

Linux Conference Australia 2022 – Virtual

NorNet – A Linux- and Open-Source-Software-based International Platform for Networking Research

Thomas Dreibholz (托马斯博士)

Simula Metropolitan Centre for Digital Engineering
dreibh@simula.no

January 15, 2022



Contents

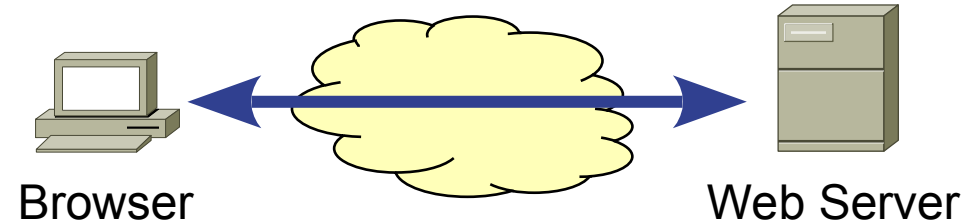
- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed
- The Software: VMs, Containers and Multi-Homed Networking
- Conclusion

Overview: Motivation

- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed
- The Software: VMs, Containers and Multi-Homed Networking
- Conclusion

„Classic“ Internet Communication

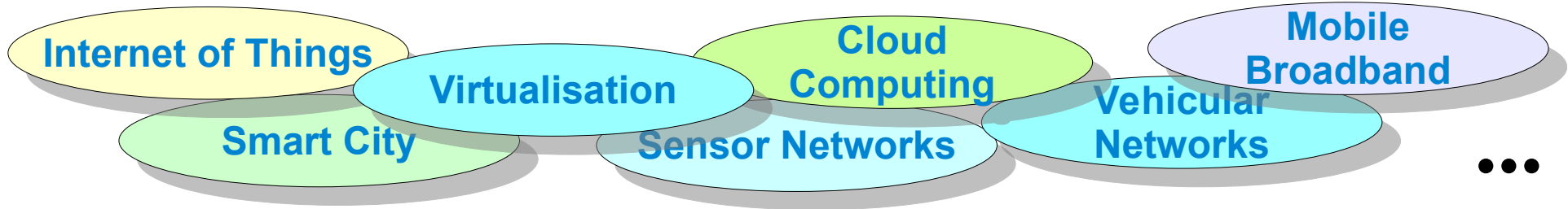
- Example: World-Wide Web



- Client ↔ Server Communication
 - 1 network interface per device → 1 **IPv4** address
 - Communication with Transmission Control Protocol (**TCP**)

The Current and Future Internet

The Big Picture



- IPv6
 - Devices are frequently IPv4/IPv6 dual stack
 - Usually multiple addresses per interface
- Mobility → address change
- Devices with multiple interfaces
 - Router
 - **Smartphone** (LTE/UMTS, WLAN, Bluetooth?)
 - **Laptop** (Ethernet, WLAN, LTE/UMTS?)



Multi-Homing and Multi-Path Transport

- Multi-Homing
 - Multiple interfaces (addresses)
 - **Redundancy** → Communication even when some paths fail
- Multi-Path Transport
 - Also utilise paths simultaneously → better throughput
 - **MPTCP**: Multi-Path TCP
 - **CMT-SCTP**: Concurrent Multi-Path Transfer for SCTP



SCTP: Stream Control Transmission Protocol
TCP: Transmission Control Protocol

Hot topic in research and standardisation!

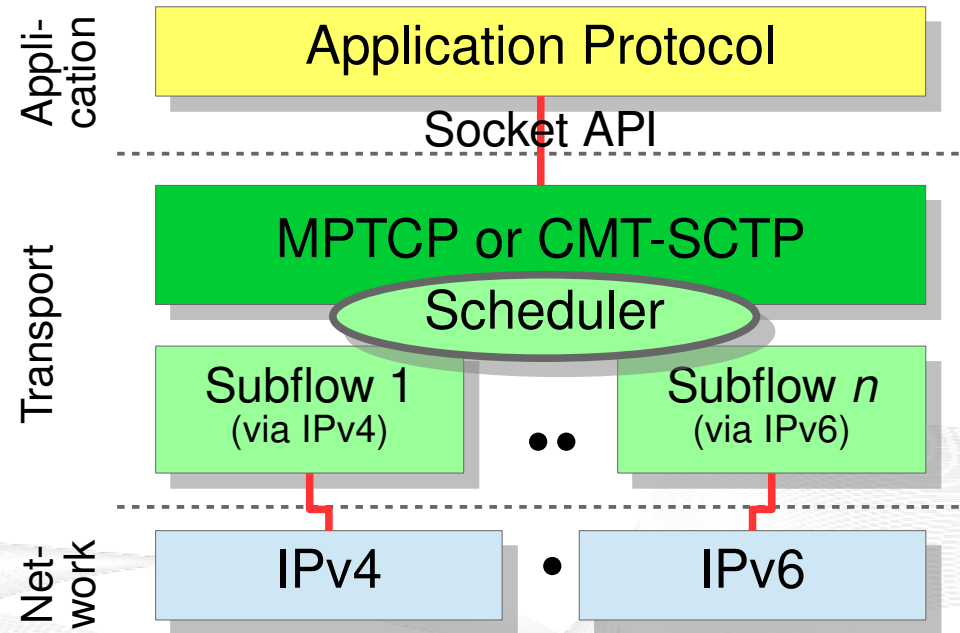
Overview:

Multi-Homing and Multi-Path Transport

- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed
- The Software: VMs, Containers and Multi-Homed Networking
- Conclusion

Multi-Path Transport with MPTCP and CMT-SCTP

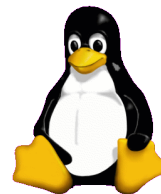
- Subflow \leftrightarrow path
- Fairness
 - Paths may overlap (fully oder partially)
- Scheduling
 - Different path characteristics
 - Bandwidth
 - Latency and jitter
 - Packet loss



How can I use multi-path transport?

Multi-Path TCP (MPTCP)

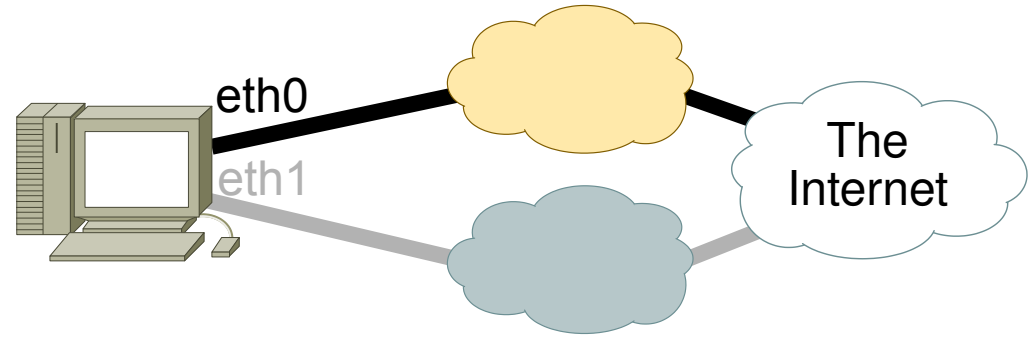
- RFC 6824
- Features: multi-homing + multi-path transport
- Backwards-compatible to TCP
 - Communicate with old TCP implementations
 - Works (mostly) even over non-MPTCP middleboxes (e.g. NAT/PAT)
- Linux:
 - UC Louvain → <https://www.multipath-tcp.org>
- FreeBSD:
 - Swinburne → <http://caia.swin.edu.au/newtcp/mptcp/>



How can I use MPTCP under Linux?

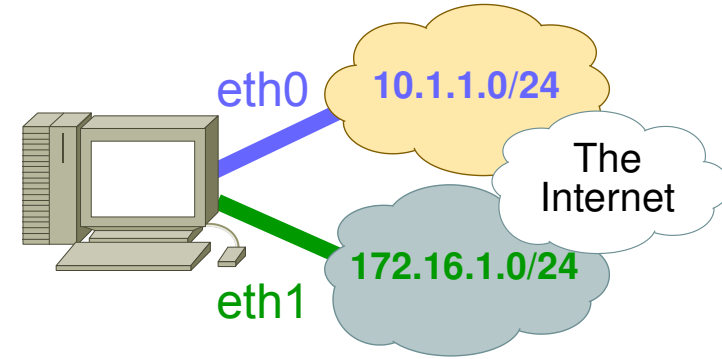
Routing Tables and Routing Rules for Multi-Path Transport

- Example:
 - Device eth0 → ISP 1
 - Device eth1 → ISP 2
- Problem:
 - First default route (with lowest metric) via ISP 1
 - All traffic uses ISP 1 ☹️
- Solution:
 - **Routing rules**
 - Separate routing tables for each ISP
 - “Selector” for actually used table, per source address



A Linux Routing Rule Example

- Configure eth0:
 - `ip addr add 10.1.1.42/24 dev eth0`
 - `ip route add default via 10.1.1.1 dev eth0`
 - `ip route add 10.1.1.0/24 scope link dev eth0 table 1`
 - `ip route add default via 10.1.1.1 dev eth0 table 1`
- Configure eth1:
 - `ip addr add 172.16.1.42/24 dev eth1`
 - `ip route add 172.16.1.0/24 scope link dev eth1 table 2`
 - `ip route add default via 172.16.1.1 dev eth1 table 2`
- We have 2 new routing tables now! Set up routing rules based on source address:
 - `ip rule add from 10.1.1.42 table 1`
 - `ip rule add from 172.16.1.42 table 2`
- Table numbers difficult to remember? Set name mapping in `/etc/iproute2/rtable`!



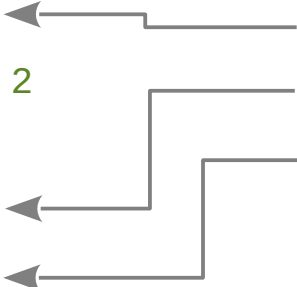
The Resulting Configuration

- The routing rules: `ip rule show`

```
0:      from all lookup local
32764:  from 172.16.1.42 lookup 2
32765:  from 10.1.1.42 lookup 1
32766:  from all lookup main
32767:  from all lookup default
```

- The name mappings: `cat /etc/iproute2/rt_tables`

```
255    local
254    main
253    default
0      unspec
```



- Table #1: `ip route show table 1`

```
default via 10.1.1.1 dev eth0
10.1.1.0/24 dev eth0 scope link
```

- Table #2: `ip route show table 2`

```
default via 172.16.1.1 dev eth1
172.16.1.0/24 dev eth1 scope link
```

- Table “main” (254): `ip route show table main`

```
default via 10.1.1.1 dev eth0
172.16.1.0/24 dev eth1 proto kernel \
scope link src 172.16.1.42
10.1.1.0/24 dev eth0 proto kernel \
scope link src 10.1.1.42
```


What about Routing Rules with IPv6?

- It works with IPv6 as well, of course!
 - `ip -6 addr add 3ffe:cafe:affe:1234::2a/64 dev eth0`
 - `ip -6 route add default via 3ffe:cafe:affe:1234::1 dev eth0`
 - `ip -6 route add 3ffe:cafe:affe:1234::/64 scope link dev eth0 table 1`
 - `ip -6 route add default via 3ffe:cafe:affe:1234::1 dev eth0 table 1`
 - `ip -6 addr add 3ffe:dead:beef:ffff::2a/64 dev eth1`
 - `ip -6 route add 3ffe:dead:beef:ffff::/64 scope link dev eth1 table 2`
 - `ip -6 route add default via 3ffe:dead:beef:ffff::1 dev eth1 table 2`
 - `ip -6 rule add from 3ffe:cafe:affe:1234::2a table 1`
 - `ip -6 rule add from 3ffe:dead:beef:ffff::2a table 2`
- Depending on source address, a packet leaves via **network 1** or **network 2**
 - MPTCP (and CMT-SCTP) will make this choice, depending on subflow
- Connections can even have IPv4- and IPv6 subflows simultaneously!

Finally: Testing MPTCP

- First, boot a MPTCP-enabled kernel
 - See <https://multipath-tcp.org> for sources
 - MPTCP enabled by default → all TCP connections are MPTCP-capable!
- Configure (and check) the routing
 - Connect to two (or more) ISPs, if possible
 - IPv4 + IPv6 may also give you partially independent paths
 - Just 1 IP address? → multiple paths to a multi-homed remote side!
- Test:
 - NetPerfMeter → <https://www.uni-due.de/~be0001/netperfmeter/>
 - Wireshark/T-Shark → <https://www.wireshark.org>
 - ...

Overview: The NorNet Testbed

- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed
- The Software: VMs, Containers and Multi-Homed Networking
- Conclusion

Testing Multi-Path Transport (1)

The First Step – A Lab Setup

- Surprisingly big effort:
 - Strange effects of cheap network components:
„It's only cheap on the paper!“
 - Debugging of SCTP in FreeBSD
- But valuable:
 - **Prior simulations were useful!**
 - Bugfixes for the FreeBSD community
 - Open Source software „**NetPerfMeter**“
 - **Learning effects** and **new ideas!**



Internet protocols → testbed in the Internet!

Testing Multi-Path Transport (2)

Real Internet: 3 Cities and 2 Continents



- 3 connected lab setups
 - Establishment of an international cooperation
 - **Essen, Burgsteinfurt** (FH Münster), **Haikou 海口** (Hainan University)
- Very interesting scenario:
 - CMT-SCTP and MPTCP evaluation
 - Very different path characteristics
 - Ideas for further experiments

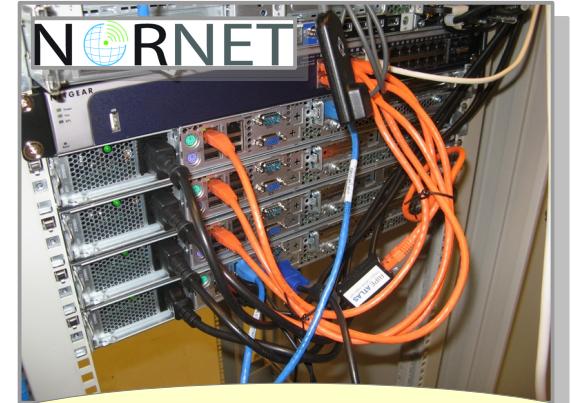
Many new ideas!

Now really big: NorNet testbed!

Testing Multi-Path Transport (3)

The NorNet Testbed

- NorNet Core
 - Cable, up to 4 providers, IPv4+IPv6 (fibre, „consumer-grade” DSL, etc.)
 - Hosts for virtual machines
 - 21 locations (11 in Norway, 10 abroad)
- [**simula** . research laboratory]
- NorNet Edge
 - Embedded systems (3 generations) running customised Debian Linux
 - Up to 4× 2G/3G/4G (+ 1× CDMA, 1× Ethernet)
 - Hundreds of locations (in Norway)



<https://www.nntb.no>



NorNet Core Site Deployment Status (October 2016)

No.	Site	ISP 1	ISP 2	ISP 3	ISP 4
1	Simula Research Laboratory	Uninett	Kvantel	Telenor	PowerTech
2	Universitetet i Oslo	Uninett	Broadnet	PowerTech	
3	Høgskolen i Gjøvik	Uninett	PowerTech		
4	Universitetet i Tromsø	Uninett	Telenor	PowerTech	
5	Universitetet i Stavanger	Uninett	Altibox	PowerTech	
6	Universitetet i Bergen	Uninett	BKK		
7	Universitetet i Agder	Uninett	PowerTech		
8	Universitetet på Svalbard	Uninett	Telenor		
9	Universitetet i Trondheim	Uninett	PowerTech		
10	Høgskolen i Narvik	Uninett	Broadnet	PowerTech	
11	Høgskolen i Oslo og Akershus	Uninett	–		
12	Karlstads Universitet	SUNET			
13	Universität Kaiserslautern	DFN			
14	Universität Duisburg-Essen	DFN	Versatel		
15	Hainan University 海南大学	CERNET	China Unicom		
16	The University of Kansas	KanREN			
17	Korea University 고려대학교	KREONET			
18	National ICT Australia (NICTA)	AARNet			
19	HAW Hamburg	DFN			
20	Technische Universität Darmstadt	DFN			
21	Haikou Cg. of Econ. 海口经济学院	China Telecom	CERNET		

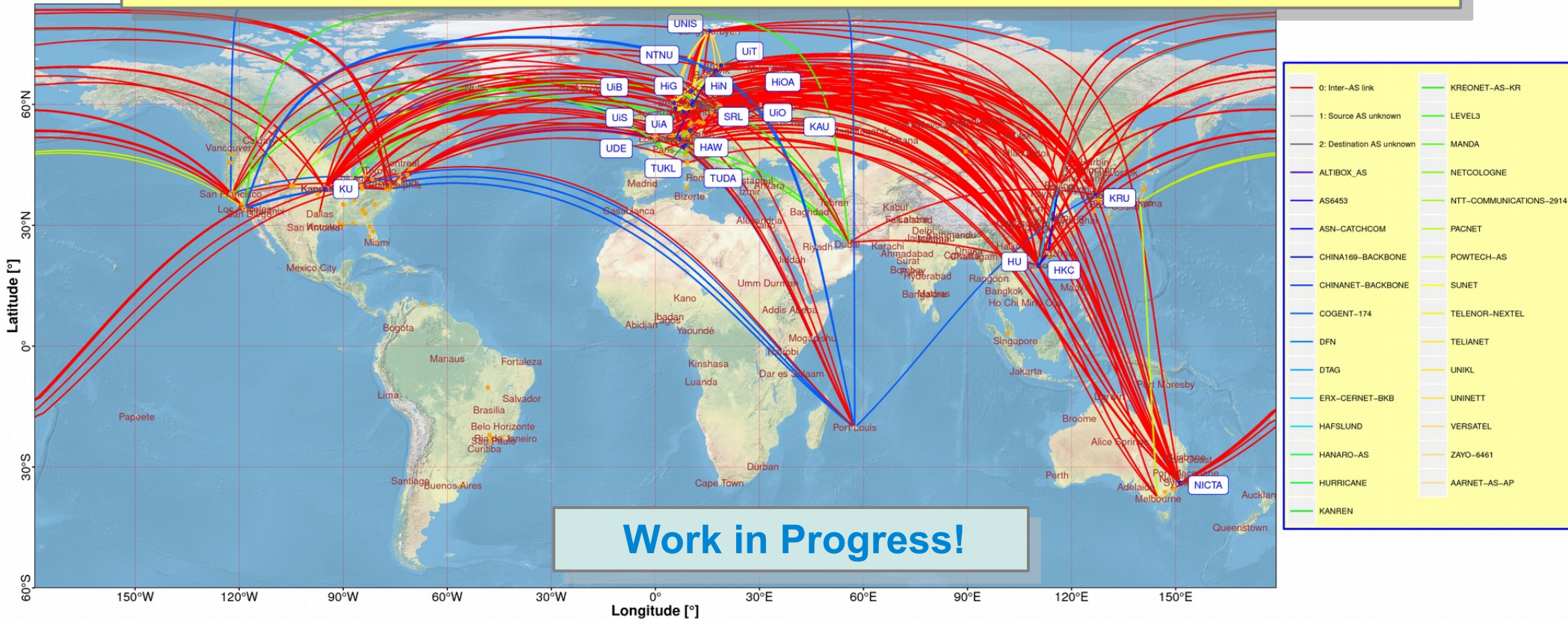
A global Internet
testbed infrastructure!

- IPv4 and IPv6
- IPv4 only (ISP without IPv6 support)
- IPv4 only (site's network without IPv6 support)
- ISP negotiation in progress

<https://www.nntb.no/pub/nornet-configuration/NorNetCore-Sites.html>

Routing Visualisation

HiPerConTracer Observed Routes and Autonomous Systems, from July 1 to July 31, 2016





Our servers may be really remote!



The “road” to Longyearbyen på Svalbard, 78.2°N

Overview:

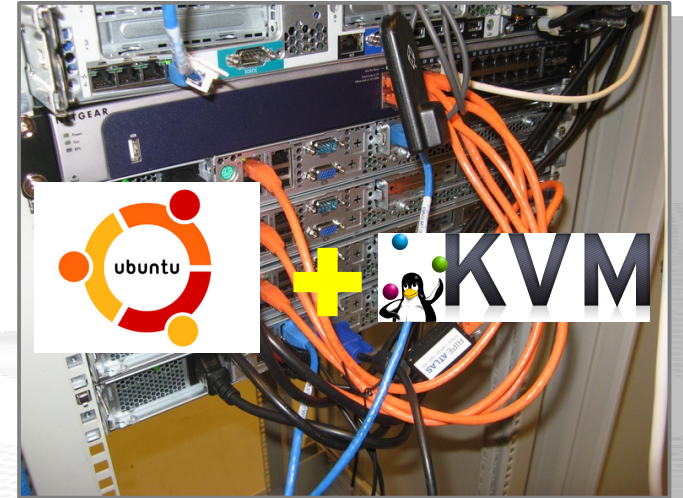
The Software: VMs, Containers and Multi-Homed Networking

- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed
- The Software: VMs, Containers and Multi-Homed Networking
- Conclusion

Virtualisation

“Anything that can go wrong, will go wrong.”
[Murphy's law]

- Experimentation software is experimental
- How to avoid software issues making a remote machine unusable?
- Idea: virtualisation
 - Lightweight, stable software setup:
Ubuntu Server 18.04 LTS → 20.04/22.04 LTS
 - **KVM (Kernel-based Virtual Machine)**
 - Other software runs in VMs:
 - Tunnelbox (router) VM
 - Research Node VMs
 - In case of problem: manual/automatic restart or reinstall of VM



Physical Machine Setup

- Ubuntu Server LTS – due to long-term support
- Customisation:
 - File system: ReiserFS 3
 - Ext4 resilience is really awful – manual fsck at remote machines ☹️
 - BTRFS has nice features, but awful performance for hosting VMs
 - => **ReiserFS!**
 - very reliable (it never killed the data of a machine)
 - good performance, also for hosting VMs
 - Newer installations: **BTRFS** (after removal of ReiserFS from installer)
 - Virtualisation: now KVM
 - Originally: VirtualBox (custom package with Open Source VNC, instead of Oracle's closed source blob)

Tunnelbox – The Router (1)

- Tunnelbox
 - Router at each site
 - Handles all network communication, over multiple ISPs
 - 1 public IPv4 address (+ 1 public IPv6 address) per site and ISP
 - Tunnels among the sites (GRE-over-IPv4, IPv6-over-IPv6)
 - Own systematic addressing scheme
 - IPv4: 10.<Provider>.<Site>.<Node>/24
 - IPv6: 2001:700:4100:<Provider><Site>::<Node>/64
 - Direct communication between sites over tunnels
 - Communication between sites and Internet over Simula's site
 - Security + avoiding legal issues (DE: "Mitstörerhaftung" ...)

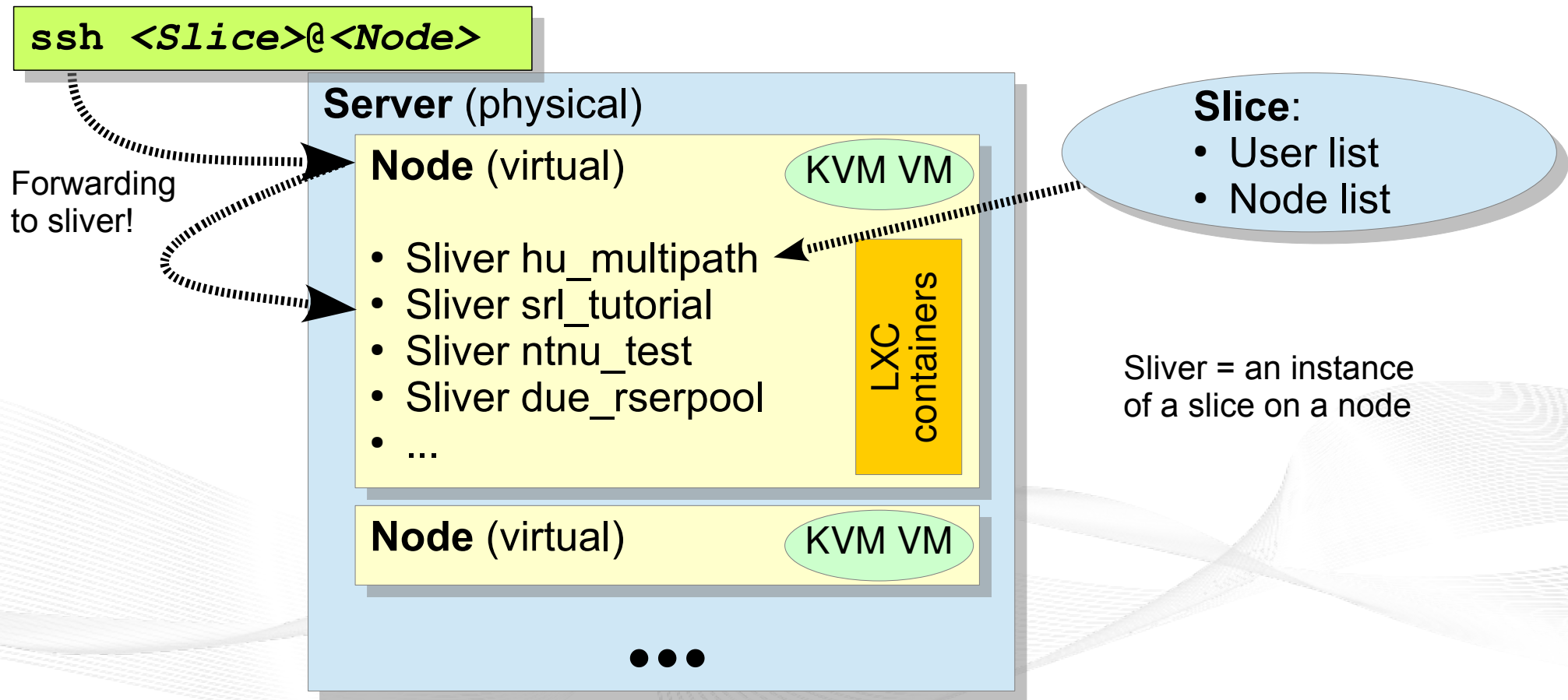
Tunnelbox – The Router (2)

- IP addresses are difficult to remember
 - DNS setup (bind9) with private TLD “.nornet”
 - Convenience 1: easy naming scheme:
 - fjellrev.telenor.unis.nornet: node “fjellrev” with ISP “Telenor” at “UNIS”
 - borbeck.ude.nornet: node “borbeck” with primary ISP at “UDE”
 - østhorn.kvantel.simula.nornet → xn--sthorn-9xa.kvantel.simula.nornet
 - 三亜 .cnunicom.hu.nornet → xn--ehqrn.cnunicom.hu.nornet
 - Convenience 2: SSHFP (SSH key fingerprints) and LOC (geolocation) RRs
- Squid HTTP proxy
 - Caching HTTP accesses (mainly: package updates)
 - If necessary: forward every request to Simula (to avoid legal issues)

Research Node

- “Usual” research node:
 - A VM (usually KVM), managed by PlanetLab Central (PLC)-based software
 - 2.5 GiB RAM, 1-2 cores
 - Fedora Core 25, Linux kernel v4.4 with MPTCP v0.94 + API patch
 - User gets a “sliver” of the research node → LXC container
 - “Own” Fedora installation, with development tools, T-Shark, ...
 - “Own” IP addresses (IPv4 + IPv6, for each ISP of the site)
 - Root permission (only within the LXC container, with limitations)
 - But: A slice is basically a BTRFS clone of a template
 - File duplicates only necessary upon changes
 - Very lightweight setup per user
- Custom VMs as research nodes are also possible (currently requires manual setup)

The Different Entities: Server, Node, Slice and Sliver



Sliver = an instance of a slice on a node

A Look into a Sliver

```
dreibh : oesthorn — Konsole <2>  
File Edit View Bokmerker Settings Help  
dreibh@nordmarka:~$ ssh -i ~/.ssh/test2 srl_test@boao.hu.nor-net  
Last login: Wed Sep 7 20:21:46 2016 from 10.1.1.51  
  
System information as of Wed Sep 7 20:22:42 UTC 2016  
  
=====
```

```
# #  
## # ##### ## # #### #  
# # # # # # # # #  
# # # # # # # # #  
# # # # ##### # # #  
# # # # # # # # #  
# # ##### # # # #  
=====
```

```
Slice:          srl_test (nor-net-f23-x86_64)  
Host:           boao.hu.nor-net  
Uptime:        21 days  
Used Memory:    1 x x86_64; 10 processes; 0 users  
System:         Fedora 23 (Twenty Three) with 4.1.27-888.fc23.x86_64  
Load:          0.24, 0.10, 0.09  
Used Memory:    28 MiB of 512 MiB ( 307 MiB free)  
Used Swap:      363 MiB of 1023 MiB ( 660 MiB free)  
Used Diskspace: 8% on /, 8% on /home  
SSH Keys:  
SHA256:Gt2q0mhIIappz2TyYdKr8l/d8m+06zlbgYoIy6jVBZc (ECDSA 256)  
SHA256:hGoMsrbU0ezy2/+sKvtr+Ijt3AKTWHqvFwL7t2JagEc (ED25519 256)  
SHA256:nod5ULV6dr8npnjzb8I+o5g+b1lyWjaLrmA5Z8lTn8 (RSA 2048)
```

```
Network:  
eth0:          10.80.88.125  10.81.88.125  
                2001:700:4100:5058::7d:66  
                2001:700:4100:5158::7d:66  
  
srl_test@boao.hu.nor-net $ su  
srl_test@boao.hu.nor-net # id  
uid=0(root) gid=0(root) groups=0(root)  
srl_test@boao.hu.nor-net # █
```

```
dreibh : oesthorn
```

- SSH login to the sliver
- Here: 2 ISPs
 - 2x IPv4
 - 2x IPv6
- Kernel with MPTCP
- For security:
 - SSH key fingerprints
- Superuser (“su” or “sudo”)
 - dnf install ...
 - tcpdump -i eth0 ...
 - tshark -i eth0 ...

Overview: Conclusion

- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed
- The Software: VMs, Containers and Multi-Homed Networking
- Conclusion

Conclusion and Future Work

- Multi-homed devices increasingly widespread → multi-path transfer
- Realistic, large-scale Internet testbed infrastructure is available: NorNet
- NorNet Core is an open testbed!
 - Interested in using NorNet? Just ask!

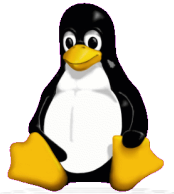


- Future work: **extend NorNet Core's scope *beyond* multi-path transport topic:**
 - Network Function Virtualisation (NFV) and Software-Defined Networking (SDN)
 - Cloud Computing and applications

See <https://www.nntb.no> for more information!

Open Source Software

- NorNet Project: <https://www.nntb.no>
 - NorNet Core Site: <https://www.nntb.no/pub/nor-net-configuration/NorNetCore-Sites.html>
 - NorNet Software: <https://www.nntb.no/software/>
 - Management: <https://github.com/simula/nor-net-control>
 - Research Node: <https://benlomond.nntb.no/releases/>
- HiPerConTracer: <https://github.com/dreibh/hipercontracer>
- NetPerfMeter: <https://www.uni-due.de/~be0001/netperf-meter/>
- RSPLIB: <https://www.uni-due.de/~be0001/rserpool/>
- Linux Multi-Path TCP: <https://multipath-tcp.org/>
- Multi-Path TCP Page: <https://www.uni-due.de/~be0001/mptcp/>
- SCTP Project Page: <http://www.iem.uni-due.de/~dreibh/sctp/>
- Simula Research Laboratory: <https://www.simula.no>



Literature (1)

- Dreibholz, T. and Mazumdar, S.: "Reliable Server Pooling Based Workload Offloading with Mobile Edge Computing: A Proof-of-Concept", in Proceedings of the 2nd International Workshop on Recent Advances for Multi-Clouds and Mobile Edge Computing (M2EC) in conjunction with the 35th International Conference on Advanced Information Networking and Applications (AINA), vol. 3, pp. 582–593, Toronto, Ontario/Canada, May 2021.
- Dreibholz, T.: "NorNet at Hainan University in 2021: Getting Started with NorNet Core – A Remote Tutorial", Tutorial at Hainan University, College of Information Science and Technology (CIST), Haikou, Hainan/People's Republic of China, January 2021.
- Dreibholz, T.: "HiPerConTracer - A Versatile Tool for IP Connectivity Tracing in Multi-Path Setups", in Proceedings of the 28th IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Hvar, Dalmacija/Croatia, September 2020.
- Dreibholz, T.: "NorNet at the University of Sydney: From Simulations to Real-World Internet Measurements for Multi-Path Transport Research", Invited Talk at University of Sydney, Sydney, New South Wales/Australia, January 2019.
- Dreibholz, T.: "NetPerfMeter: A Network Performance Metering Tool", in Multipath TCP Blog, September 2015.

Literature (2)

- Gran, E. G.; Dreibholz, T. and Kvalbein, A.: "NorNet Core – A Multi-Homed Research Testbed", in Computer Networks, Special Issue on Future Internet Testbeds, vol. 61, pp. 75–87, March 2014.
- Ford, A.; Raiciu, C.; Handley, M. and Bonaventure, O.: "TCP Extensions for Multipath Operation with Multiple Addresses", RFC 6824, IETF, January 2013.
- Dreibholz, T.: "Evaluation and Optimisation of Multi-Path Transport using the Stream Control Transmission Protocol", Habilitation Treatise, University of Duisburg-Essen, Faculty of Economics, Institute for Computer Science and Business Information Systems, March 2012.
- Stewart, R. R.: "Stream Control Transmission Protocol", RFC 4960, IETF, September 2007.
- Dreibholz, T.: "Reliable Server Pooling – Evaluation, Optimization and Extension of a Novel IETF Architecture", University of Duisburg-Essen, Faculty of Economics, Institute for Computer Science and Business Information Systems, March 2007.

Also see <https://www.nntb.no/publications/> !



Thank you for your attention!
Any questions?

Thomas Dreibholz
dreibh@simula.no
<https://www.simula.no/people/dreibh>