Software Defined Networking

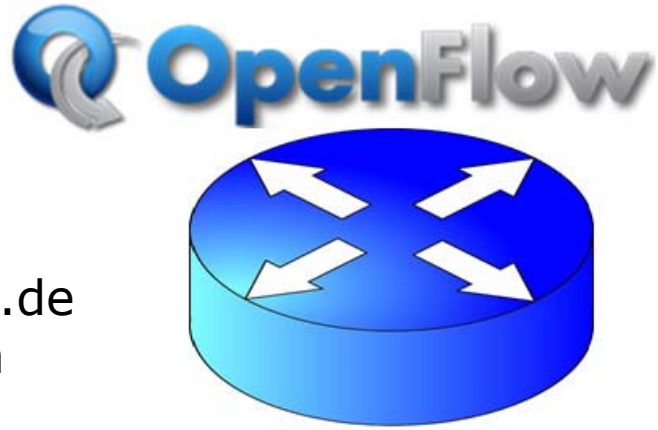
SDN Hardware and Use Case

TECHNISCHE UNIVERSITÄT DARMSTADT

Jeremias Blendin, Julius Rückert, David Hausheer

Department of Electrical Engineering
and Information Technology
Technische Universität Darmstadt

E-Mail: firstname.lastname@ps.tu-darmstadt.de
http://www.ps.tu-darmstadt.de/teaching/sdn



*Original slides for this lecture provided by Jeremias Blendin, Julius Rückert, and David Hausheer (TU Darmstadt)

SDN Exam

TECHNISCHE UNIVERSITÄT DARMSTADT

- Date/Time: 09.03.16, 14:30 - 16:30
- Rooms: S101/A1 (Audimax), A4 and A5

Lecture Overview

TECHNISCHE UNIVERSITÄT DARMSTADT

- OpenFlow Hardware Switches
- SDM: Software-Defined Multicast

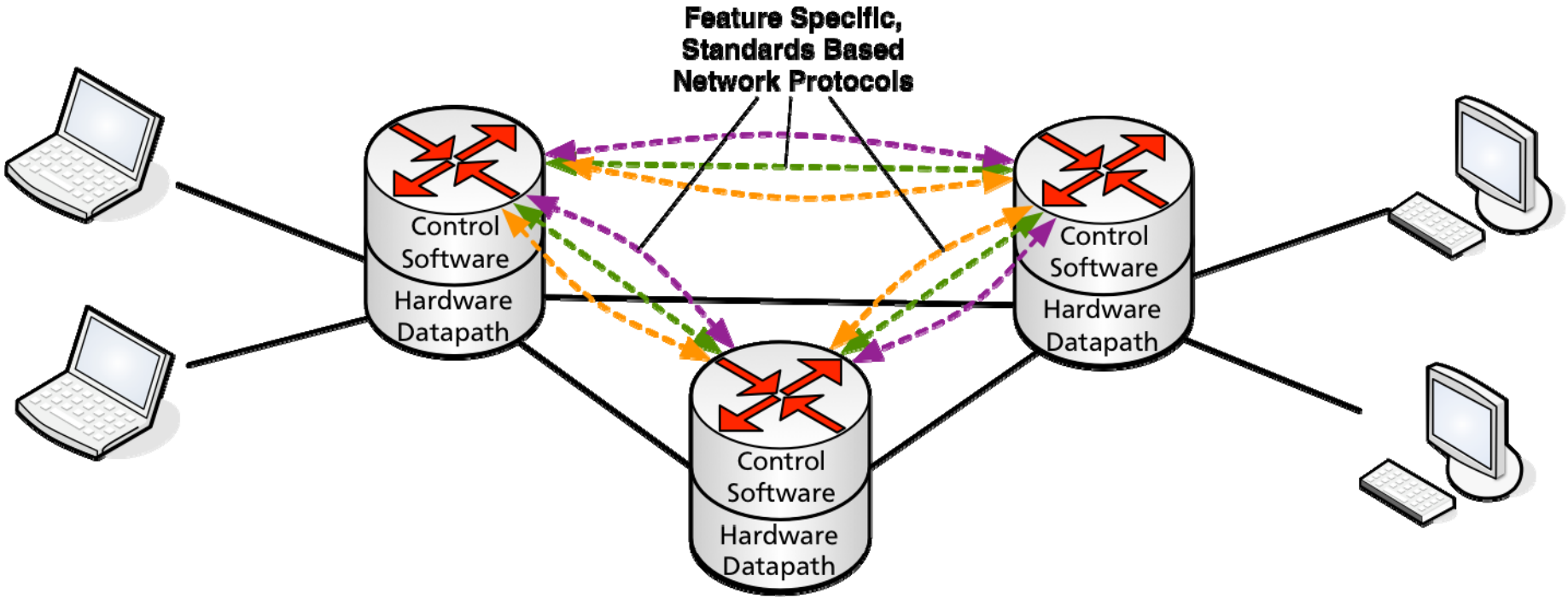
OpenFlow Hardware Switches

David Hausheer, Jeremias Blendin, Fabian Kaup, Leonhard Nobach, Julius Rückert, Matthias Wichtlhuber

Reminder: The OpenFlow Concept in a Nutshell

TECHNISCHE UNIVERSITÄT DARMSTADT

- Traditional:

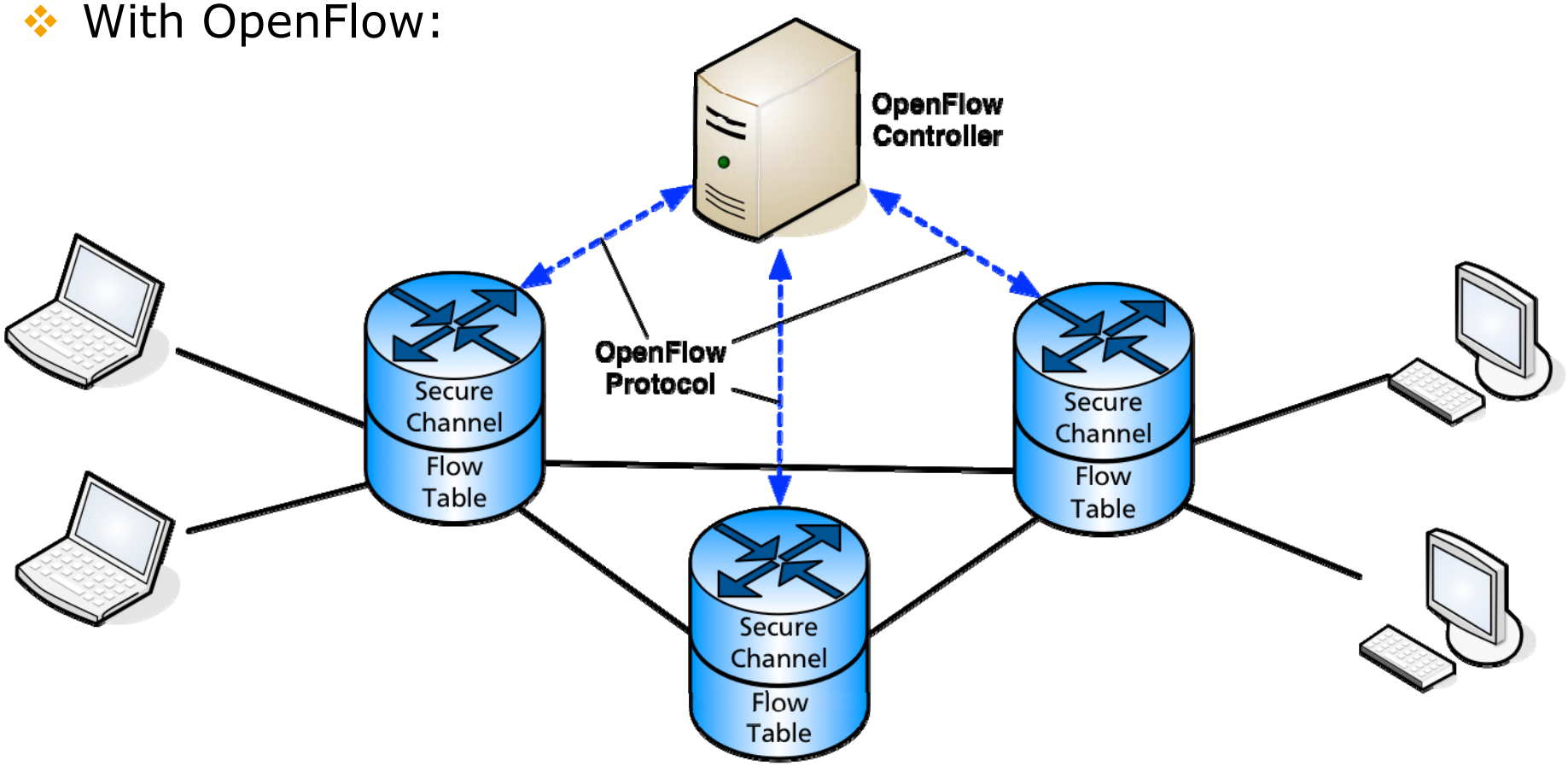


Source: [2]

Reminder: The OpenFlow Concept in a Nutshell

TECHNISCHE UNIVERSITÄT DARMSTADT

- With OpenFlow:



Source: [2]

The Promise of OpenFlow

TECHNISCHE UNIVERSITÄT DARMSTADT

- Centralized control of multi-vendor environments
  - SDN control software can control any OpenFlow-enabled network device
  - From any vendor, including switches, routers, and virtual switches. [1]

- Make the software independent of the hardware
  - Create a big hardware market
  - Forwarding hardware as a commodity
  - Market mechanisms should drive prices down

OpenFlow Hardware

TECHNISCHE UNIVERSITÄT DARMSTADT

- If the forwarding hardware is a commodity, why do we care?

- Situation today
  - Hardware is not a commodity (yet)
  - Huge differences in the performance of OF devices
  - Even differences in the performance of single OpenFlow features on a single device

For centralized control and programmability, we need to understand the devices‘ performance characteristics

OpenFlow Switch Performance

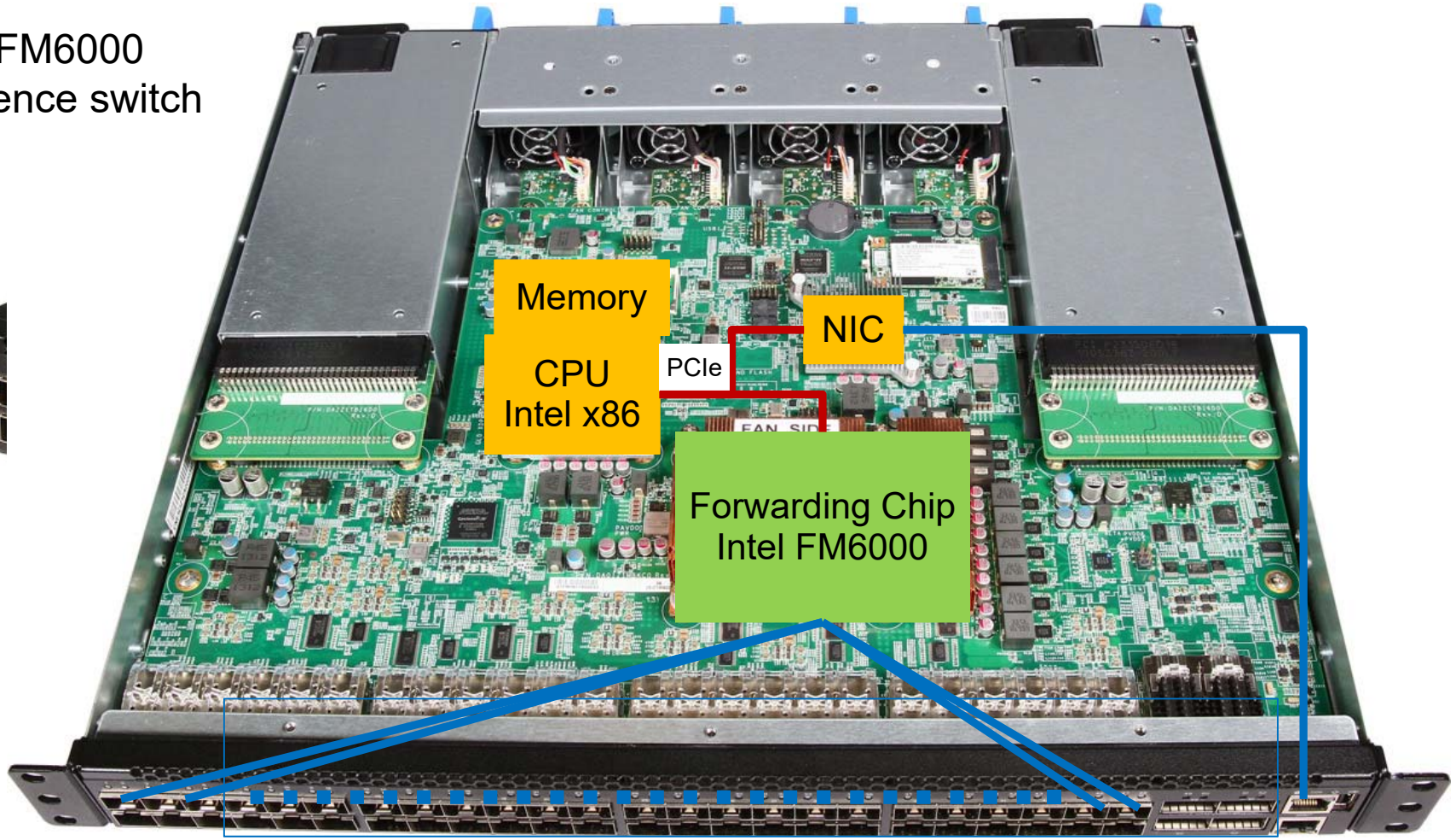
TECHNISCHE UNIVERSITÄT DARMSTADT

- Management
  - Flow table changes per second
- Matcher
  - How many matchers can be installed?
  - Do all matchers perform the same?
    - Depend on the number of existing matchers?
- Actions
  - How many actions can be put in a list?
  - Do all actions perform the same?
    - Depend on the position in the action list?
    - Depend on the action itself?

How does a Switch / Router Look on the inside?

TECHNISCHE UNIVERSITÄT DARMSTADT

Intel FM6000
reference switch

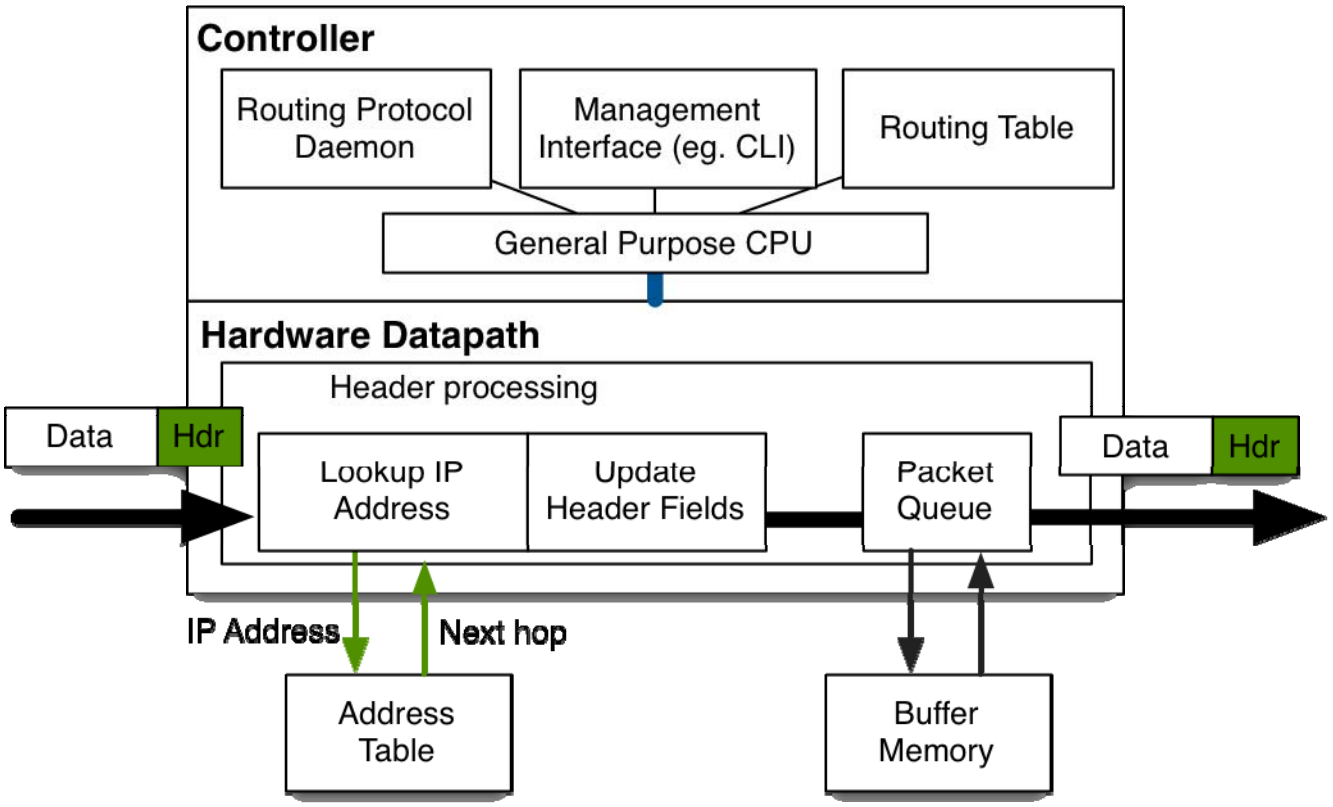


Source: http://sysmagazine.com/posts/212639/

IP Router Architecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- Controller
  - „Normal" CPU
  - Runs the management software
- Datapath
  - Specialized hardware
  - Configured by the controller

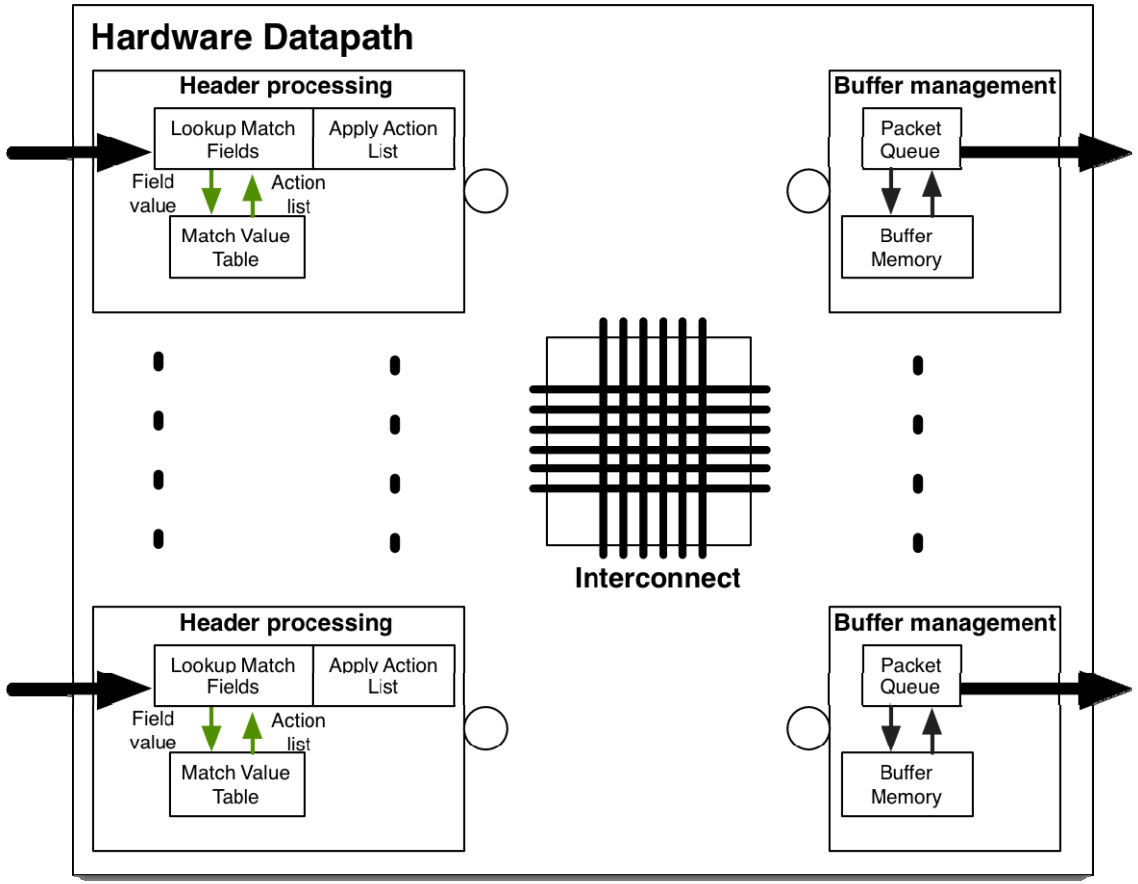


Adapted from: McKeon, CS343 slides, Stanford University, 2003

IP Router Architecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- Scaling
  - Multiple header processing units
  - Multiple buffer manager
  - Interconnection

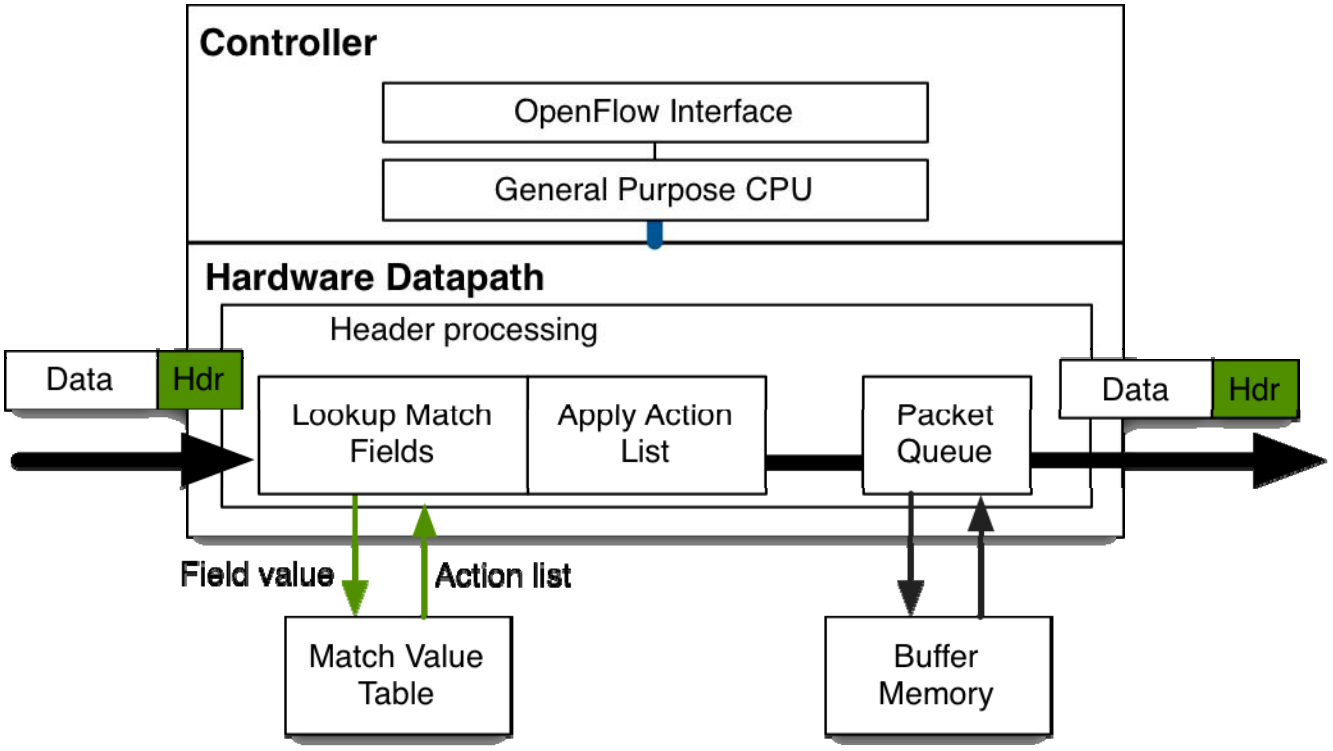


Adapted from: McKeon, CS343 slides, Stanford University, 2003

OpenFlow Switch Architecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- Controller
  - „Normal“ CPU
  - Runs the OpenFlow interface
- Datapath
  - Specialized hardware
  - Configured by OpenFlow rules



Adapted from: McKeon, CS343 slides, Stanford University, 2003

OpenFlow Switch Architecture

TECHNISCHE UNIVERSITÄT DARMSTADT

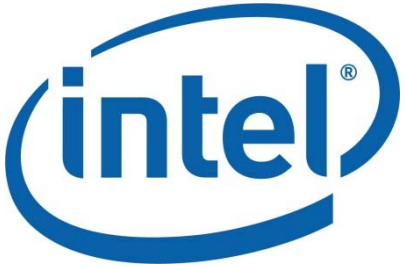
- OpenFlow 1.0 closely resembles the IP router architecture
  - Queuing is not implemented by OpenFlow
  - OpenFlow 1.1 introduces a more refined approach

- Crucial components
  - Matcher memory
    - Size
    - Performance
  - Actions
    - Performance
    - List length

Packet Processing: Memory

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example: Intel FM6000 Series
  - E.g. Arista 7150S, often used by stock traders
  - 1 gigapacket/s throughput
- Header bits look-up performance
  - $\frac{1}{1E9} s = 1ns$ per packet
- Memory performance
  - DDR3 2Ghz
  - $\sim 4ns$ access latency
  - Not fast enough



Content-Addressable Memory

TECHNISCHE UNIVERSITÄT DARMSTADT

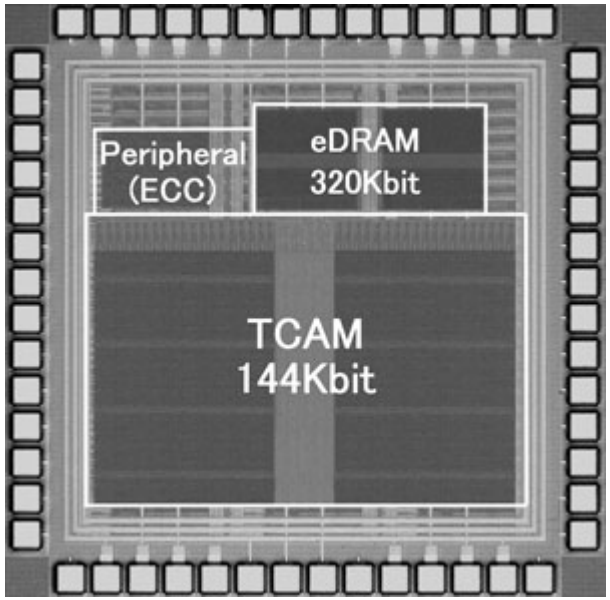
- General idea
  - Memory that matches a given bit string in O(1)
  - Lookup only takes a few clock cycles
  - Returns an associated value
- TCAM: Ternary content-addressable memory
  - Matches a given bit string with don‘t care bits in O(1)
  - Examples:
    - IP routing: Longest prefix match
    - Access control lists: compare IP addresses
  - Used widely today

TCAM Limits

TECHNISCHE UNIVERSITÄT DARMSTADT

- High power consumption
- High price
- Large (on die)

- Capacity
  - E.g. 36Mbit in high end device
    - 1,125,000 IPv4 entries
    - 281,250 IPv6 entries
  - Full Internet routing table
    - 481,743 IPv4 routes
    - 15,398 IPv6 routes
    - Both tables: ~17,4 Mbit



Source: Renesas

OpenFlow and TCAM

TECHNISCHE UNIVERSITÄT DARMSTADT

- OpenFlow increases the flexibility of TCAM usage
- Not only ACLs or IP prefixes are matched
- Different OpenFlow matchers
  - IPv4 address: 32bit
  - IPv6 address: 128bit
  - MAC address: 48bit

- Available OpenFlow versions do not allow to query the switch for memory capacity
- Approaches if TCAM is full
  - Add OF matcher to low performance memory
  - Refuse additional rules

Processor Options

TECHNISCHE UNIVERSITÄT DARMSTADT

- CPU
  - X86, ARM, MIPS
- NPU
  - Similar to GPU, but for networking applications
  - Large number of specialized network processors
- FPGA
  - Programmable Hardware
- Switch ASIC
  - Most of the relevant functions are implemented in hardware

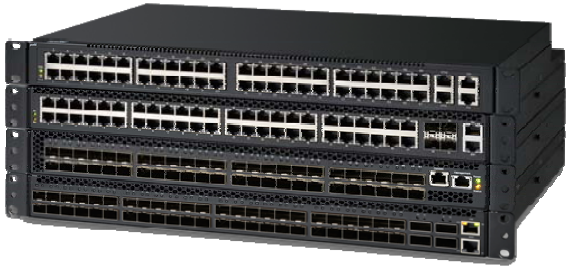
OpenFlow Hardware Classes

TECHNISCHE UNIVERSITÄT DARMSTADT

- Class: SOFTWARE
  - OpenFlow features added to existing devices
  - Most features in software, only some in hardware
  - Most available devices today

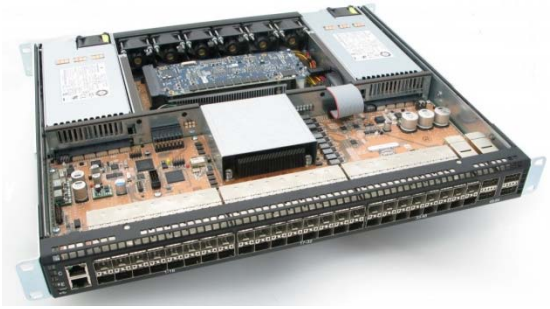
Example: HP ProCurve 3500yl

- Class: DEVICE
  - Purpose built device for OpenFlow
  - Based on existing processors
  - Only few devices today



Example: Pica 8

- Class: HARDWARE
  - ASIC adapted to OpenFlow
  - Very few devices today



Example: Intel FM6700

Commercial OpenFlow Device Examples (2013)

TECHNISCHE UNIVERSITÄT DARMSTADT

Class	Name	Processor		Performance (Gb/s)		Matcher		Change Fields		
		Type	Model	max.	OpenFlow	All in TCAM	Max.	L2	L3	L4
Software	HP 3500 yl	ASIC	HP ProVision	150	?	no	3055	hw/sw	sw	sw
Software	IBM 8264T	ASIC	Broadcom Trident II	1280	?	Yes	750	hw	no	no
Software	Arista 7050									
Device	NEC PF5240	?	?	176	?	Yes	160,000	hw	sw	sw
Device	Pica8 P-3290	?	?	176	?	no	4000	hw	no	no
Device	Centec V330	ASIC	CTC6048	176	?	yes	2500	hw	hw/no	hw
Device	NoviSwitch 1132	NPU	EzChip NP-4	320	100	yes	500,000	hw	hw	hw
Hardware	Netronome Network Flow Processing Platform	CPU, NPU, ASIC	Intel Xeon, NFP-3240	? 200 480	? 200 ?	no	5,000,000	hw	hw	hw
Hardware	Intel FM6700 Reference	ASIC	Intel FM6764	640	?	yes	?	hw	hw	hw
Reference for comparison: Software Switch based on PC hardware										
N/A	Open vSwitch	CPU	Intel Core i7	20	N/A	no	N/A	sw	sw	sw

Sources: Vendor information, [2]

Example: HP ProCurve OpenFlow Series

TECHNISCHE UNIVERSITÄT DARMSTADT

- HP took existing ProCurve switches and added OpenFlow features
- VLAN and STP processing before OpenFlow
- Wildcard rules or non-IP pkts processed in s/w
- MAC address rewriting is done in software

HP ProCurve 3500yl

HP ProCurve 5400

HP ProCurve 8200zl

Example: NEC Switches

TECHNISCHE UNIVERSITÄT DARMSTADT

- Specifically designed for OpenFlow
- OpenFlow takes precedence
- Many actions processed in hardware
- MAC address rewriting in HW
- IP address rewriting in SW
- Complex OpenFlow configuration, many feature dependencies

NEC IP8800

NEC PF5240

Further OpenFlow Hardware Examples

TECHNISCHE UNIVERSITÄT DARMSTADT

Juniper MX-series

Netgear 7324

WiMax (NEC)

Ciena Coredirector

Pronto 3240/3290

PC Engines

Adapted from: Xenofontas Dimitropoulos, ETH Zurich

Literature

[1] Software-Defined Networking: The New Norm for Networks, Open Network Foundation, 2012

[2] G. Pongrácz, L. Molnár, Z. L. Kis, Z. Turányi: Cheap Silicon: a Myth or Reality? Picking the Right Data Plane Hardware for Software Defined Networking, HotSDN 2013

TECHNISCHE UNIVERSITÄT DARMSTADT

SDM: Software-Defined Multicast

Using OpenFlow to Push Overlay Streams into the Underlay / An OpenFlow-based Cross-Layer Approach for Overlay Live Streaming

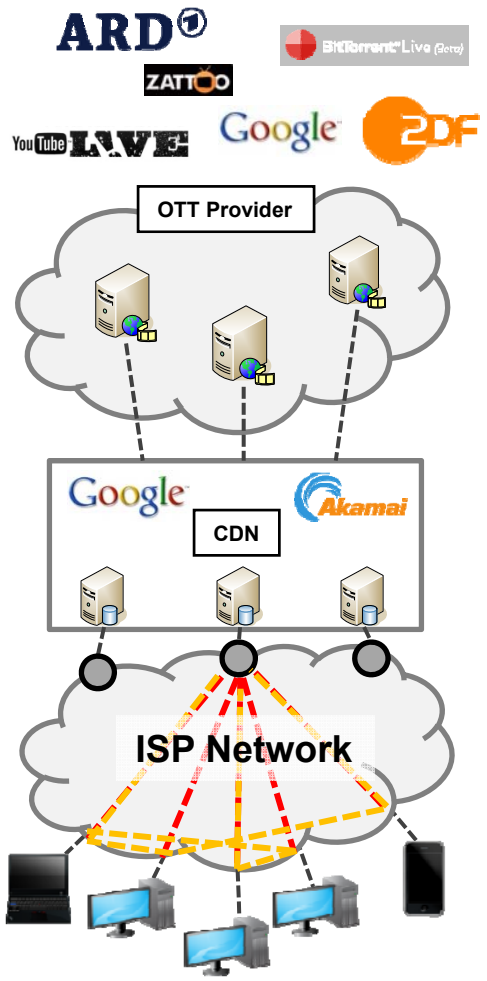
Julius Rückert, Jeremias Blendin, David Hausheer

SDM: Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Current Situation on the Internet
  - Increasing number of OTT live streaming services
  - Content Delivery Networks (CDNs)
    - Used to improve global content delivery process
    - Usually end at edge of Internet Service Provider (ISP)
- Delivery of Live Content Today
  - Locally: IP multicast (e.g. ISP-internal IPTV)
  - Globally: IP unicast with CDN support
    - ✘ High load on ISP border and internal network
    - ✘ P2P delivery can only help to reduce content provider and CDN load

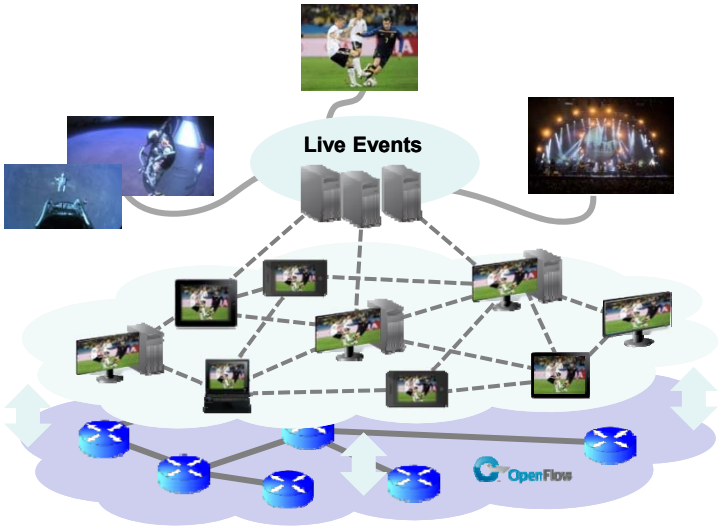
Research Question
How to support efficient large-scale OTT and P2P live video streaming inside ISP networks?

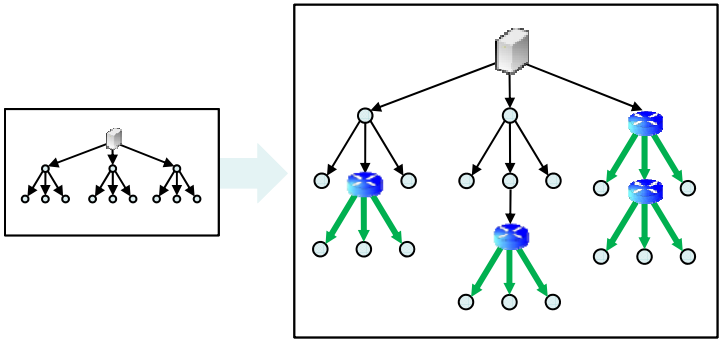


SDM: Software-Defined Multicast

TECHNISCHE UNIVERSITÄT DARMSTADT

- OpenFlow-based cross-layer approach
- Allows ISPs to provide network-layer multicast as service for global OTT and P2P live streaming [BI13, RBH13a, RBH13b]
- High-level Features
  - Network layer multicast data delivery with delivery performance and costs at the level of IP multicast
  - Virtual presence of traffic source inside ISP network (with super peer capabilities)
  - Multicast delivery transparent to receivers
  - Management of service and traffic under full control of ISP

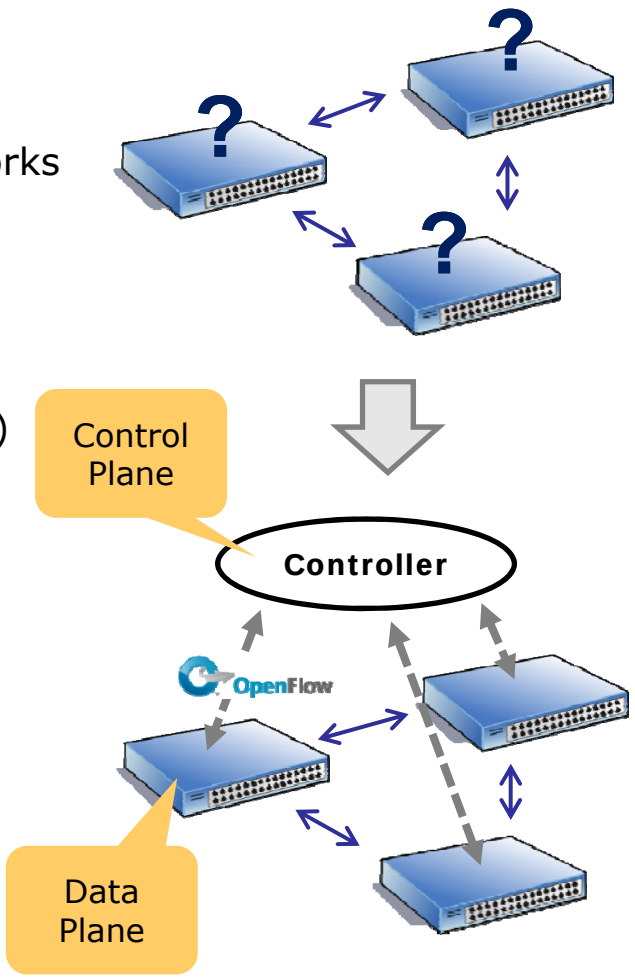




Reminder: Software-Defined Networking (SDN)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Network hardware often black boxes
  - Closed-source, vendor-specific solutions
  - Complex to manage, control, and monitor large networks
- Idea of SDN and OpenFlow:
  - Separate control and data forwarding functionality
  - Provide standard interface to interact with forwarding functionality of network hardware (OpenFlow protocol)
  - Logically central control of the network infrastructure
- Two planes, different objectives
  - Control plane:
    - Network topology discovery
    - Network-wide traffic management
  - Data plane:
    - Line-speed data switching/routing
    - Run as much as possible in hardware



SDM: Architecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- SDM Controller
  - Provides external SDM service API
  - Service provisioning and admission
- NL-SDM Component
  - Network-layer multicast functionality
  - Planning and management of multicast tree instances
- Virtual Peer Component
  - Network-layer proxy functionality
  - Gives outside OTT or peer a virtual presence inside the ISP network
- ISP Network
  - OpenFlow-enabled switches/routers

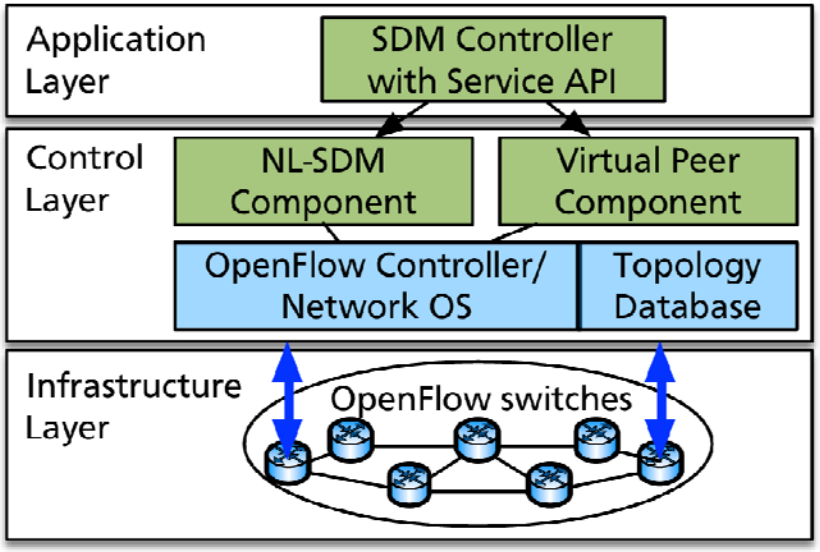
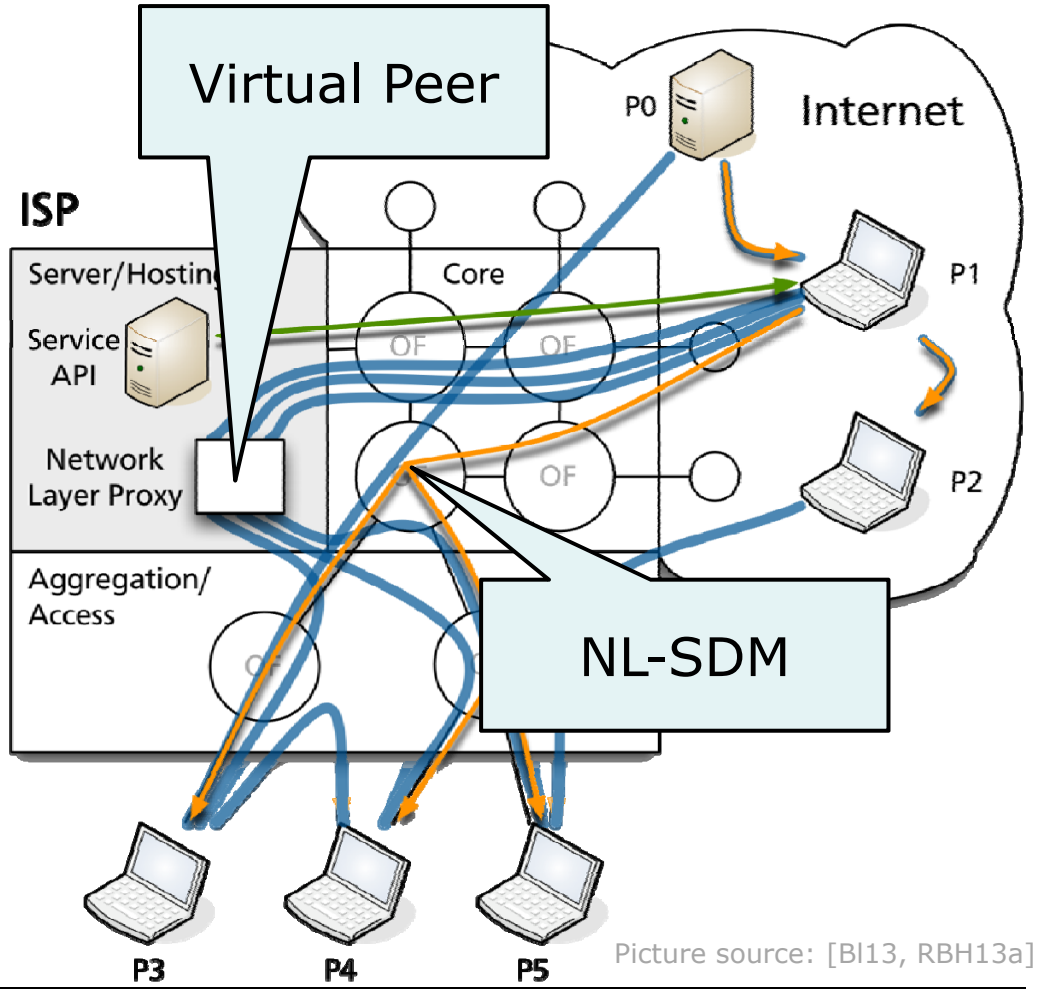


Figure source: [Bl13, RBH13a]

SDM: High-level Concept

TECHNISCHE UNIVERSITÄT DARMSTADT

- Overlay Properties
  - P2P streaming systems highly rely on tree overlay topologies
  - OTT delivery can be seen as tree overlay topology with a single level
- Process
  1. Establish API connection
  2. Create SDM Instance
  3. Promote virtual identity to peers/clients inside ISP
  4. Overlay connections pass through network-layer proxy (virtual peer)
  5. Push-based content delivery to ISP-internal peers/clients through network layer multicast (NL-SDM)



Picture source: [Bl13, RBH13a]

SDM: Interaction between entities

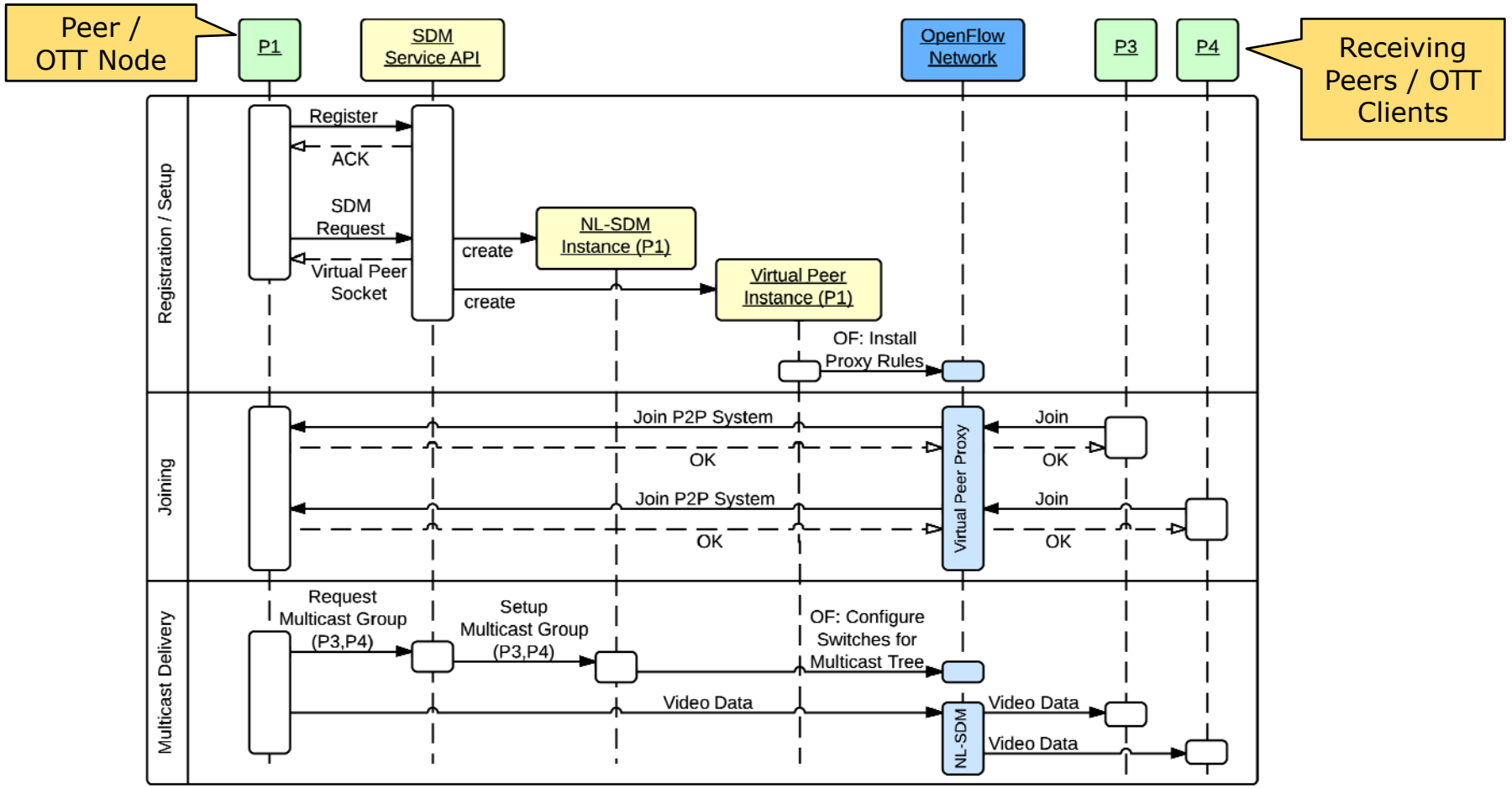


Figure source: [Bl13]

NL-SDM Component

TECHNISCHE UNIVERSITÄT DARMSTADT

1. SDM application calculates multicast tree to reach a set of given group members
2. Source peer/OTT node sends packets to specific IP unicast address as defined by ISP
3. Ingress switches mark packet with group ID (e.g. by rewriting destination MAC address)
4. Internal switches are configured according to multicast tree (on matching group ID: forward or duplicate packet)
5. Egress switches remove mark and rewrite destination MAC and IP to receiver's address

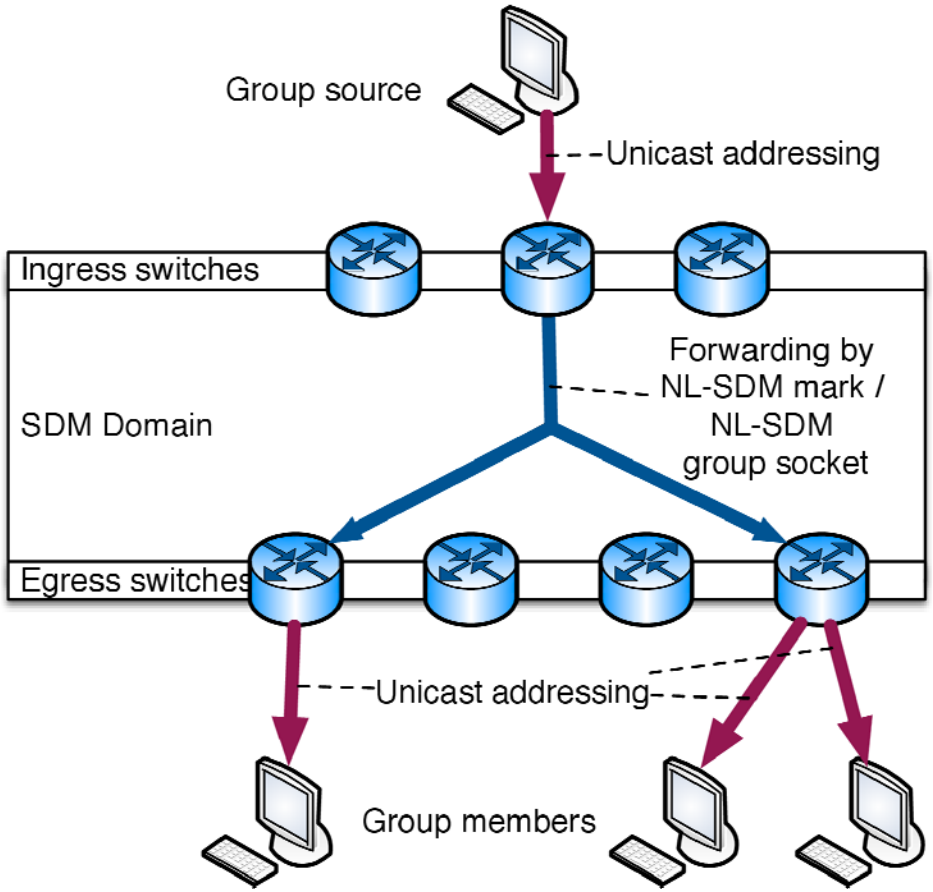


Figure source: [Bl13]

SDM: Qualitative Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

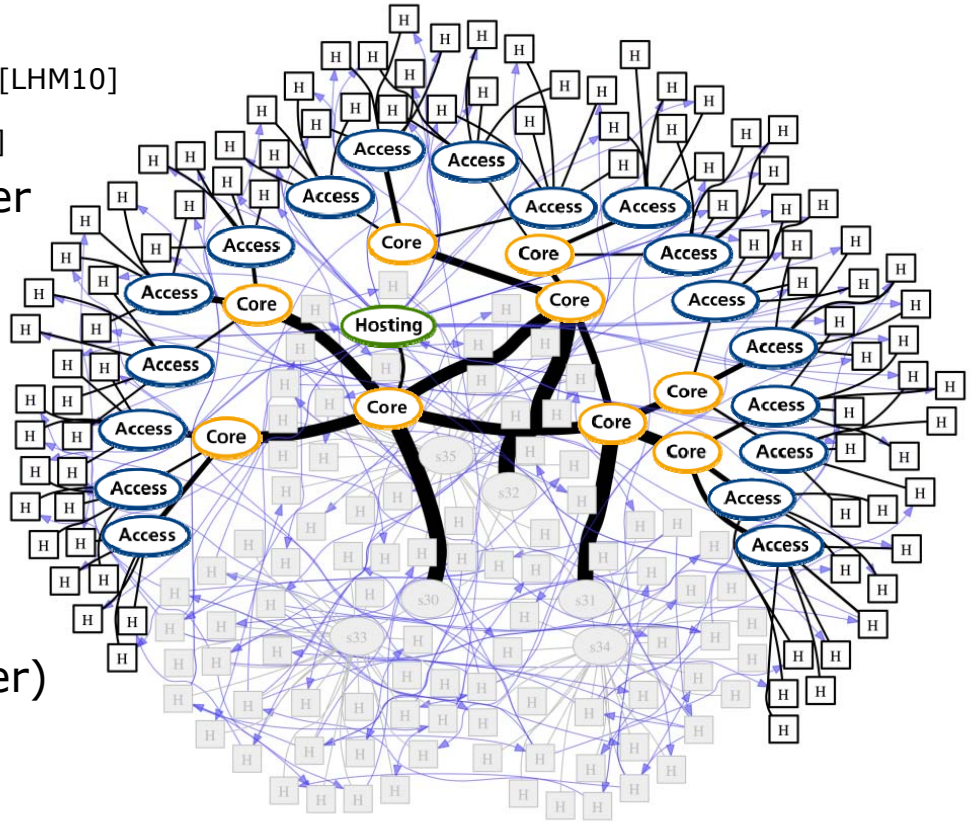
	IP Multicast	Overlay Multicast	SDM
ISP control	• Traffic control and admission hard to achieve	• None / very limited: High volumes of hard to control traffic	• Full traffic and admission control
Delivery efficiency	• **High**: Traffic duplication at routers (L3)	• **Low**: Traffic duplication at clients	• **Very high**: Traffic duplication at switches (L2)
Network Deployment Requirements	• Requires multicast-capable routers	• Directly deployed on the current Internet (overlay)	• Requires OpenFlow support within the ISP network
Content Provider Requirements	• Use of multicast protocol	• Use of overlay multicast protocol	• usage of API by Content Provider or OTT node
Client-side Requirements	• Clients need to support IP multicast protocol	• Clients run overlay multicast application	• **None**: Transparent IP unicast delivery
Transport Protocol	• UDP only	• No limitation	• UDP only
Scalability	• **Low:** limited scalability of multicast-capable routers	• **High:** fully distributed and scalable protocols are available	• Depends on OpenFlow hardware and management architecture

Based on [Bl13,RBH13b]

Quantitative Evaluation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Setting and tools
  - Mininet-based virtual network [LHM10]
  - Open vSwitch instances [PGP+10]
  - Ryu[1]-based OpenFlow controller
  - Real UDP-based video stream
  - ISP-like network structures
- Focus of the evaluation
  - Influence of ISP/overlay topology
  - Transmission efficiency (network traffic per served peer)
    - Intra-ISP network traffic
    - Network traffic at ISP border
  - Costs: Number of OpenFlow rules and messages



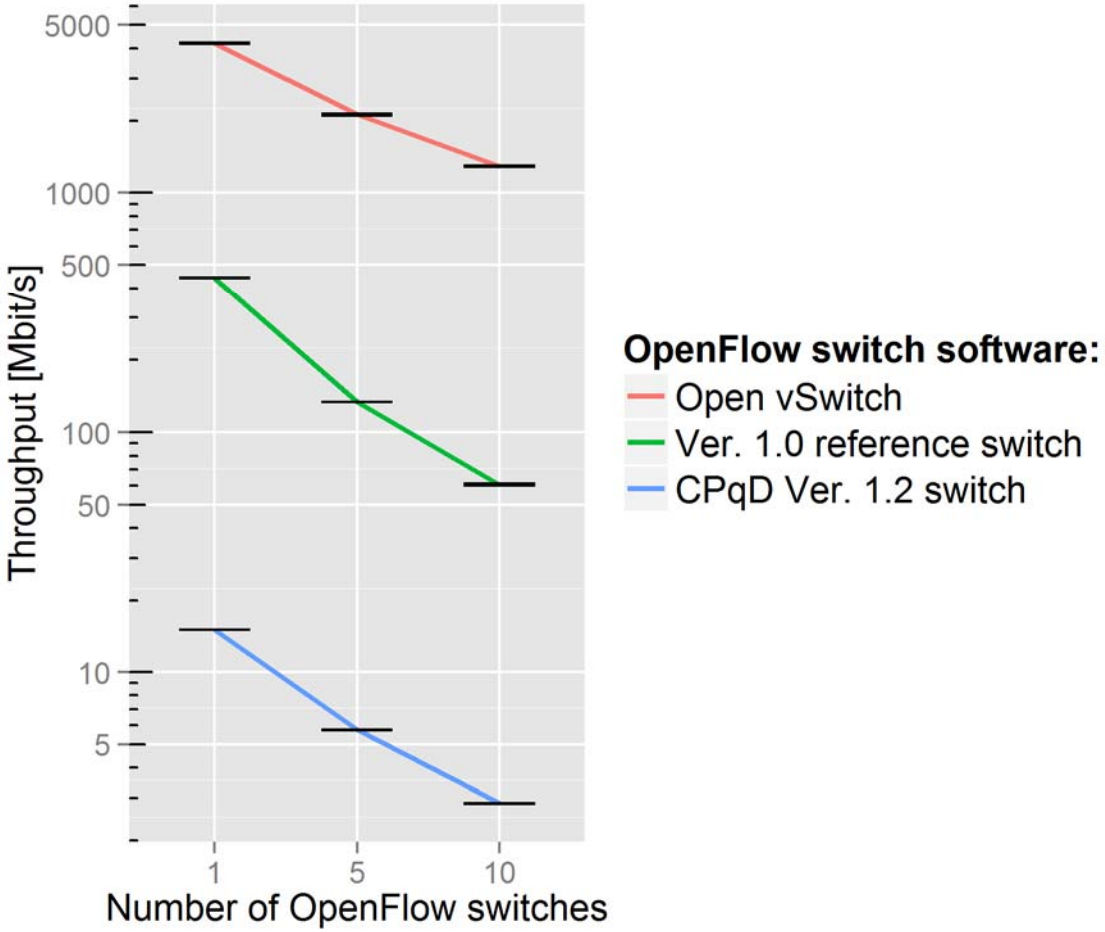
[1]Available from: http://osrg.github.io/ryu/ [Accessed 21.05.2013]

Figure source: [Bl13]

OpenFlow Software Switch Performance

TECHNISCHE UNIVERSITÄT DARMSTADT

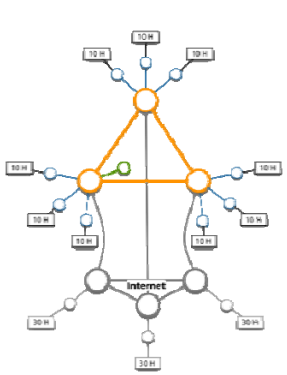
- Requirements
  - Mininet integration
  - ≥ OpenFlow 1.0
  - High throughput
- Results
  - Use Open vSwitch
  - Results consistent with [LHM2010]
  - OpenFlow 1.2 is not an option
  - Use OpenFlow 1.0



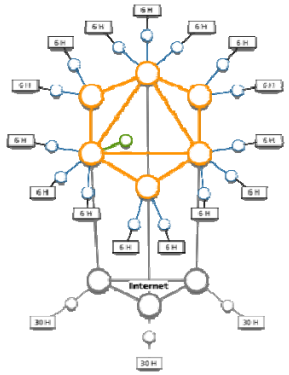
Quantitative Evaluation

TECHNISCHE UNIVERSITÄT DARMSTADT

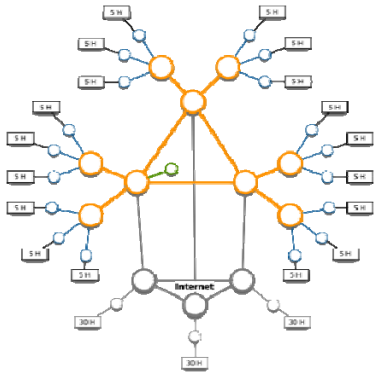
- ISP and Overlay Networks
  - Three different ISP topology variants
  - Three types of overlay topologies
  - 30 repetitions per ISP and overlay



DT Hosts

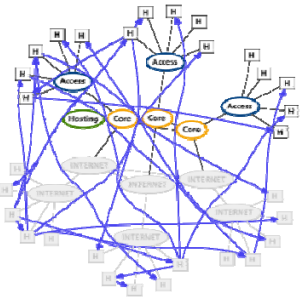


DT Core



DT Depths

- Parameters
  - 180 peers: 90 inside, 90 outside the ISP
  - Video bitrate: 150 kbps, length: 60s



Random

Underlay-aware

SDM

Figure source: [Bl13]

Evaluation Setup

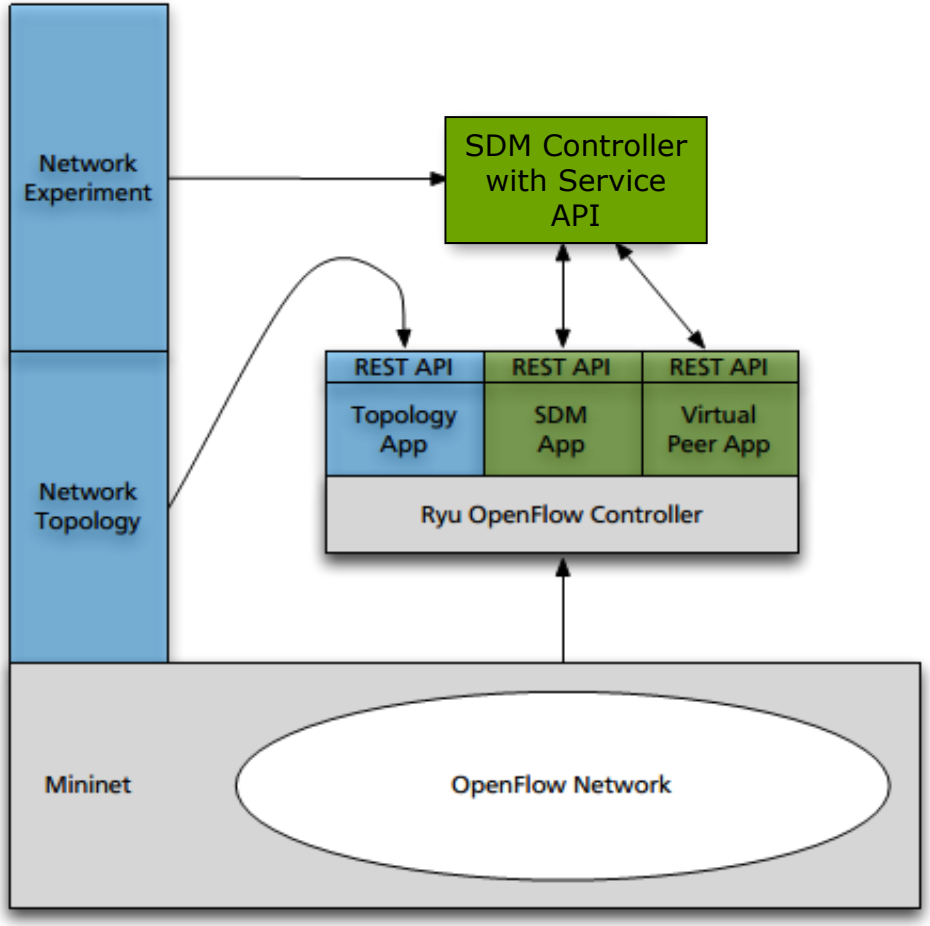


Figure source: [Bl13]

Evaluation Results
Intra-ISP Traffic Volume

TECHNISCHE UNIVERSITÄT DARMSTADT

- ISP topology shows clear influence on volumes
- Order of overlay variants the same for ISP topologies
- SDM traffic is at level of IP multicast

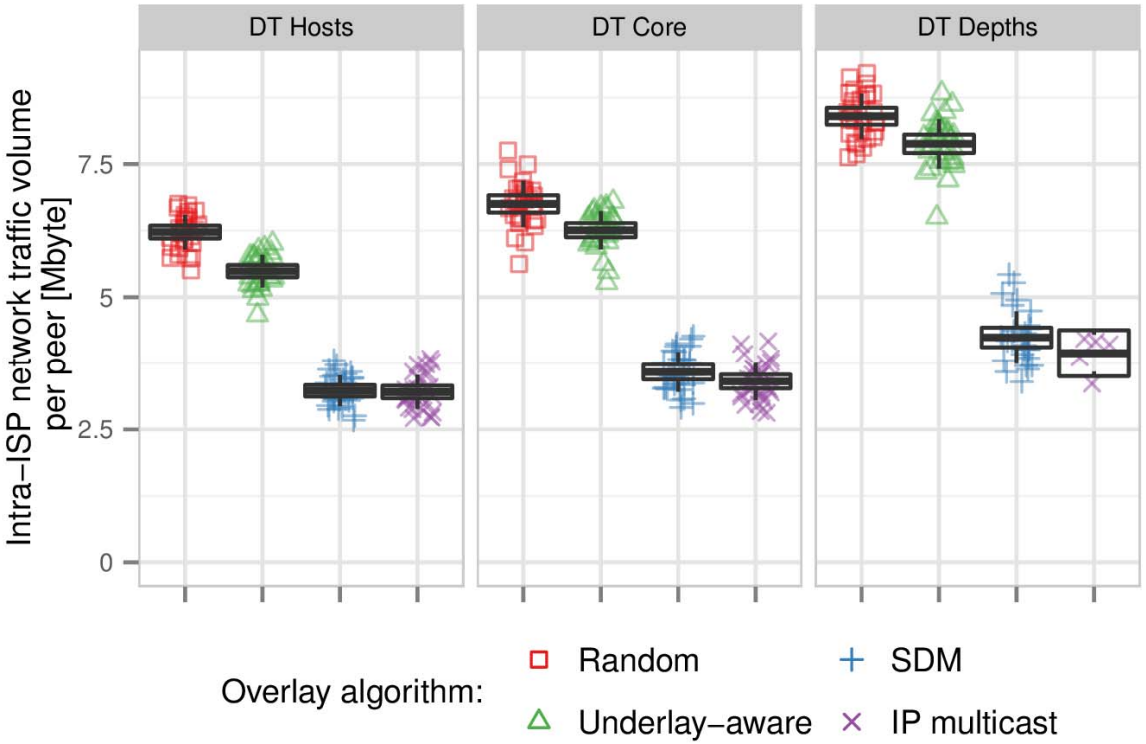


Figure source: [RBH13b]

Evaluation Results
Intra-ISP Normalized Volume

TECHNISCHE UNIVERSITÄT DARMSTADT

- SDM and IP multicast over 30% "better" than underlay-aware
- ISP network topology has no significant influence on relative difference between mechanisms

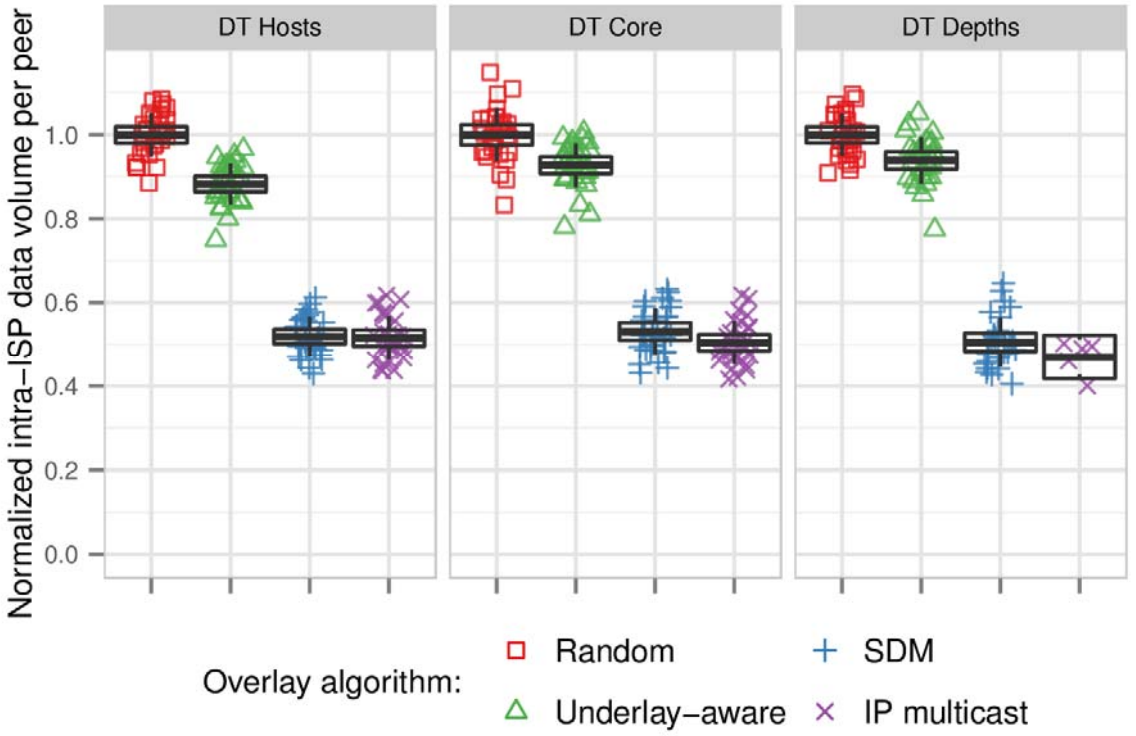


Figure source: [RBH13b]

Evaluation Results
Intra-ISP Traffic Volume

TECHNISCHE UNIVERSITÄT DARMSTADT

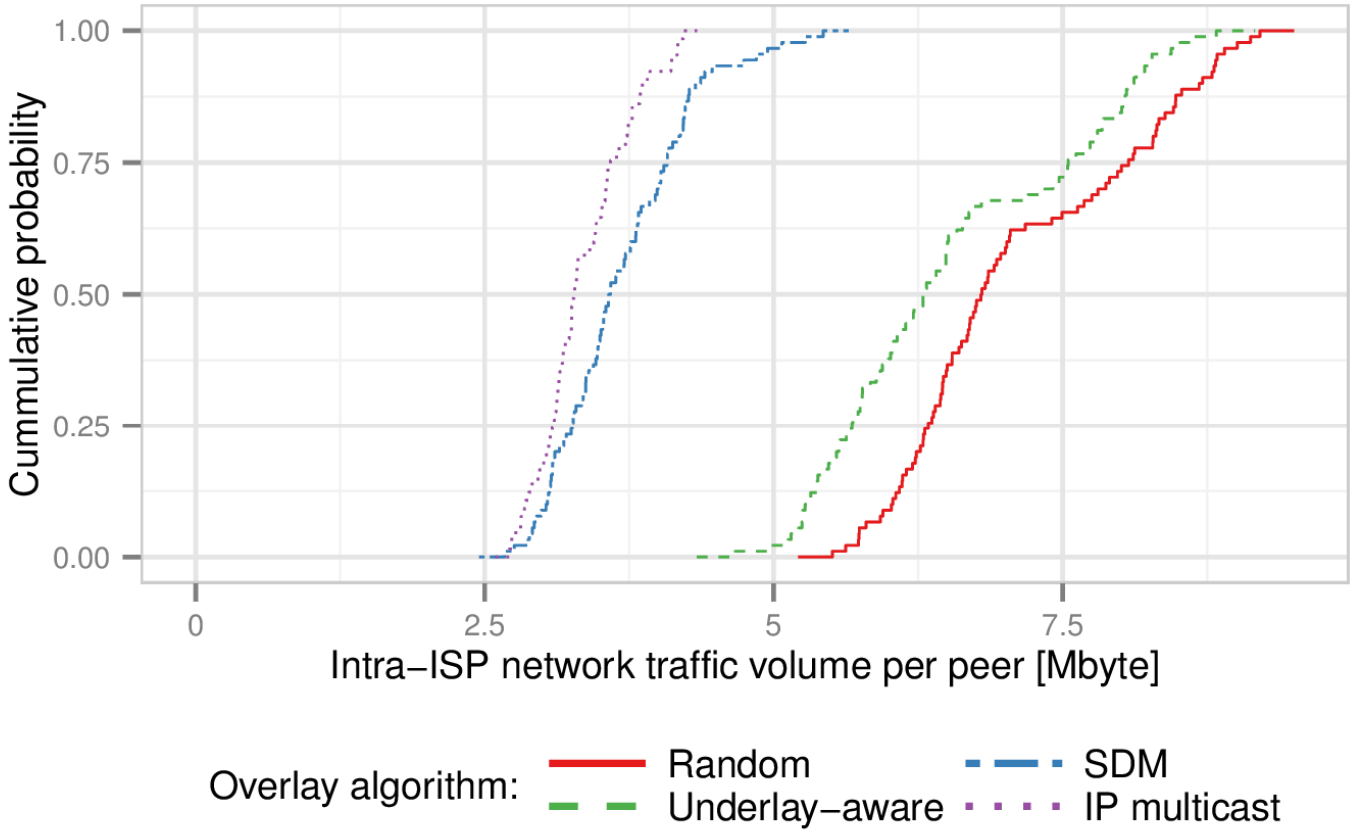


Figure source: [RBH13b]

Evaluation Results
Cross-border Traffic Volume

TECHNISCHE UNIVERSITÄT DARMSTADT

- No significant influence of ISP topology
- SDM traffic is at level of IP multicast and below underlay-aware for DT Hosts
- Low number of valid measurements for IP multicast in case of DT Depths

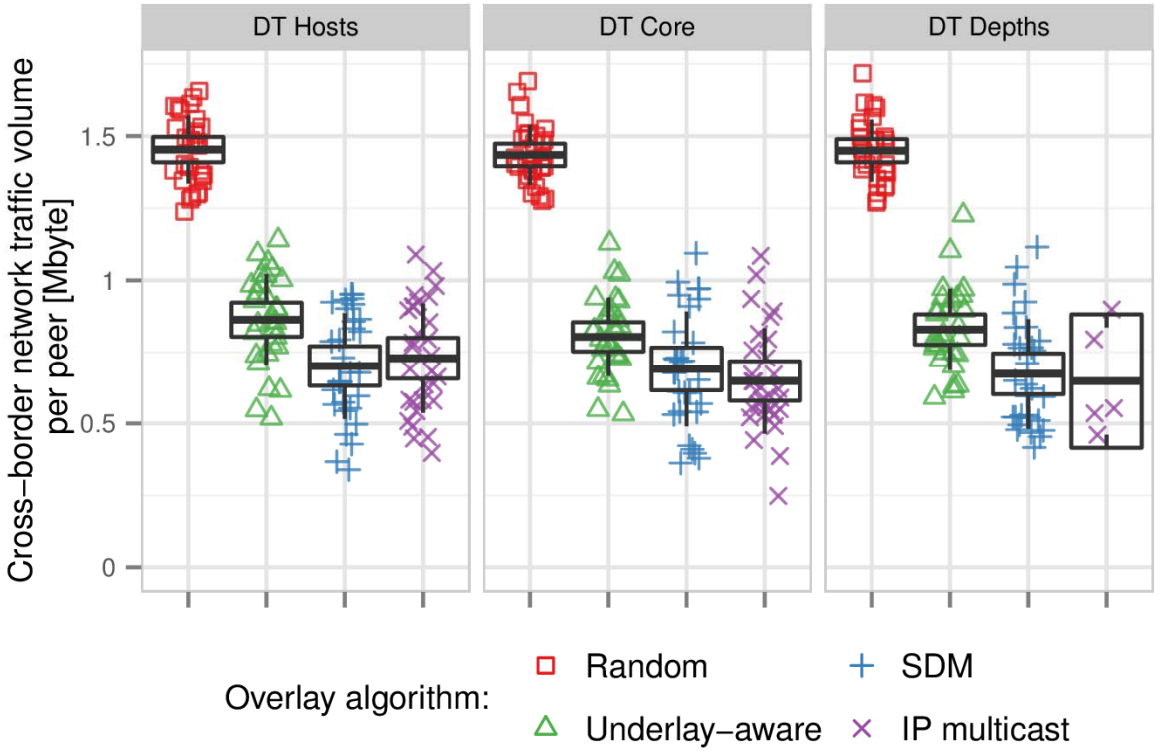


Figure source: [RBH13b]

Evaluation Results: Link Stretch of Peers inside the ISP Network

TECHNISCHE UNIVERSITÄT DARMSTADT

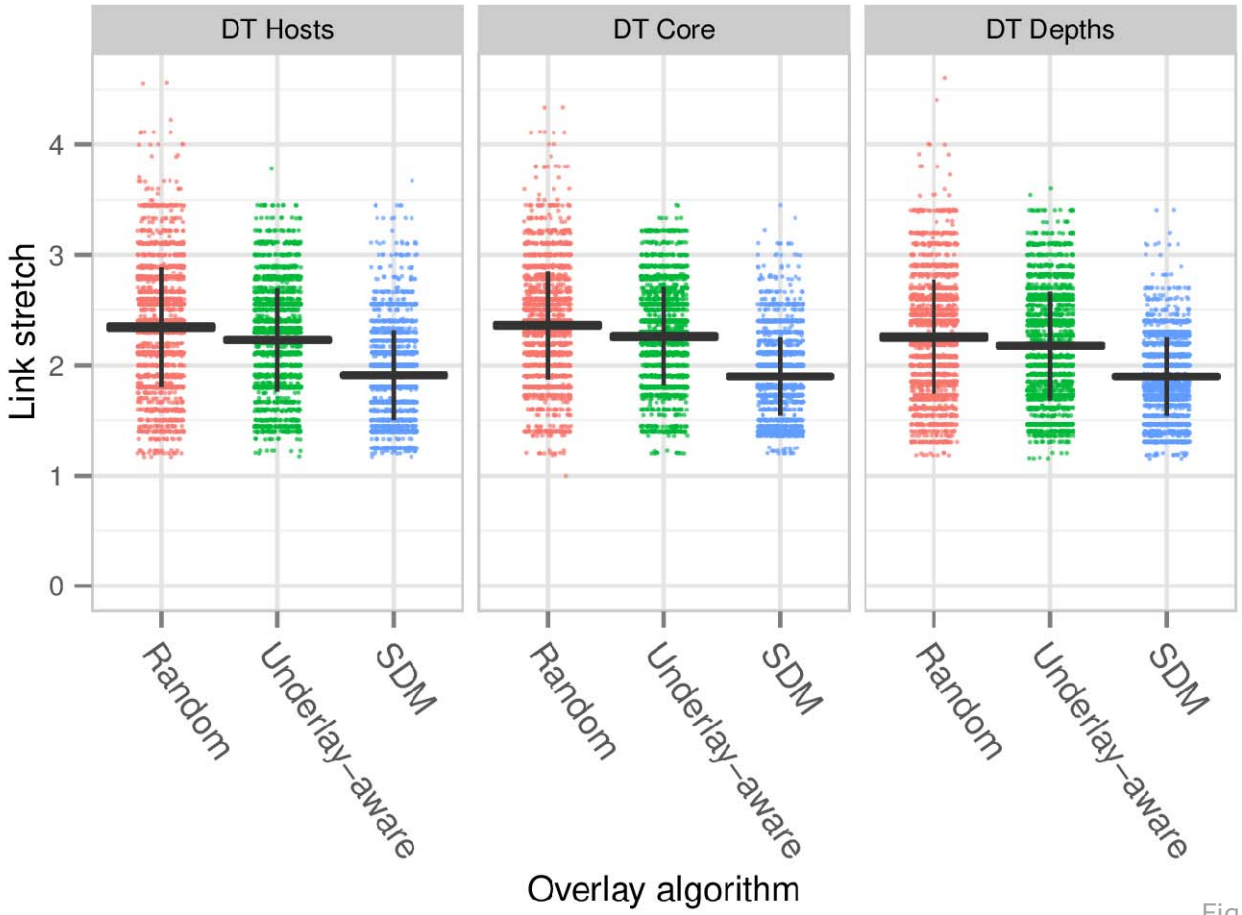
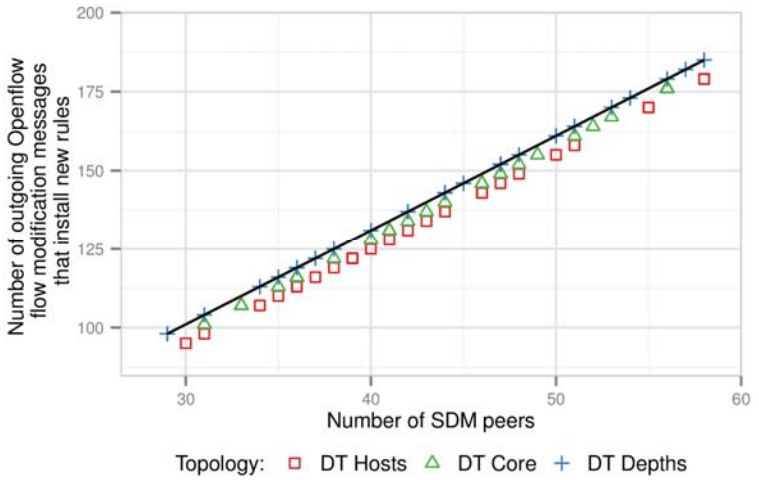


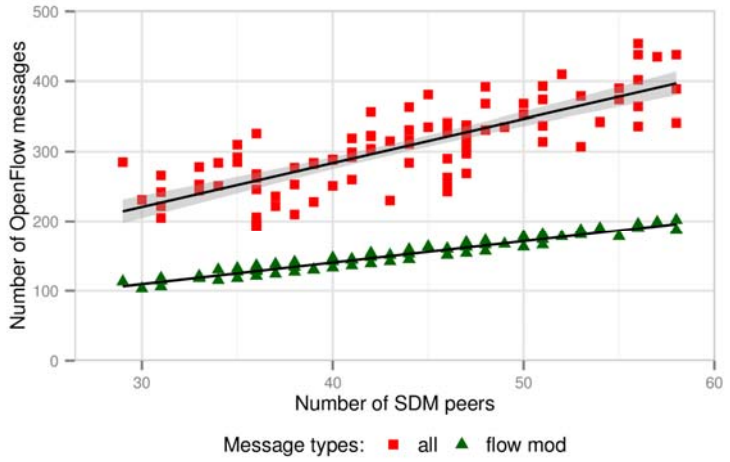
Figure source: [RBH13b]

Evaluation Results: Costs

TECHNISCHE UNIVERSITÄT DARMSTADT

- OpenFlow Rules
  - OpenFlow is rule based
  - Number of supported rules is main resource limitation of devices
  - In average 3 new rules per peer
  - Upper bound given by the longest path from SDM instance switch to any egress switch
- OpenFlow Messages
  - Number of flow mod messages similar to number of rules
  - Additional PacketIn and PacketOut messages due to reactive nature of virtual peer component
  - In average 3.5 additional messages





Picture source: [RBH13b]

References

- [Bl13] J. Blendin: Cross-layer Optimization of Peer-to-Peer Video Streaming in OpenFlow-based ISP Networks. Diploma Thesis, Technische Universität Darmstadt, Supervisor: J. Rückert, 2013.
- [LHM10] B. Lantz, B. Heller, and N. McKeown: A Network in a Laptop. In: ACM Workshop on Hot Topics in Networks (HotNets), 2010.
- [MSG+12] C. Marcondes, T. P. Santos, A. P. Godoy, C.C. Viel, C. A. Teixeira: CastFlow: Clean-slate Multicast Approach using in-advance Path Processing in Programmable Networks. In: IEEE Symposium Computers and Communications (ISCC), 2012.
- [PSS09] I. Papafili, S. Soursos, G. D. Stamoulis: Improvement of BitTorrent Performance and Inter-domain Traffic by Inserting ISP-Owned Peers. In: Network Economics for Next Generation Networks, 5539(10), pp. 97–108, Springer, 2009.
- [PGP+10] J. Pettit, J. Gross, B. Pfaff, M. Casado: Virtual Switching in an Era of Advanced Edges. In: Workshop on Data Center - Converged and Virtual Ethernet Switching (DC-CAVES), 2010.
- [RBH13b] J. Rückert, J. Blendin, D. Hausheer: Software-Defined Multicast for Over-the-Top and Overlay-based Live Streaming in ISP Networks. In: JNSM Special Issue on Management of Software Defined Networks, 2014. http://dx.doi.org/10.1007/s10922-014-9322-8
- [SNZ00] I. Stoica, T.S.E. Ng, H. Zhang: REUNITE: a recursive unicast approach to multicast. In: IEEE INFOCOM, 2000.
- [YGM03] B. Yang, H. Garcia-Molina: Designing a Super-Peer Network. In: IEEE International Conference on Data Engineering, 2003.