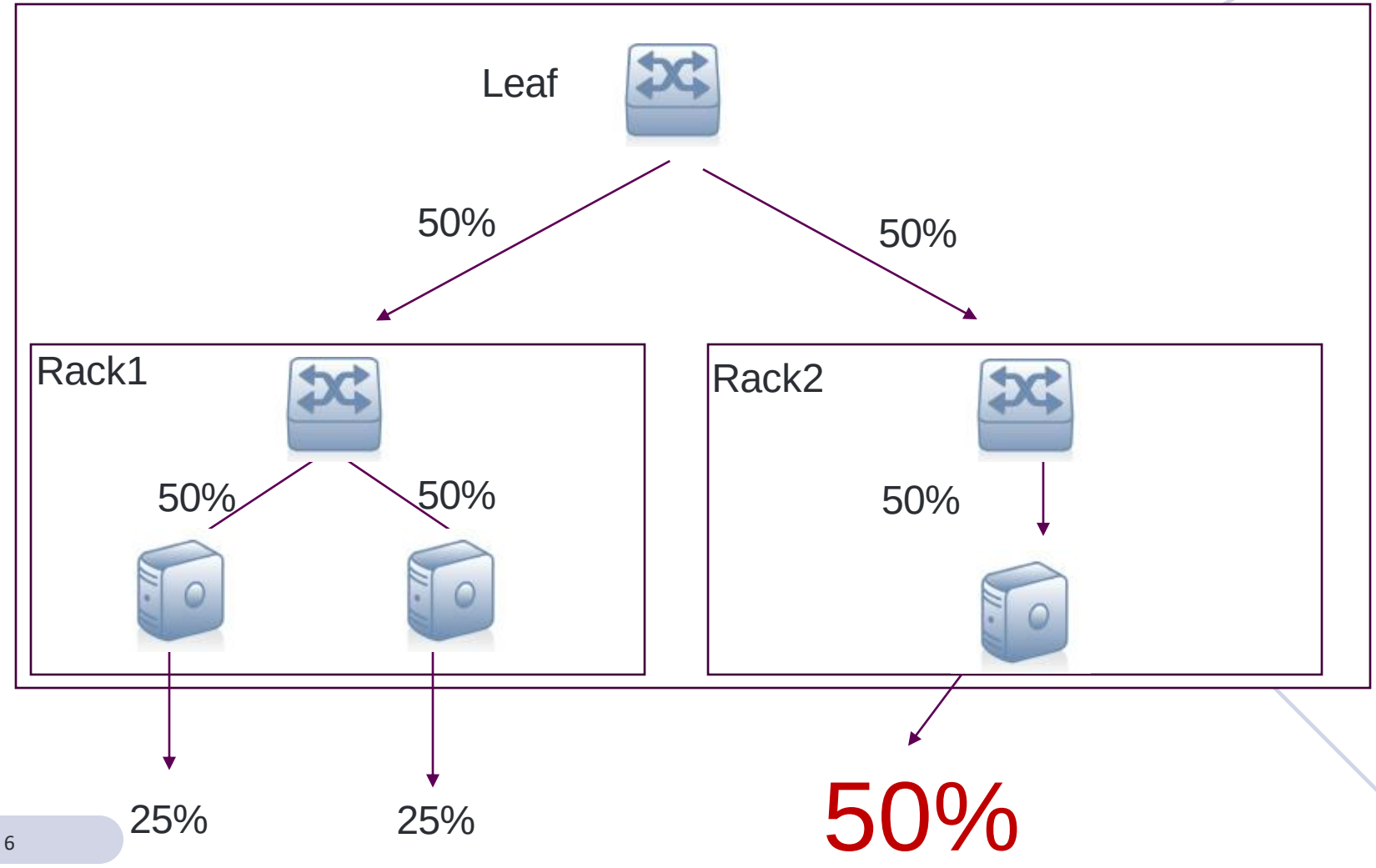
CRITEO

ECMP on a Clos matrix



Not easy to do equal cost

ECMP on a Clos matrix



BGP aggregation

CRITEO

L4 LB roles ?

Weighted load balancing

Keep TCP sessions alive

Consistent hashing (Maglev)

Handle L7 LBs maintenance (traffic drain)

We need L4 LBs

CRITEO

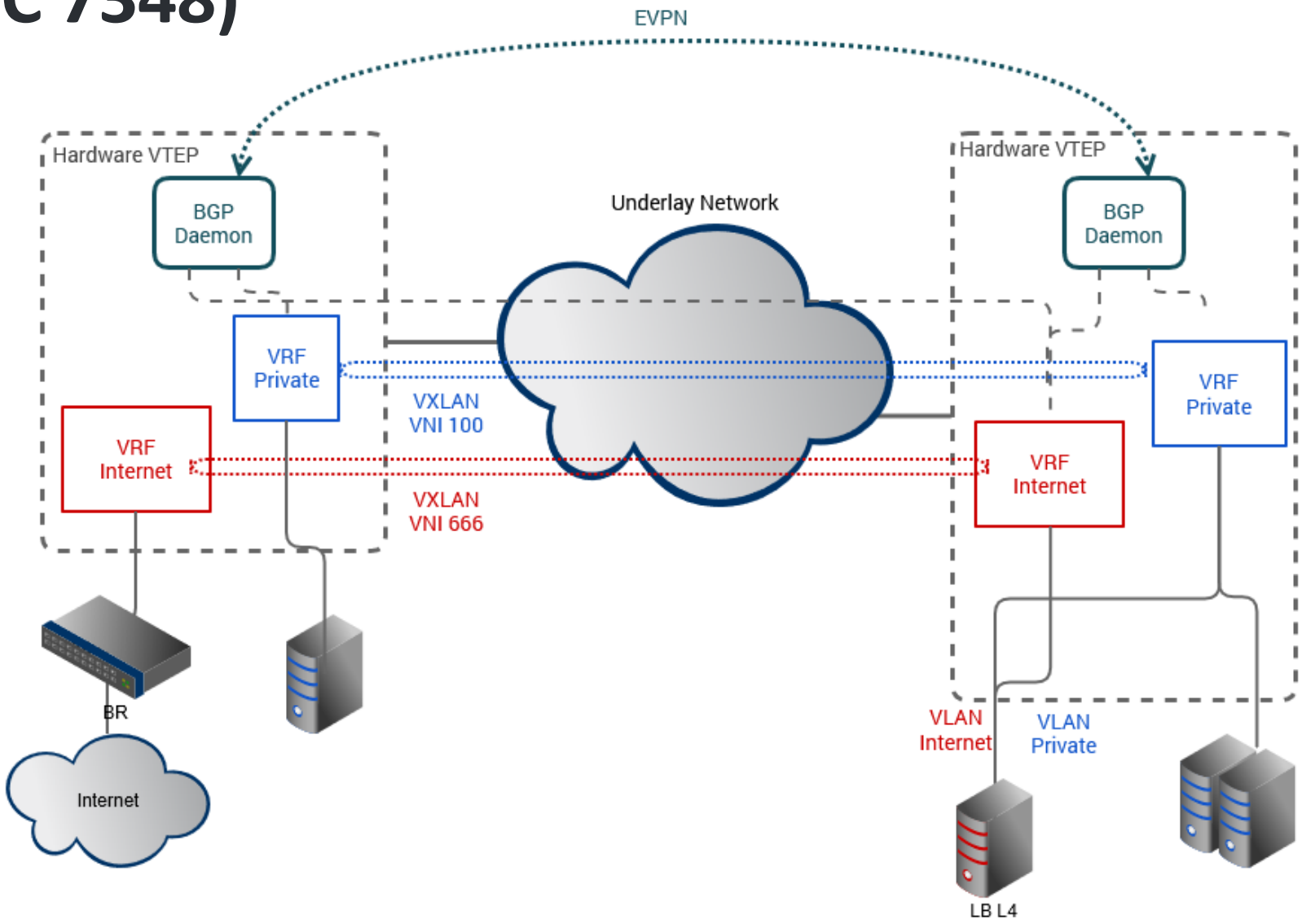
L4 LB

- **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**

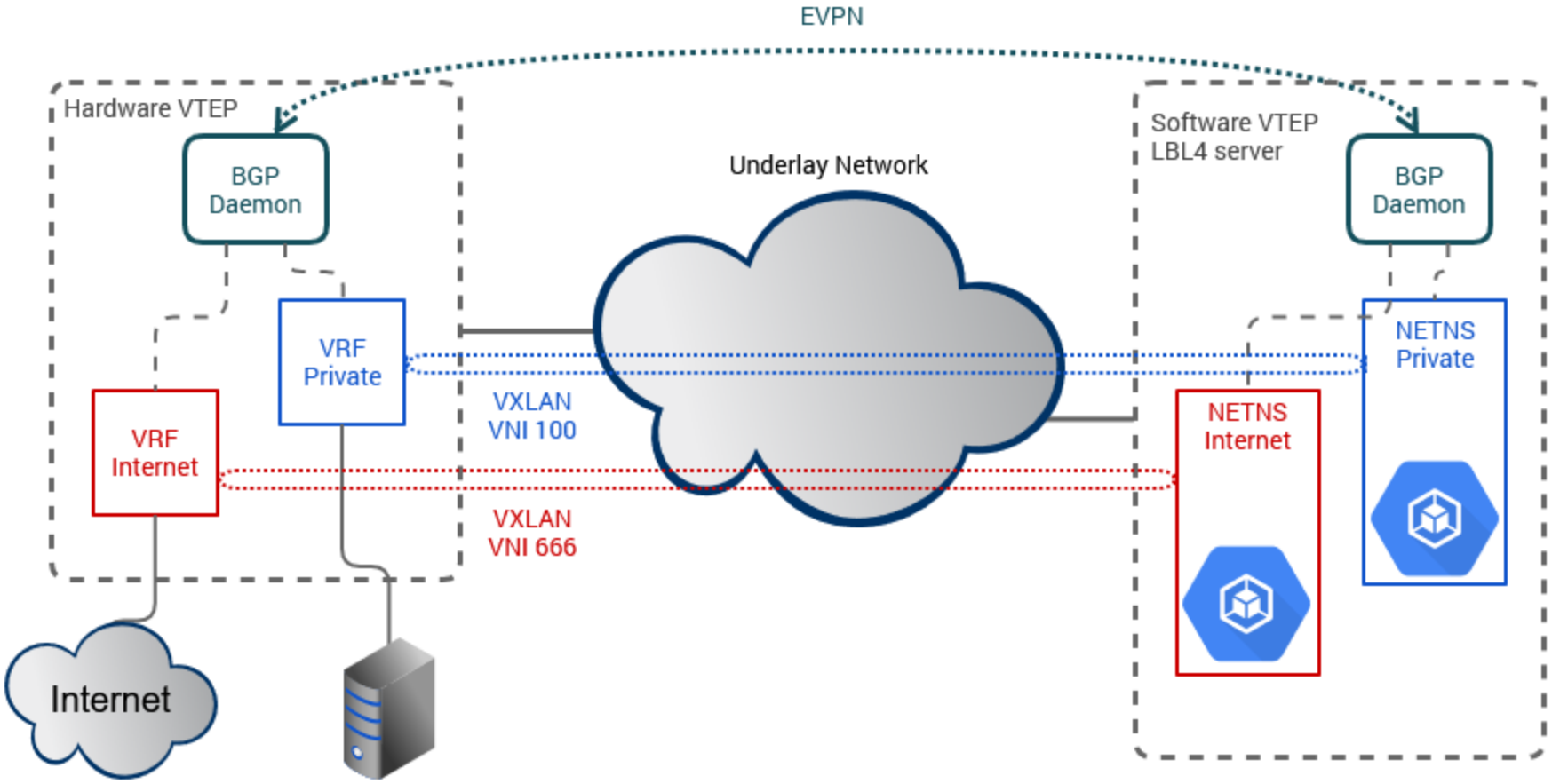
What is an L4 LB ?

CRITEO

VXLAN (RFC 7348)



Software VTEP



How it started

How it's going

CRITEO

Questions ?

Thank you!

Cédric Paillet

c.paillet@criteo.com

CRITEO