LinuxCon Europe 2016

NorNet
–
Building an Inter-Continental Internet Testbed based on Open Source Software

Thomas Dreibholz
Simula Research Laboratory
dreibh@simula.no

October 5, 2016

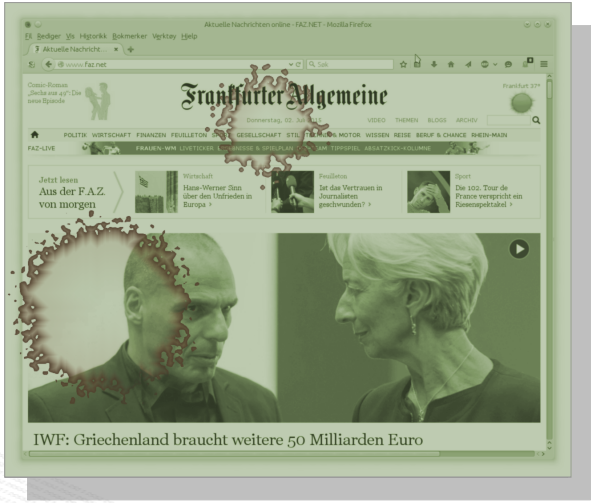
Contents

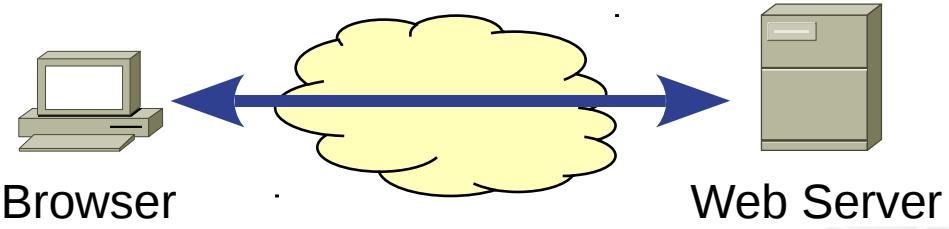
Overview:
Motivation

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

„Classic“ Internet Communication

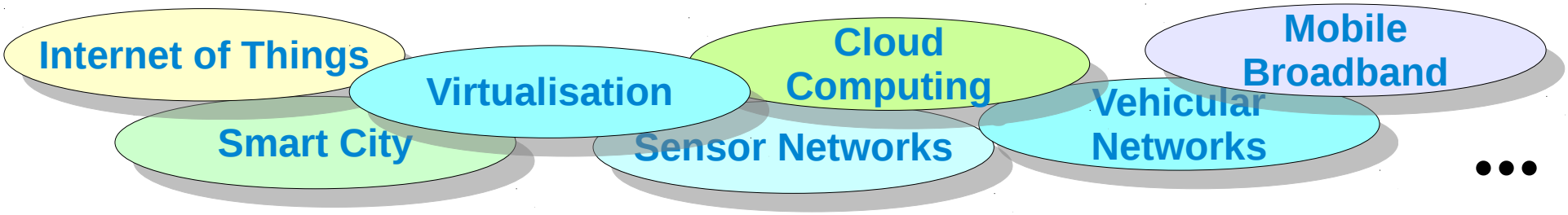
- Example: World-Wide Web





- 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

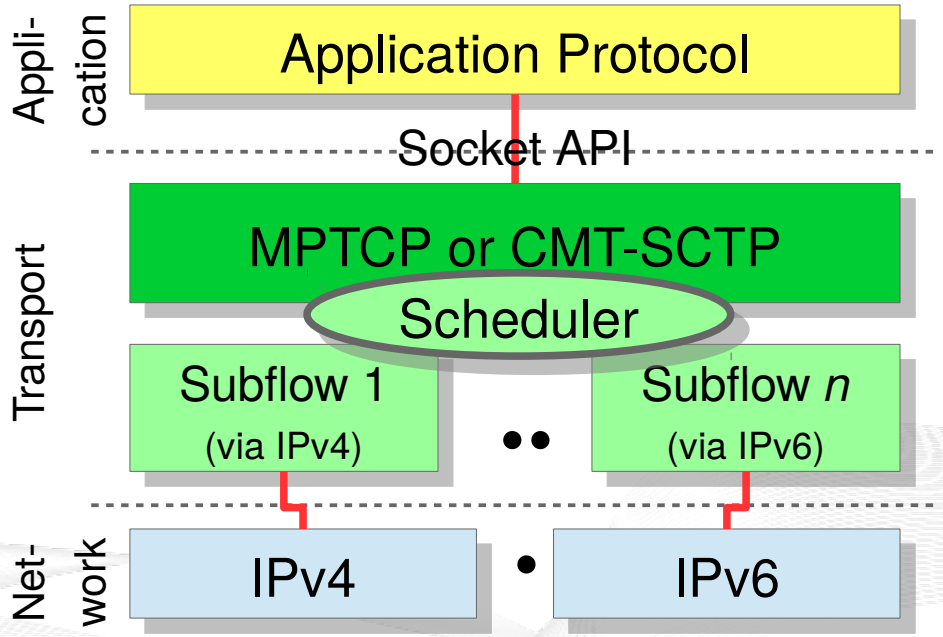
Hot topic in research and standardisation!

Overview:
Multi-Homing and Multi-Path Transport

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

Multi-Path Transport with MPTCP and CMT-SCTP

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



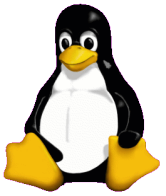
How can I use multi-path transport?

Stream Control Transmission Protocol (SCTP)

- RFC 4960
- Main features:
  - **Multi-homing**
  - **Multi-streaming** (independent streams over one connection)
  - Many **extensions**, e.g. partial reliability, address reconfiguration ...
- Linux: in mainline kernel → works out of the box!
  - Lacks of many features, unfortunately – like CMT-SCTP ☹
  - Interested in helping → master student projects possible!
- FreeBSD: SCTP reference implementation → included in kernel already!
  - All the nice features
  - May be somebody could port it to Linux?

Multi-Path TCP (MPTCP)

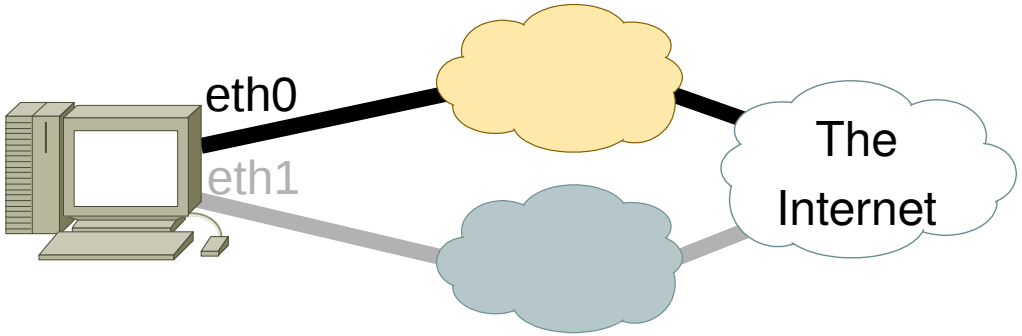
- RFC 6824
- Features: multi-homing + multi-path transport
- Backwards-compatible to TCP
  - Communicate with old TCP implementations
  - Works (mostly) even over non-MPTCP middleboxes (e.g. NAT/PAT)
- Linux:
  - UC Louvain → http://www.multipath-tcp.org
- FreeBSD:
  - Swinburne → http://caia.swin.edu.au/newtcp/mptcp/



How can I use MPTCP under Linux?

Routing Tables and Routing Rules for Multi-Path Transport

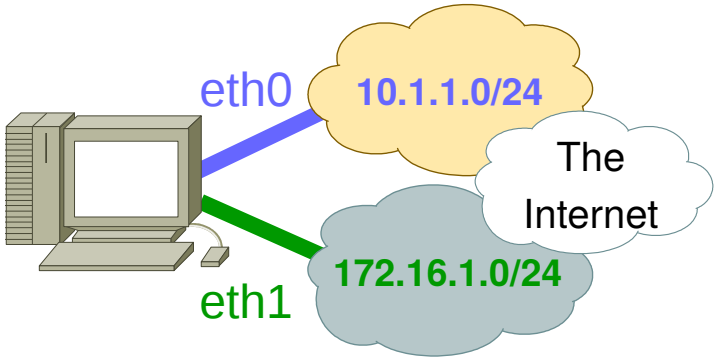
- Example:
  - Device eth0 → ISP 1
  - Device eth1 → ISP 2



- 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/rt_tables!



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
```

- 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 eth0
172.16.1.0/24 dev eth0 scope link
```

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

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

- 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:
  - Wireshark/T-Shark → https://www.wireshark.org
  - NetPerfMeter → https://www.uni-due.de/~be0001/netperfmeter/
  - ...

Overview:
The NorNet Testbed Setup

- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed Setup
- 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

UNIVERSITÄT DUISBURG ESSEN

Fachhochschule Münster University of Applied Sciences

海南大学 Hainan University

- 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!

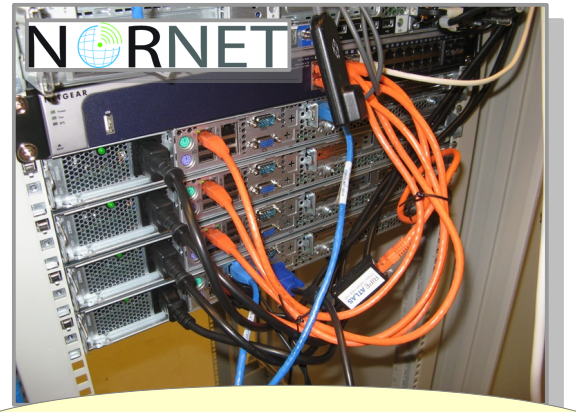
Now really big: NorNet testbed!

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 system „Ufoboard“
  - Up to 4x 2G/3G/4G, 1x CDMA, 1x 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

Work in Progress!

❄ -30°C
Longyearbyen

Our servers may be really remote!

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

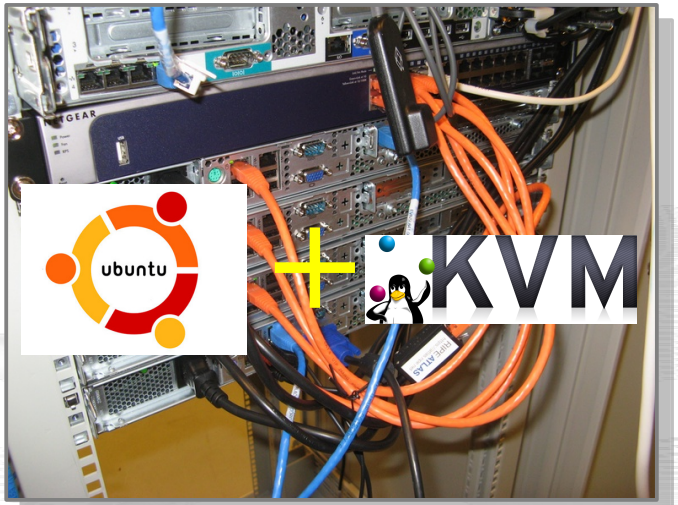
Overview:
The Software: VMs, Containers and Multi-Homed Networking

- Motivation
- Multi-Homing and Multi-Path Transport
- The NorNet Testbed Setup
- 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 14.04 LTS → 16.04 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
    - Not tried ReiserFS 4 (unfortunately not in mainline kernel) or ZFS, yet
  - Virtualisation: now KVM
    - Formerly: 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" ...)

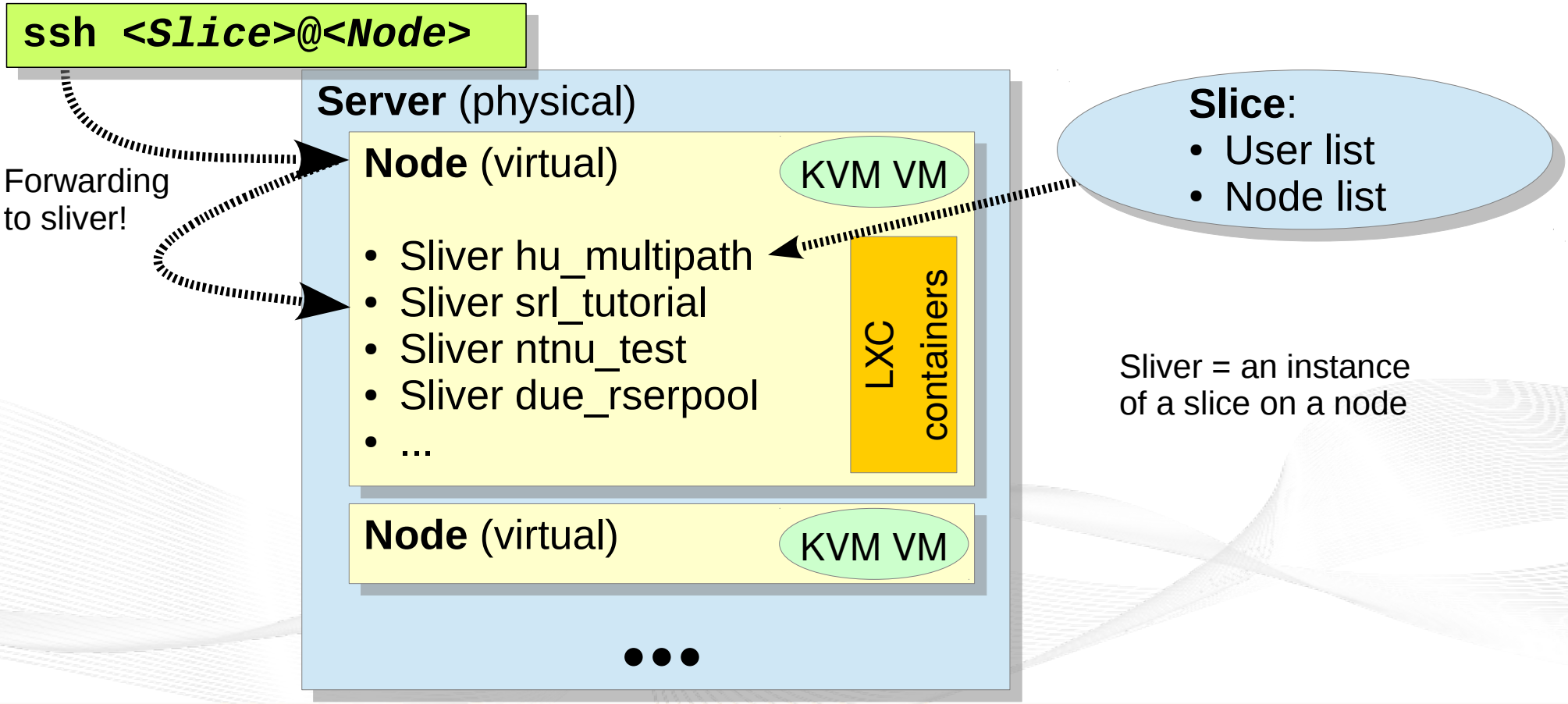
Tunnelbox – The Router (2)

- 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
    - 三亚 .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)

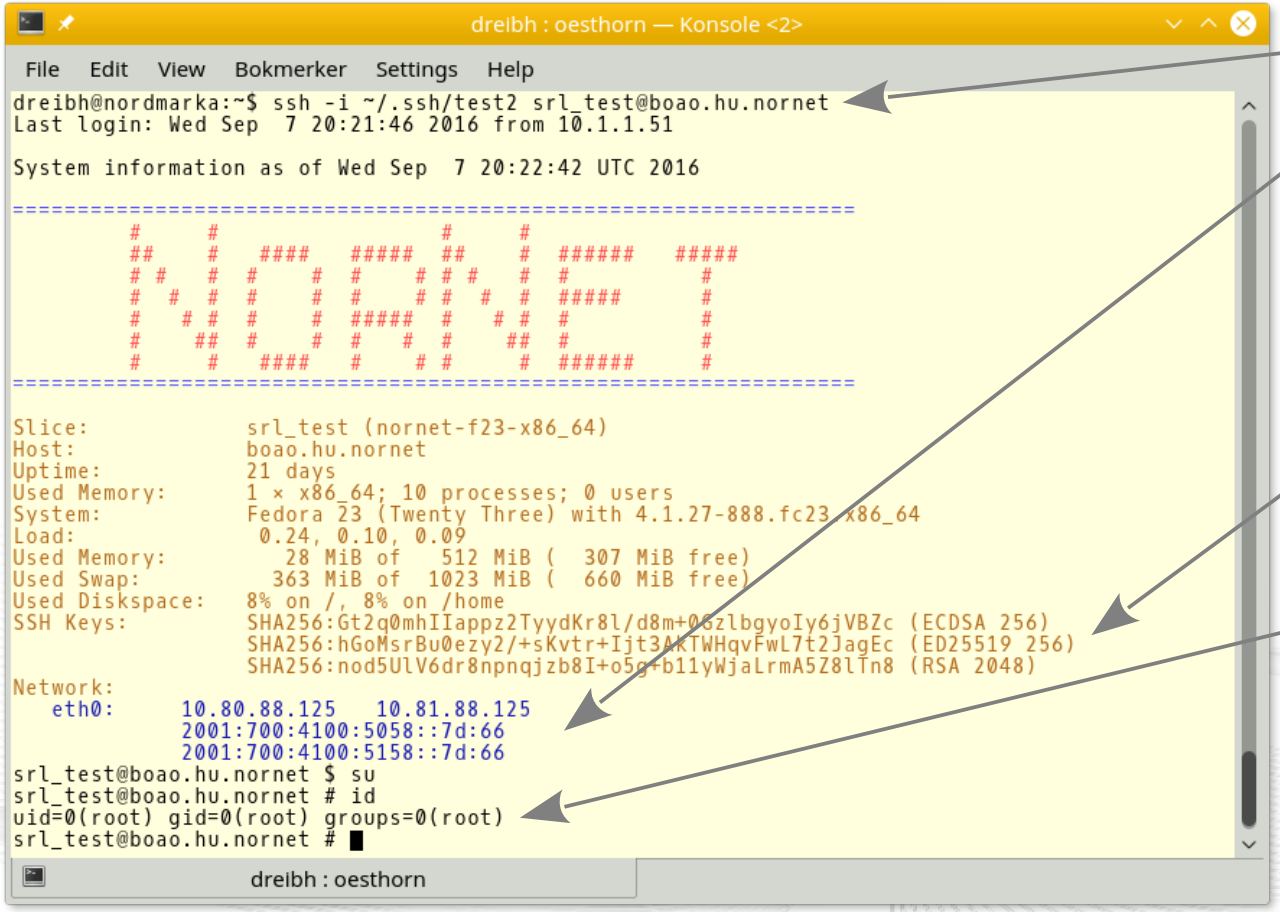
Research Node

- “Usual” research node:
  - A VM (usually KVM), managed by PlanetLab Central (PLC)-based software
    - 2.5 GiB RAM, 1-2 cores
    - Fedora Core 23, Linux kernel v4.1.32 with MPTCP v0.91 + API patch
  - User gets a “sliver” of the research node → LXC container
    - “Own” Fedora installation, with development tools, T-Shark, ...
    - “Own” IP addreses (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



A Look into a Sliver



- 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 Setup
- 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!

Any Questions?

NORNET

https://www.nntb.no

Thomas Dreibholz, dreibh@simula.no

Links

- HiPerConTracer: https://github.com/dreibh/hipercontracer
- Linux Multi-Path TCP: http://multipath-tcp.org
- Multi-Path TCP Page: http://www.iem.uni-due.de/~dreibh/mptcp/
- NetPerfMeter: https://www.uni-due.de/~be0001/netperfmeter/
- NorNet Project: https://www.nntb.no
- NorNet Core Sites: https://www.nntb.no/pub/nornet-configuration/NorNetCore-Sites.html
- NorNet Software: https://www.nntb.no/software/
  - Management: https://github.com/simula/nornet-control
  - Research Node: https://benlomond.nntb.no/releases/
- SCTP Project Page: http://www.iem.uni-due.de/~dreibh/sctp/
- Simula Research Laboratory: https://www.simula.no
- Wireshark/T-Shark: https://www.wireshark.org

Literature (1)

- Dreibholz, T.: "NorNet – The Internet Testbed for Multi-Homed Systems", in Proceedings of the Multi-Service Networks Conference (MSN, Coseners), Abingdon, Oxfordshire/United Kingdom, July 2016.
- Amer, P. D.; Becke, M.; Dreibholz, T.; Ekiz, N.; Iyengar, J. R.; Natarajan, P.; Stewart, R. R. and Tüxen, M.: "Load Sharing for the Stream Control Transmission Protocol (SCTP)", Internet Draft draft-tuexen-tsvwg-sctp-multipath-12, IETF, Individual Submission, June 2016.
- Dreibholz, T.: "An Experiment Tutorial for the NorNet Core Testbed at Hainan University", Tutorial at Hainan University, College of Information Science and Technology (CIST), Haikou, Hainan/People's Republic of China, May 2016.
- Dreibholz, T.: "An Experiment Tutorial for the NorNet Core Testbed at HAW Hamburg", Tutorial at Hochschule für Angewandte Wissenschaften Hamburg (HAW Hamburg), Hamburg/Germany, March 2016.
- Dreibholz, T.: "An Experiment Tutorial for the NorNet Core Testbed at NICTA", Tutorial at National Information Communications Technology Australia (NICTA), Sydney, New South Wales/Australia, January 2016.
- Fa, F.; Zhou, X.; Dreibholz, T.; Wang, K.; Zhou, F. and Gan, Q.: "Performance Comparison of Congestion Control Strategies for Multi-Path TCP in the NorNet Testbed", in Proceedings of the 4th IEEE/CIC International Conference on Communications in China (ICCC), pp. 607–612, Shenzhen, Guangdong/People's Republic of China, November 2015.

Literature (2)

- Livadariu, I. A.; Ferlin, S.; Alay, Ö.; Dreibholz, T.; Dhamdhere, A. and Elmokashfi, A. M.: "Leveraging the IPv4/IPv6 Identity Duality by using Multi-Path Transport", in Proceedings of the 18th IEEE Global Internet Symposium (GI) at the 34th IEEE Conference on Computer Communications (INFOCOM), pp. 312–317, Hong Kong/People's Republic of China, April 2015.
- Golkar, F.; Dreibholz, T. and Kvalbein, A.: "Measuring and Comparing Internet Path Stability in IPv4 and IPv6", in Proceedings of the 5th IEEE International Conference on the Network of the Future (NoF), pp. 1–5, Paris/France, December 2014.
- 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.

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