



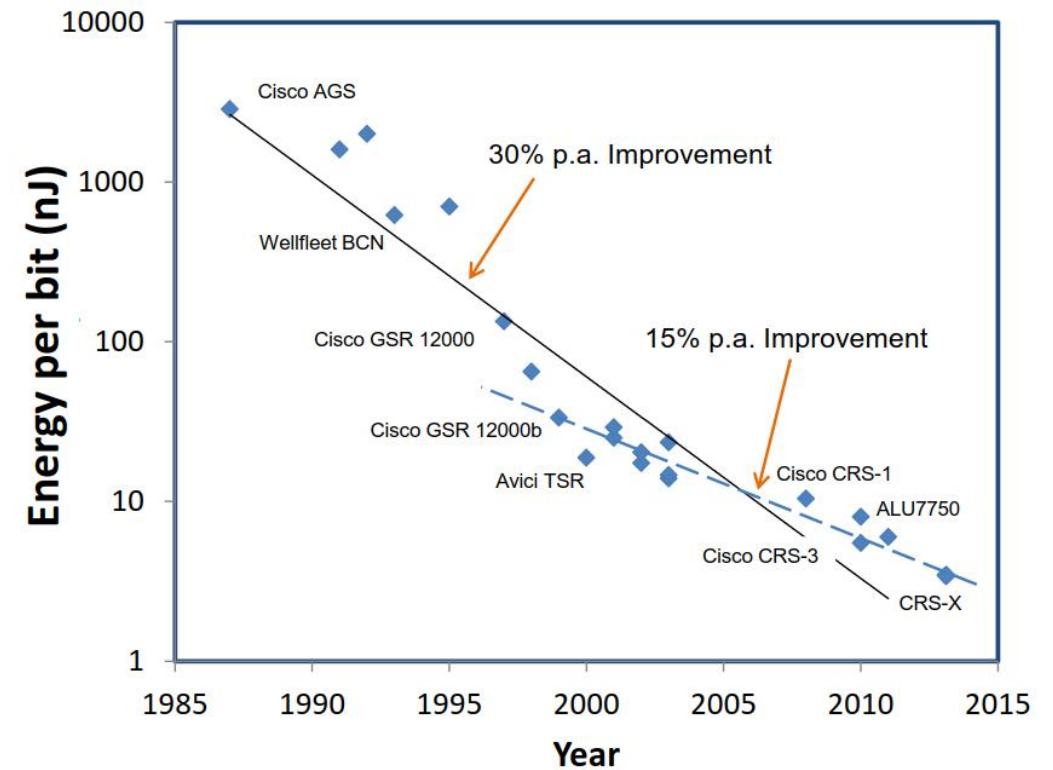
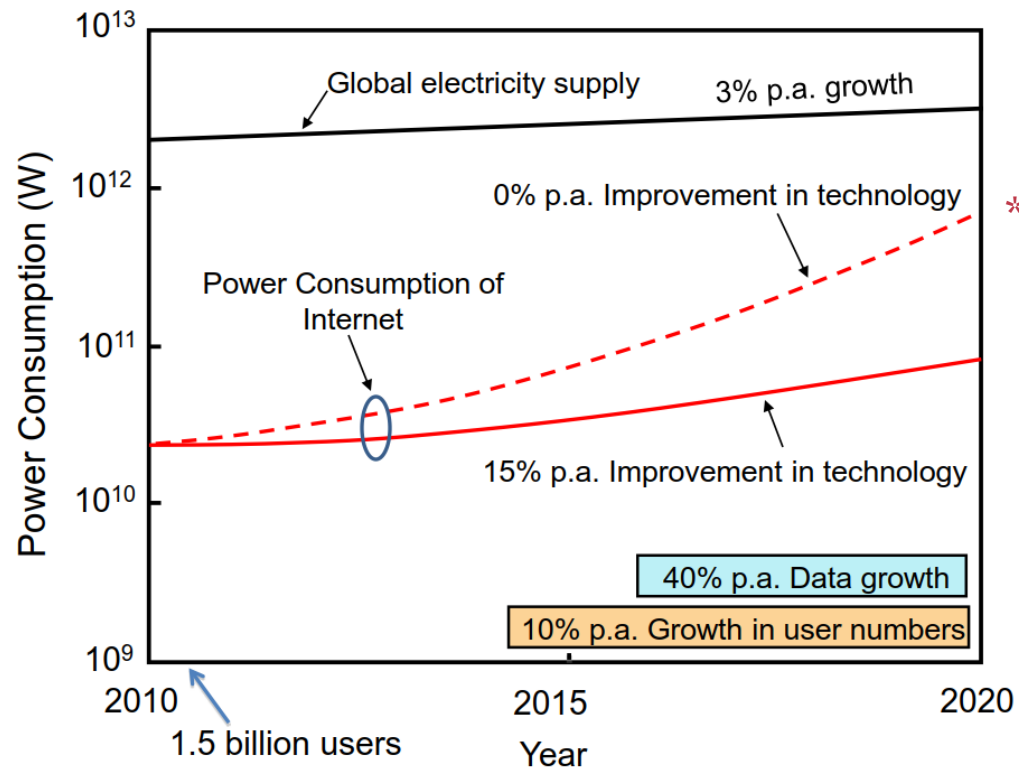
# SILVERENGINE

Consultancy, R&D Services and Products for  
Mission Critical Information and Communication Technology

## energy efficient data plane evolution



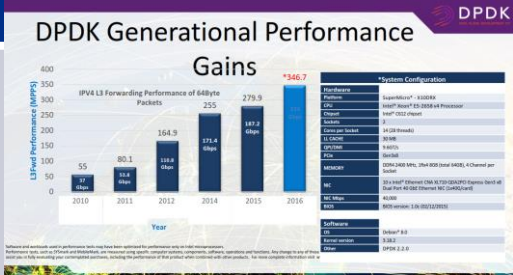

or why ETSI Network Function Virtualization is not sustainable

# current many-hop network architecture not sustainable applications must move closer to the users and the network layer



**\*) Note: after ~2022 CMOS scaling ends!**

a VSwitch is ~10 times less efficient than a TOR or NIC switch, and even less efficient than a latest edge router

TOR-Switch	NIC	VSwitch (DPDK)	Edge Router
			
Edge-Core Networks Wedge 100-32x	2x100G Mellanox NIC MCX516A-CCAT	2 x Intel E5-2658V4, 2x14 cores, 14nm, 347 Mpps, 64byte packets	Juniper MX 2010: 10 slot, 16x100GbE/slot
3,2 Tb/s, 436 W	200 Gb/s, ~ 17.6 W	233 Gb/s, ~300W 210W CPUs, 40W NICs, 50W other	16 Tb/s, ~15,5 kW, 25°C
~ 136 pico-Joule/bit	~ 89 pico-Joule/bit	~ 1286 pico-Joule/bit	~ 950 pico-Joule/bit
@ 100% load with 64B packets			

sources: respective product data sheets, NIC and VRouter power assumes efficiency of 0.91  
<https://dpdksummit.com/Archive/pdf/2017India/DPDK-India2017-Ramajain-ArchitectureRoadmap.pdf>

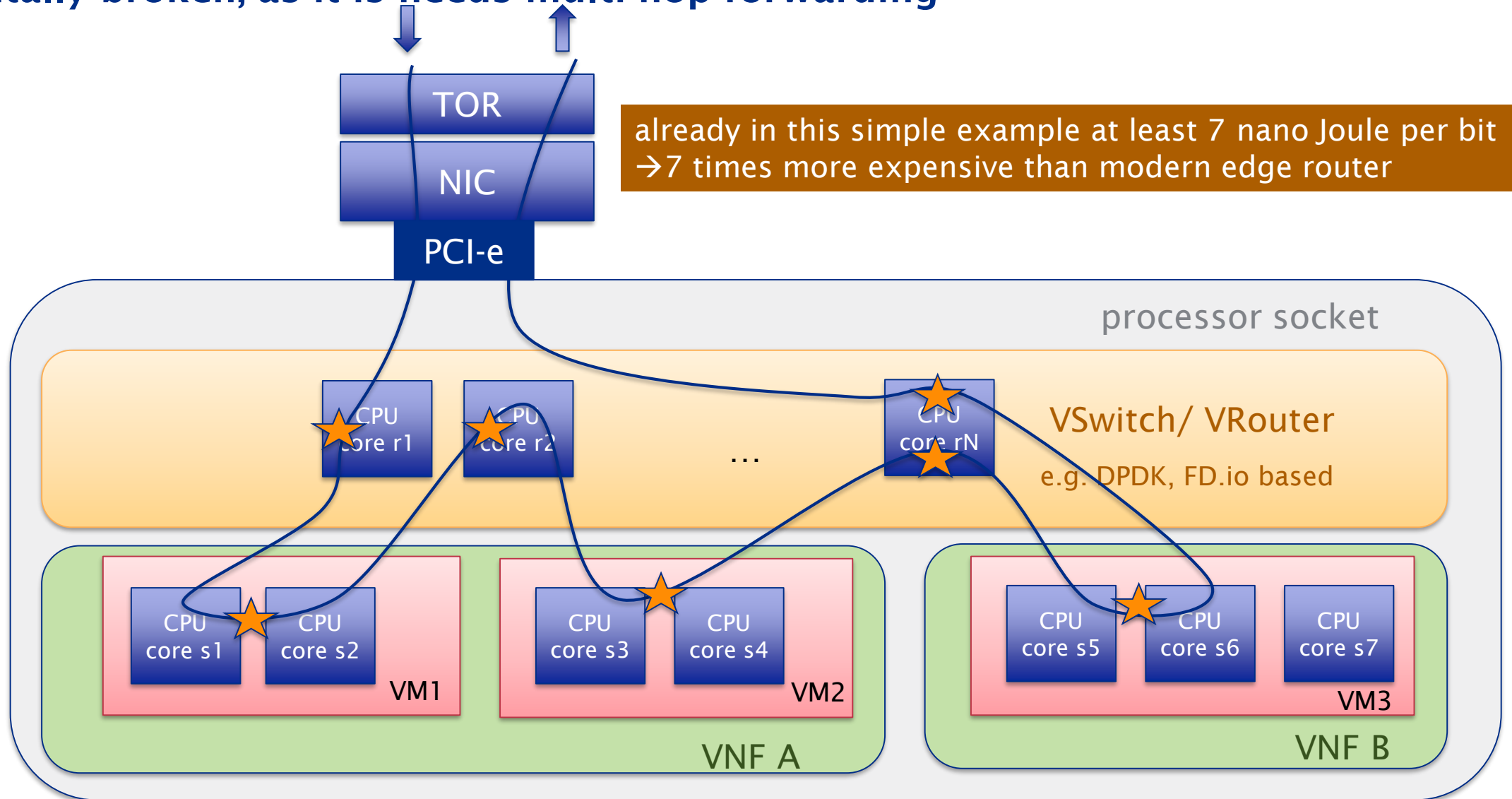
# 'classical' IaaS virtualization of a data plane with VRouter/VSwitch is fundamentally broken, as it needs multi hop forwarding

already in this simple example at least 7 nano Joule per bit  
 → 7 times more expensive than modern edge router

DC fabric underlay

network virtualization overlay

application layer, e.g. VNFs



## problem is recognized in industry and academic

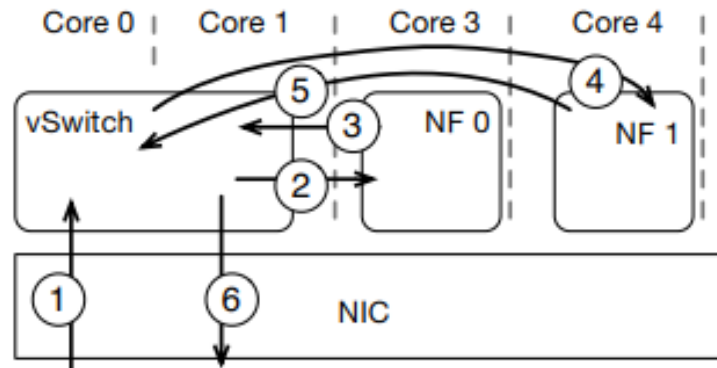
two major mitigation approaches are proposed:

- apply network functions without moving packets through a VRouter, e.g.
  - latest proposal of Scott Shenker: Zero Copy Software Isolation  
network functions exchange references to packets, not the packet data itself
  - requires move from IaaS (HW abstraction) to PaaS (SW-functional abstractions)
  - Xpress Data Path: packets stay in kernel/driver, network functions (eBPF) are loaded into the data path
  - Identifier Locator Addressing (ILA<sup>\*)</sup>) as NVO
- offload the General Purpose Processor from bulk data transfer tasks
  - well known from SAN and HPC technologies, e.g. remote DMA

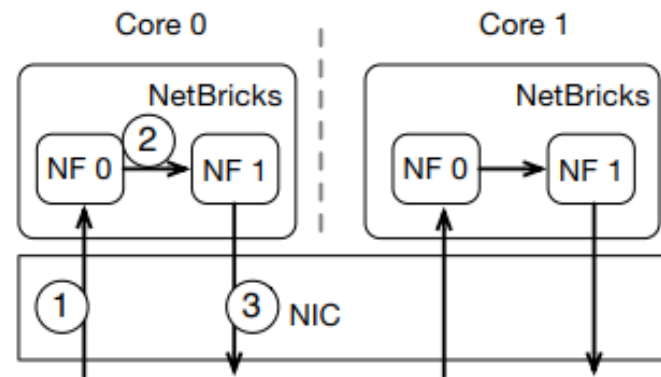
<sup>\*)</sup> IETF draft-herbert-nvo3-ila-04

## Zero Copy Software Isolation (ZCSI) / 'NetBricks' \*)

applying network functions without moving packets in-between



VRouter approach



ZCSI approach

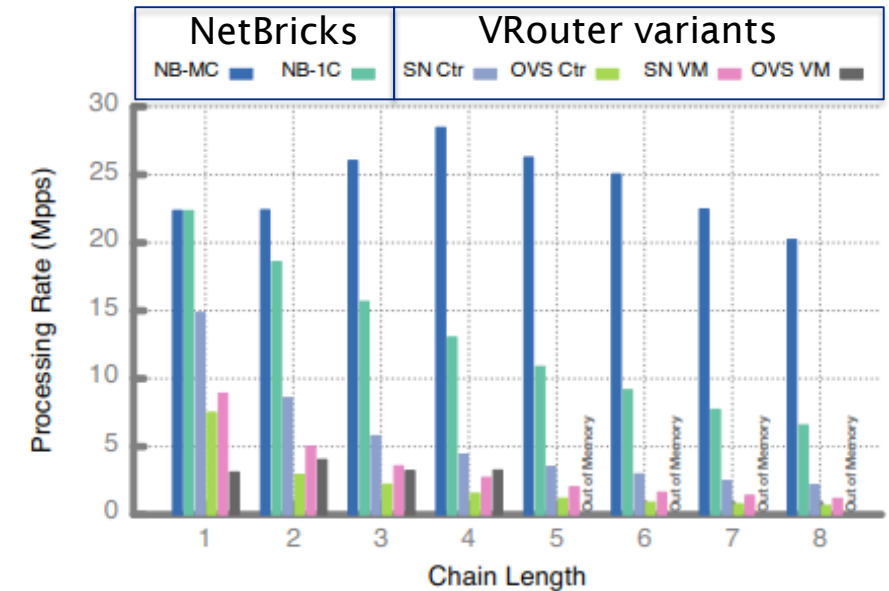


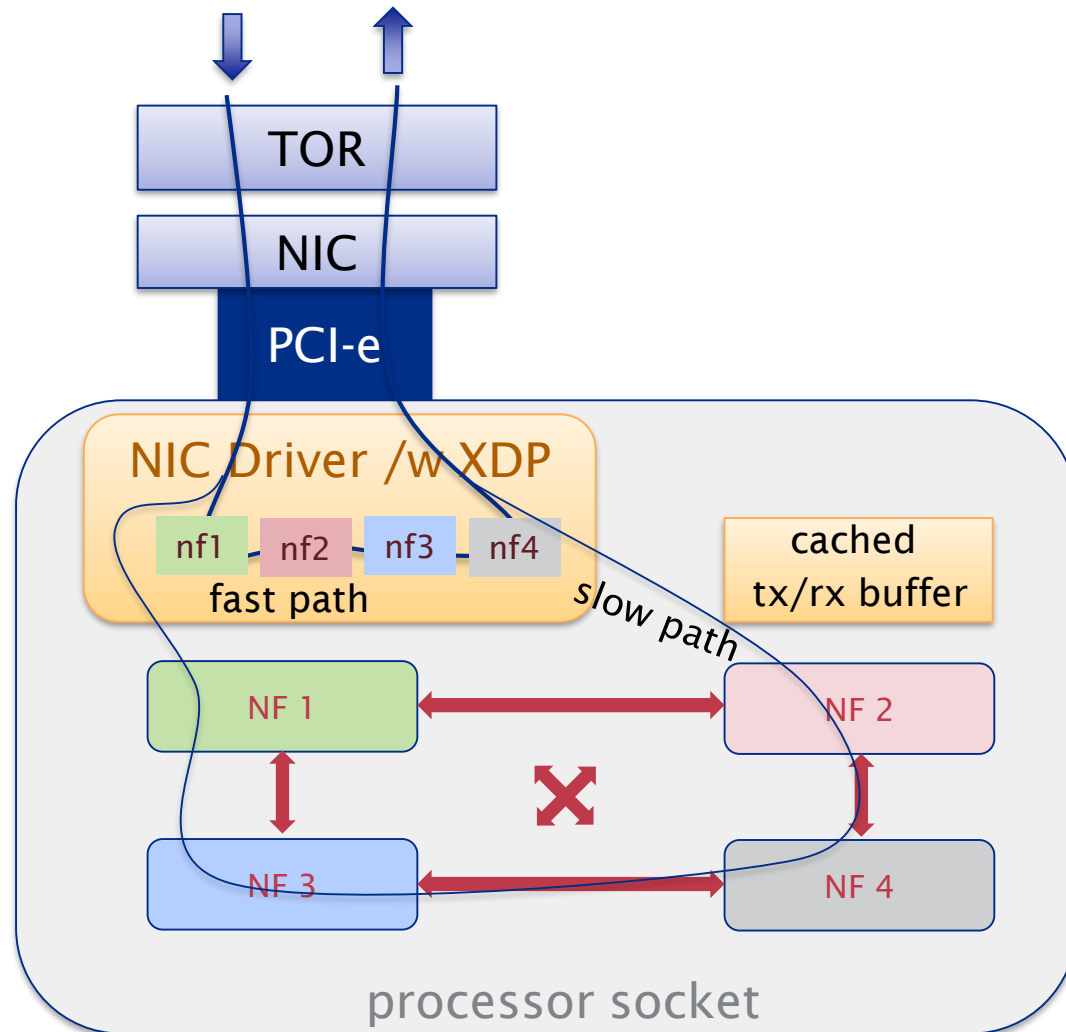
Figure 7: Throughput with increasing chain length when using 64B packets. In this figure NB-MC represents NetBricks with multiple cores, NB-1C represents NetBricks with 1 core.

- I/O handled by DPDK, network functions implemented in safe programming language Rust
- concurrent code checked by Rust compiler for synchronization issues
- requires Rust based execution platform for network functions, no virtualization(!)
- so far, no control plane in Netbricks (e.g. ARP, ICMP, ...)

\*) A. Panda, S. Han, K. Jang, M. Walls, S. Ratnasamy, S. Shenker, ODSI 11/2016, "NetBricks: Taking the V out of NFV", <https://people.eecs.berkeley.edu/~apanda/assets/papers/osdi16.pdf>, <http://netbricks.io>

## Xpress Data Path

a Linux based programmable data path with DPDK-like performance

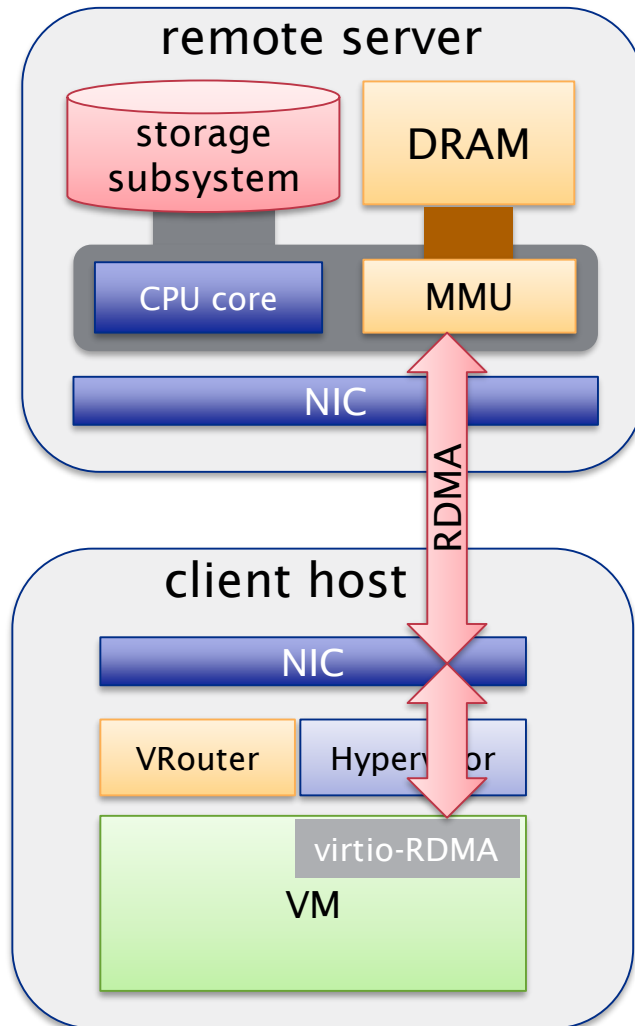


- XDP offload for forwarding traffic, whenever possible
- chaining of network functions in an XDP enabled NIC driver
- no need for proprietary DPDK and extra licenses
- XDP as a service, as part of an NFV PaaS



## an offload example from HPC industry: Remote Direct Memory Access

offloading the virtual network overlay from bulk data transfers



- remote CPU bypass with RDMA for inter server bulk traffic
- native network: Infiniband
- RDMA over converged IP/Ethernet fabric
  - iWARP (RFC 5040-5044)
  - RoCE
- 100G throughput per server possible today
- scalable Network File System over RDMA
  - parallel NFS over RDMA (RFC 5661-5664)
  - supported in upstream Linux kernel



## Conclusion

- we expect applications to drive a massive IT'zation of edge and core
  - only with a common harmonized network- and application-architecture energy efficiency targets are achievable
- for network-bound applications classical IaaS following ETSI NFV is fundamentally broken
- Zero Copy Soft Isolation looks like a very promising solution approach
- HPC network offload technologies mitigate the problem, but impact the application architecture
- all approaches must be based on services offered to applications and network functions by a Telco-PaaS
- need to replace ETSI NFV architecture with a NF over PaaS architecture