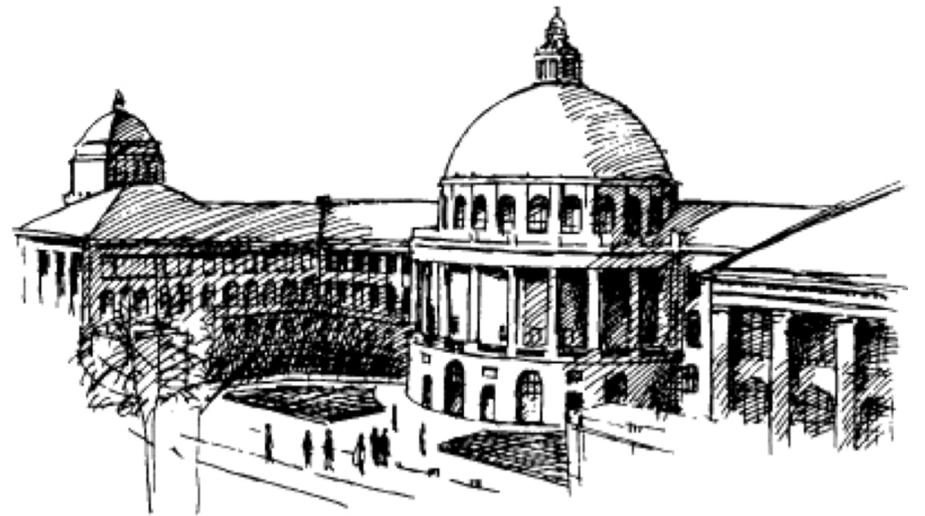
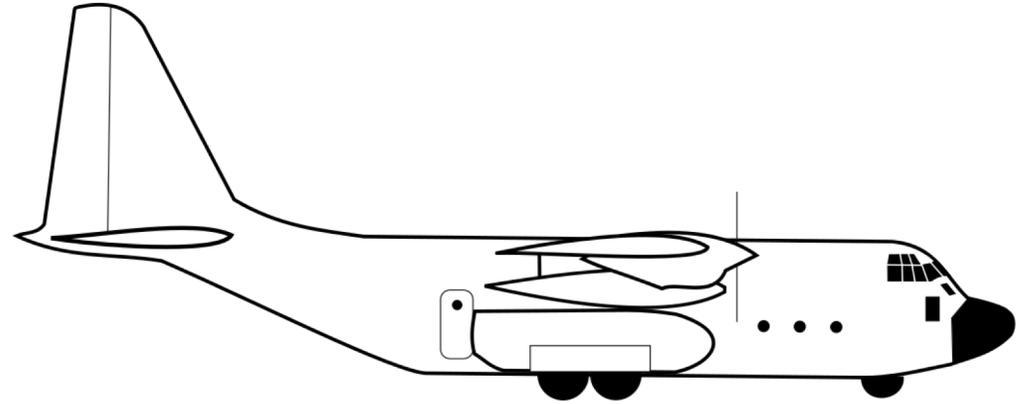


Hercules

Bulk Data Transfer over SCION

Presented by
Matthias Frei and François Wirz



Project Scope

High-speed large file transfer over Internet

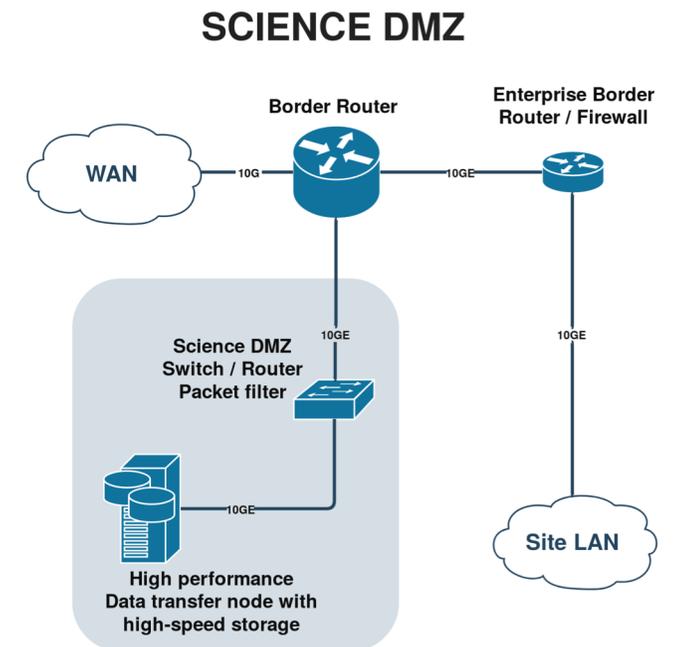
- Large = Terabyte-scale data transfers

Use Cases

- Data-intensive science: healthcare, physics, big data, etc.
- Remote processing, data needs to be transmitted beforehand
- Remote backup

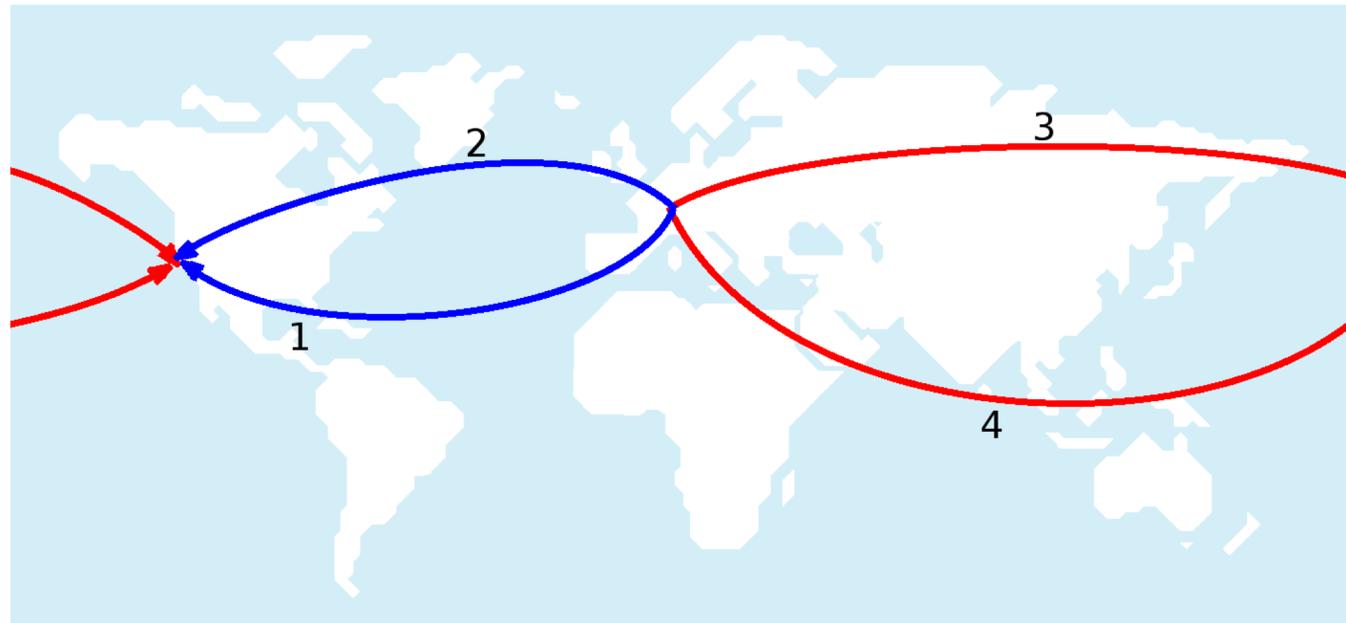
Traditional Approaches

- FTP over TCP/IP
 - TCP suffers from degraded performance with high latency and random losses
 - Poor multipath support
 - Open many TCP streams and hope and pray
 - Multipath TCP in the future
 - Poor utilisation of available capacity
- Science DMZ
 - Designated data transfer infrastructure, in front of enterprise firewall
 - Simple packet filter, whitelist source IPs



How can SCION Speed up File Transfer?

- Clean multipath communication
 - Multiple disjoint paths
 - Utilize local backup links

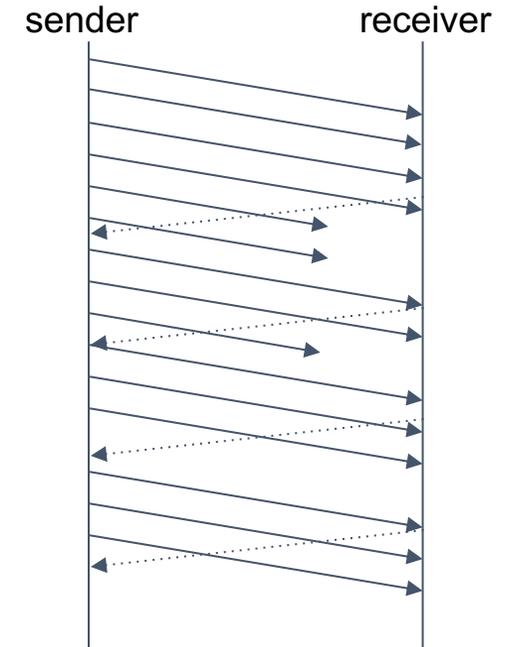


How can SCION Speed up File Transfer?

- Clean multipath communication
 - Multiple disjoint paths
 - Utilize local backup links
- Simplified congestion control & low loss thanks to COLIBRI quality-of-service system
- LightningFilter: packet filter for Science DMZ *with* strong cryptographic packet authentication

Hercules

- SCION/UDP packet blasting + retransmits
 - “Reliable Blast UDP”^[1]
- Range ACKs at fixed frequency
- Performance-oriented congestion control^[2]
 - Link empirical performance to actions taken



[1] "[Reliable Blast UDP : Predictable High Performance Bulk Data Transfer](#)", Eric He, Jason Leigh, Oliver Yu and Thomas A. DeFanti, Proceedings of IEEE Cluster Computing, Chicago, Illinois, September, 2002

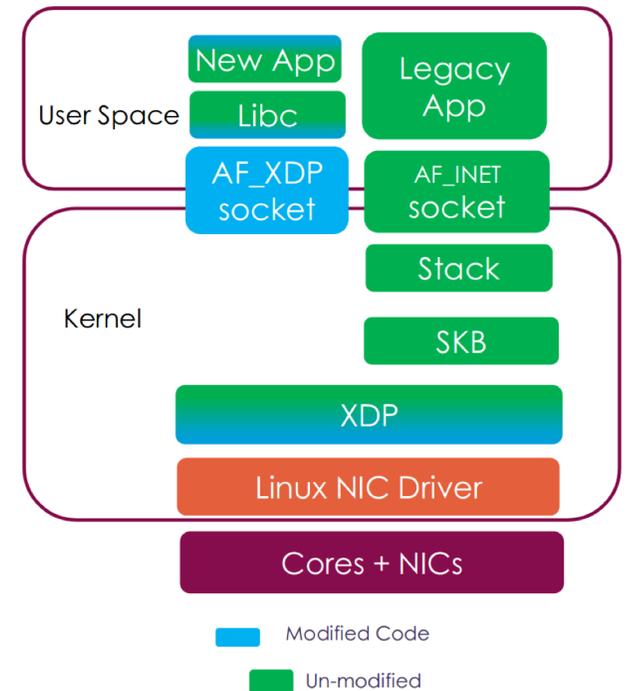
[2] "[PCC: Re-architecting Congestion Control for Consistent High Performance](#)", Mo Dong, Qingxi Li, Doron Zarchy, P. Brighten Godfrey, and Michael Schapira, 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15)

Hercules

AF_XDP^[3] for high performance SCION/UDP

- Published in December 2018
available in Linux ≥ 4.18
zero-copy mode in Linux ≥ 5.1
- Bypass Linux networking stack for send/receive
- Bypass SCION dispatcher

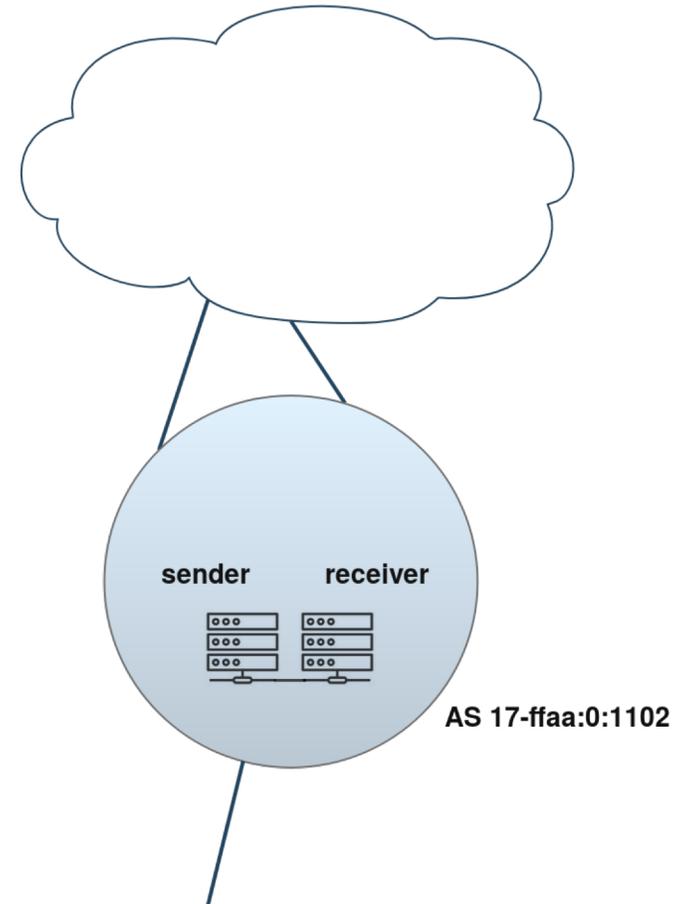
[3] "[Accelerating networking with AF_XDP](#)", Jonathan Corbet, LWN.net, 2018



PMD for AF_XDP: Zhang Qi, Li Xiaoyun

Demo

- Transfer file between two SCION hosts in *same AS*
- Directly connected, 40GbE
- *Not* the target use case, but high-performance SCION links are being established



Demo

```
Alacritty
[0] 0:demo* 1:src- "tag" 15:35 04-Nov-19
matfrei@sender$ sudo numactl -l --cpunodebind=netdev:ens787f1 -- ./hercules -i ens787f1 -q 0 -l 17-ffaa:0:1102,[172.16.0.1]:10000 -d 17-ffaa:0:1102,[172.16.0.2]:10000 -t data-20G.bin -pcc=false
Waiting for receiver to get ready... OK
```

Time	Completion	Goodput	Throughput now	Throughput target	Throughput avg	Pkts sent	Pkts rcvd			
1.1s	0.00%	25.4Gbps	26.9Gbps	2.2Mpps	40.0Gbps	3.3Mpps	25.5Gbps	2.1Mpps	2244922	2
2.1s	34.92%	30.6Gbps	32.5Gbps	2.7Mpps	40.0Gbps	3.3Mpps	31.6Gbps	2.6Mpps	5411840	11
3.1s	53.98%	31.9Gbps	33.8Gbps	2.8Mpps	40.0Gbps	3.3Mpps	33.2Gbps	2.8Mpps	8456448	19
4.1s	73.32%	32.7Gbps	34.7Gbps	2.9Mpps	40.0Gbps	3.3Mpps	34.2Gbps	2.8Mpps	11562961	28
5.1s	95.80%	33.4Gbps	35.4Gbps	2.9Mpps	40.0Gbps	3.3Mpps	35.0Gbps	2.9Mpps	14742250	36
5.3s	100.00%	33.0Gbps	35.0Gbps	2.9Mpps	40.0Gbps	3.3Mpps	34.6Gbps	2.9Mpps	15192433	38

```
Transfer completed:
Duration:      5.272s
Filesize:     20.0GiB
Rate:         32.6Gb/s (3.8GiB/s)
Sent/Chunk:   1.000
Rcvd/Chunk:   0.000
matfrei@sender$ shasum data-20G.bin
722299740a266f799a2460b40cf4e00b0d772b46  data-20G.bin
matfrei@sender$
```

```
matfrei@receiver$ sudo numactl -l --cpunodebind=netdev:ens786f1 -- ./hercules -i ens786f1 -q 0 -l 17-ffaa:0:1102,[172.16.0.2]:10000 -o data/rcv.bin
Preparing file for receive... OK
```

Time	Completion	Goodput	Throughput now	Throughput avg	Pkts rcvd	Pkts sent		
1.0s	13.68%	23.5Gbps	24.9Gbps	2.1Mpps	24.8Gbps	2.1Mpps	2077949	8
2.0s	34.53%	29.7Gbps	31.5Gbps	2.6Mpps	31.4Gbps	2.6Mpps	5243493	16
3.0s	54.60%	31.3Gbps	33.2Gbps	2.8Mpps	33.1Gbps	2.8Mpps	8291945	25
4.0s	75.01%	32.2Gbps	34.2Gbps	2.8Mpps	34.1Gbps	2.8Mpps	11392481	33
5.0s	95.94%	33.0Gbps	35.0Gbps	2.9Mpps	34.9Gbps	2.9Mpps	14570073	42
5.2s	100.00%	32.8Gbps	34.8Gbps	2.9Mpps	34.8Gbps	2.9Mpps	15187297	44

```
Transfer completed:
Duration:      5.237s
Filesize:     20.0GiB
Rate:         32.8Gb/s (3.8GiB/s)
Sent/Chunk:   0.000
Rcvd/Chunk:   1.000
matfrei@receiver$ shasum data/rcv.bin
722299740a266f799a2460b40cf4e00b0d772b46  data/rcv.bin
matfrei@receiver$
```

Demo Summary

- Hercules achieves ~30Gbps transfer rate
 - Disk I/O not included, much slower on demo host
- Comparison
 - `iperf3` with TCP achieves ~20Gbps (one thread)
 - `iperf3` with UDP, ~4Gbps
 - FTP achieves ~8Gbps

