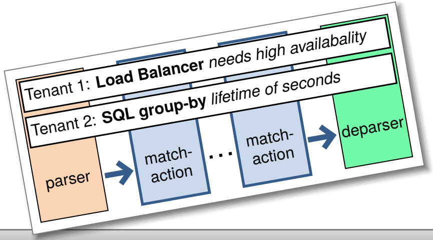


Tenant 1: **Load Balancer** *needs high availabality*

Tenant 2: **SQL group-by** *lifetime of seconds*

parser → match-action … match-action → deparser

Online Reprogrammable Multi Tenant Switches

Johannes Krude[1], Jaco Hofmann[2], Matthias Eichholz[2], Klaus Wehrle[1], Andreas Koch[2], Mira Mezini[2]

[1] RWTH Aachen University, [2] Technische Universität Darmstadt

DFG Collaborative Research Centre 1053 – MAKI
Multi Mechanism Adaptation for the Future Internet

https://comsys.rwth-aachen.de/

ENCP '19, 2019-12-09

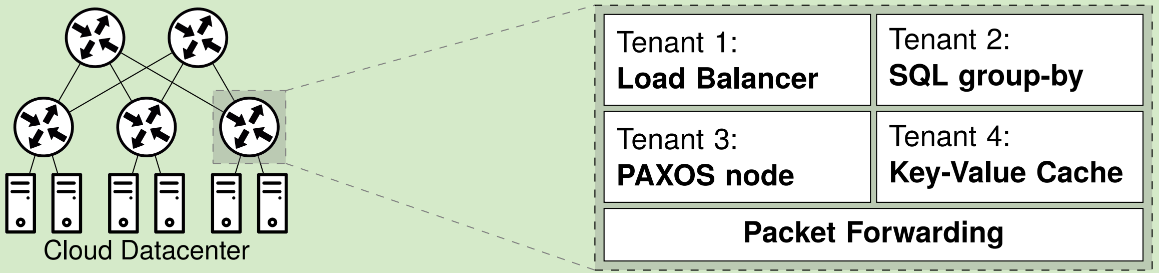
Programmable Switch as a Service

- **On switch …**
  - …stateful load balancer **replaces hundreds of servers** [SilkRoad 2017]
  - …data aggregation **speeds up databases** [Lerner et.al. 2019, …]
  - …paxos **reduces coordination overhead** [NetChain 2018, …]
  - …key-value caching **improves throughput and latency** [NetCache 2017, …]

Programmable Switch as a Service

- **On switch …**
  - …stateful load balancer **replaces hundreds of servers** [SilkRoad 2017]
  - …data aggregation **speeds up databases** [Lerner et.al. 2019, …]
  - …paxos **reduces coordination overhead** [NetChain 2018, …]
  - …key-value caching **improves throughput and latency** [NetCache 2017, …]

Programmable Switch as a Service



Tenant 1: **Load Balancer**

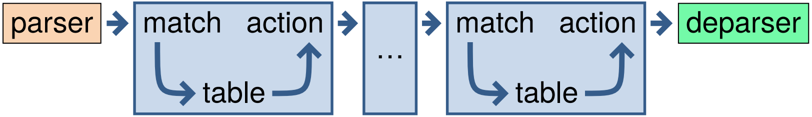
Tenant 2: **SQL group-by**

Tenant 3: **PAXOS node**

Tenant 4: **Key-Value Cache**

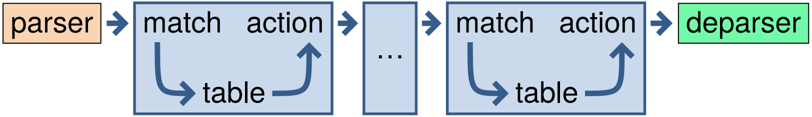
Packet Forwarding

Current Programmable Switch Pipeline [RMT 2013, Barefoot Tofino]



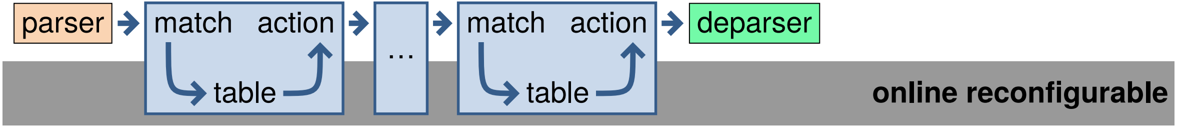
parser → match action table → … → match action table → deparser

Current Programmable Switch Pipeline [RMT 2013, Barefoot Tofino]



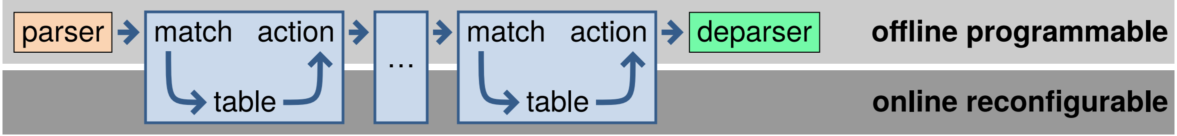
- **Runs a single P4 program**

Current Programmable Switch Pipeline [RMT 2013, Barefoot Tofino]



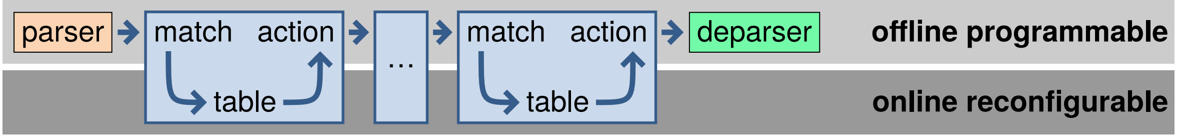
- **Runs a single P4 program**

Current Programmable Switch Pipeline [RMT 2013, Barefoot Tofino]



- Runs a single P4 program
- Reprogramming causes switch and network downtime

Current Programmable Switch Pipeline [RMT 2013, Barefoot Tofino]



- **Runs a single P4 program**
- **Reprogramming causes switch and network downtime**

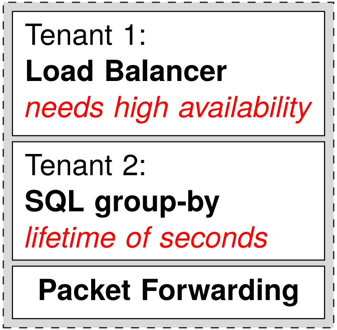
We propose to modify the programmable switch architecture

- To enable hot-pluggability of on-switch functions

Hot-Pluggability

Definition

The ability to insert, modify, and remove on-switch functions without affecting other on-switch functions and packet forwarding.



Programmable Switch

Hot-Pluggability

Definition

The ability to insert, modify, and remove on-switch functions without affecting other on-switch functions and packet forwarding.

Related Work

- **Use a dedicated switch for each application** [PPS 2019]
- **Put generalized functionality permanently onto switches** [NetAccel 2019, Ports et al. 2019]
- **Emulate P4 in Match-Action Tables** [Hyper4 2016, HyperVDP 2019]
  - ▶ Excessive Resource Consumption

Hot-Pluggability

Definition

The ability to insert, modify, and remove on-switch functions without affecting other on-switch functions and packet forwarding.

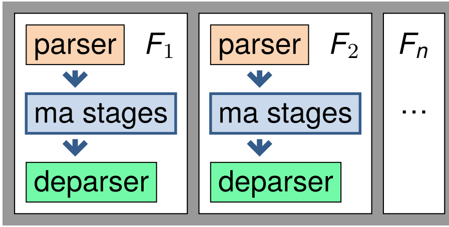
Related Work

- **Use a dedicated switch for each application** [PPS 2019]
- **Put generalized functionality permanently onto switches** [NetAccel 2019, Ports et al. 2019]
- **Emulate P4 in Match-Action Tables** [Hyper4 2016, HyperVDP 2019]
  - ► Excessive Resource Consumption

We want: **Switch Sharing** & **On-Demand Instantiation** & **Individual Customization**

Switch Architecture Requirements

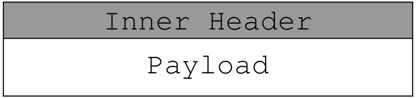
Switch Architecture Requirements

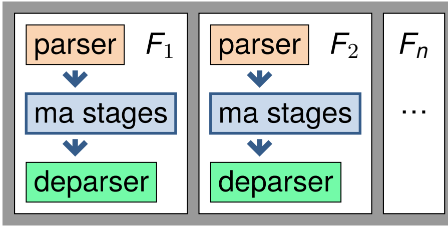


- **Multiple functions**
  - Hot-pluggable

Switch Architecture Requirements

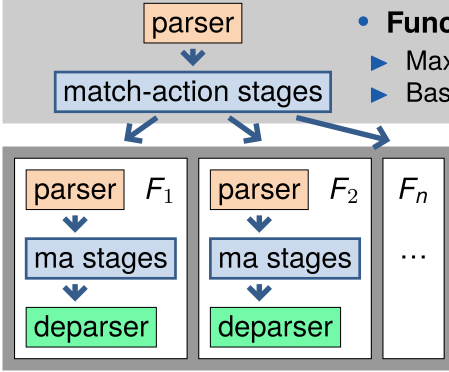
Packet Structure





- **Multiple functions**
  - Hot-pluggable
  - Custom (de)parser

Switch Architecture Requirements

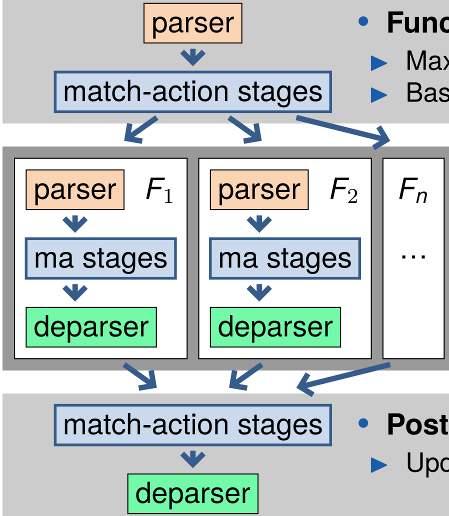


- **Function selection**
  - ▶ Max one function per packet
  - ▶ Based on VXLAN, IP dst, …
- **Multiple functions**
  - ▶ Hot-pluggable
  - ▶ Custom (de)parser

Packet Structure

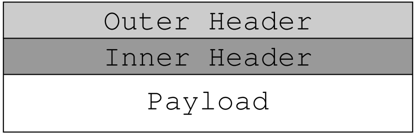
`Outer Header`
`Inner Header`
`Payload`

Switch Architecture Requirements

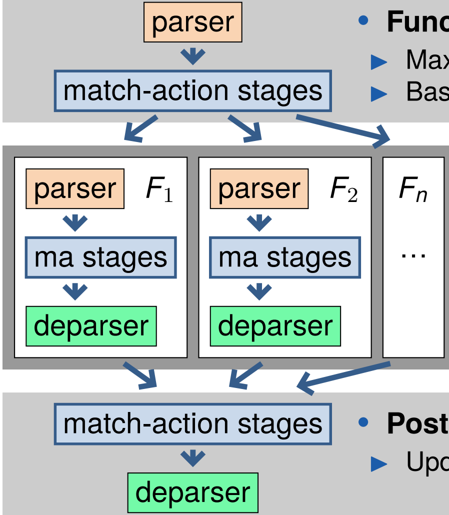


- **Function selection**
  - Max one function per packet
  - Based on VXLAN, IP dst, …
- **Multiple functions**
  - Hot-pluggable
  - Custom (de)parser
- **Post processing**
  - Updating dst/output port

Packet Structure



Switch Architecture Requirements



parser

match-action stages

parser F_1

ma stages

deparser

parser F_2

ma stages

deparser

F_n

…

match-action stages

deparser

- **Function selection**
  - Max one function per packet
  - Based on VXLAN, IP dst, …

- **Multiple functions**
  - Hot-pluggable
  - Custom (de)parser

- **Post processing**
  - Updating dst/output port

Packet Structure

Outer Header

Inner Header

Payload

- **Program isolation**
  - Access to only own packets
  - Limit access to outer header
  - Control plane virtualization

Realizing Online Reprogrammability

- **We present three different possible implementations**
  - None of them yet implemented

Multiple Switching ASICs

✓ Easily realizable

✗ No statefull functions

Using FPGAs

✓ Realizable with FPGA knowledge

✗ Reduced Throughput

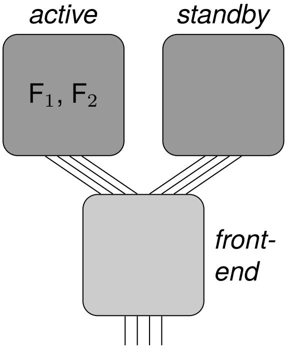
An ASIC extension

✗ To be done by switching ASIC vendors

✓ High performance

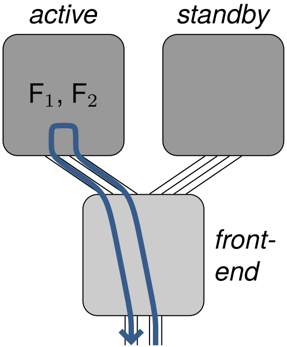
Multiple Programmable Switching ASICs

- **Alternate between two switching ASICs**



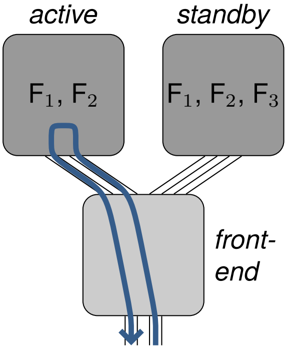
Multiple Programmable Switching ASICs

- Alternate between two switching ASICs



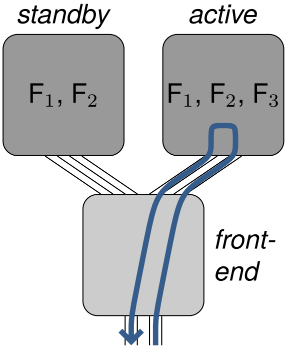
Multiple Programmable Switching ASICs

- **Alternate between two switching ASICs**



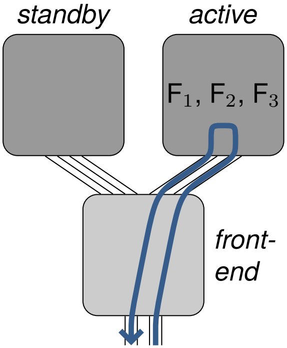
Multiple Programmable Switching ASICs

- Alternate between two switching ASICs



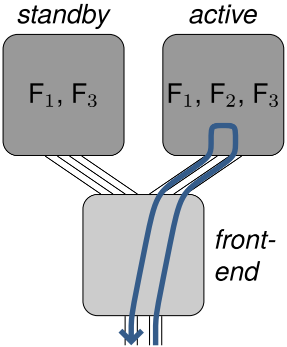
Multiple Programmable Switching ASICs

- **Alternate between two switching ASICs**



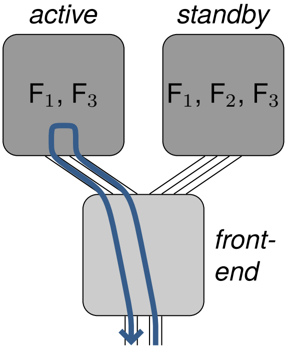
Multiple Programmable Switching ASICs

- **Alternate between two switching ASICs**



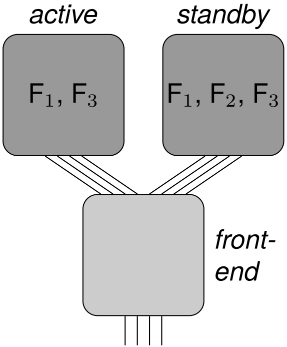
Multiple Programmable Switching ASICs

- **Alternate between two switching ASICs**



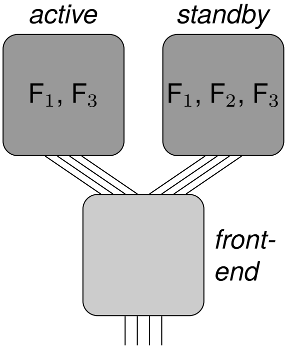
Multiple Programmable Switching ASICs

- Alternate between two switching ASICs
- Merge functions into single program



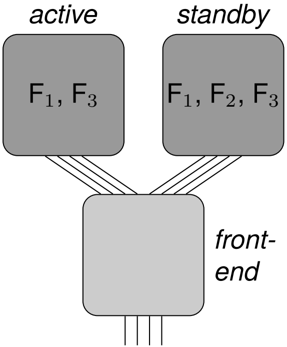
Multiple Programmable Switching ASICs

- **Alternate between two switching ASICs**
- **Merge functions into single program**
- **Compiler provides isolation**
  - ▶ Restricts access to outer headers
  - ▶ Control plane mapping from table memory to function



Multiple Programmable Switching ASICs

- **Alternate between two switching ASICs**
- **Merge functions into single program**
- **Compiler provides isolation**
  - ▶ Restricts access to outer headers
  - ▶ Control plane mapping from table memory to function



Advantages

✓ Based on available hardware

Limitations

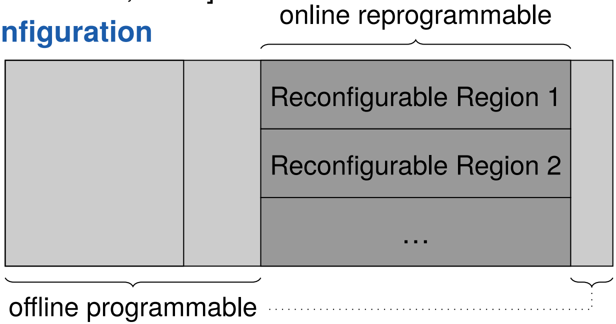
✗ Problematic for statefull functions

Using FPGAs

- **P4 can be executed on FPGAs** [P4→ NetFPGA, 2019]
- **FPGAs support dynamic partial reconfiguration**

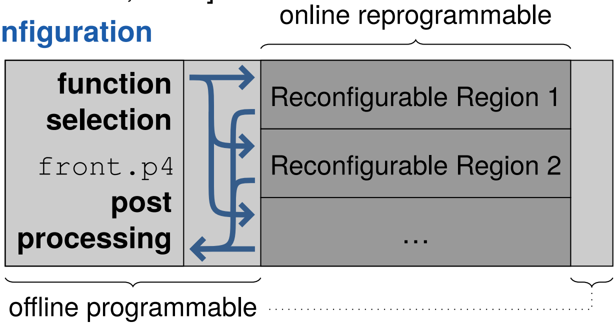
Using FPGAs

- **P4 can be executed on FPGAs** [P4→ NetFPGA, 2019]
- **FPGAs support dynamic partial reconfiguration**



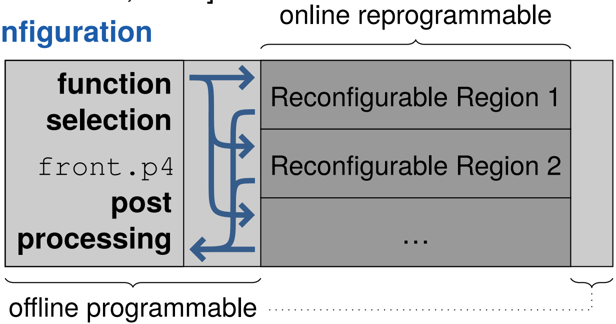
Using FPGAs

- **P4 can be executed on FPGAs** [P4→ NetFPGA, 2019]
- **FPGAs support dynamic partial reconfiguration**



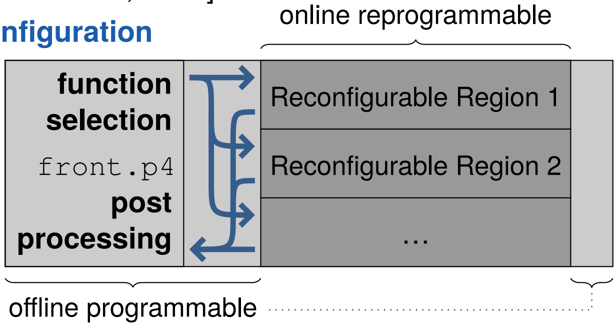
Using FPGAs

- **P4 can be executed on FPGAs** [P4→ NetFPGA, 2019]
- **FPGAs support dynamic partial reconfiguration**
- **Isolation on FPGA level**
  - ▶ Forward each packet to single region
  - ▶ Remove outer headers



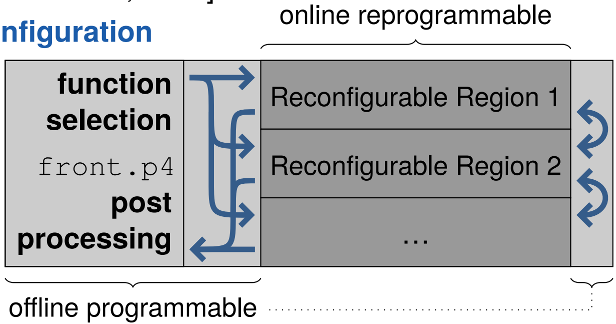
Using FPGAs

- **P4 can be executed on FPGAs** [P4→ NetFPGA, 2019]
- **FPGAs support dynamic partial reconfiguration**
- **Isolation on FPGA level**
  - ▶ Forward each packet to single region
  - ▶ Remove outer headers
- **Fixed sized reconfigurable regions**



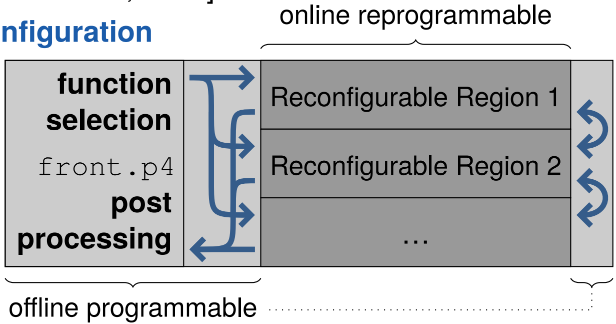
Using FPGAs

- **P4 can be executed on FPGAs** [P4→ NetFPGA, 2019]
- **FPGAs support dynamic partial reconfiguration**
- **Isolation on FPGA level**
  - ▶ Forward each packet to single region
  - ▶ Remove outer headers
- **Fixed sized reconfigurable regions**
  - ▶ Forward packets between regions
  - ▶ Split function across multiple regions



Using FPGAs

- **P4 can be executed on FPGAs** [P4→ NetFPGA, 2019]
- **FPGAs support dynamic partial reconfiguration**
- **Isolation on FPGA level**
  - ▶ Forward each packet to single region
  - ▶ Remove outer headers
- **Fixed sized reconfigurable regions**
  - ▶ Forward packets between regions
  - ▶ Split function across multiple regions



Advantages

✓ Readily available hardware

✓ Non-reconfigured regions keep state

Limitations

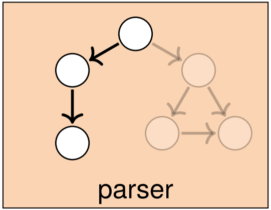
✗ Limited throughput

An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]

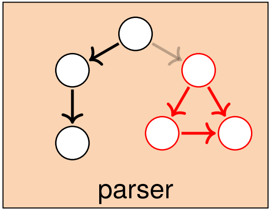
An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**



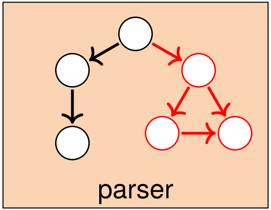
An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**



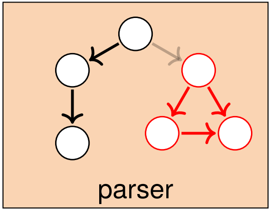
An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**



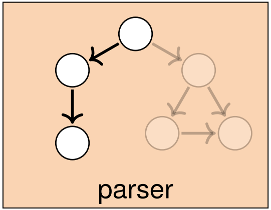
An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**



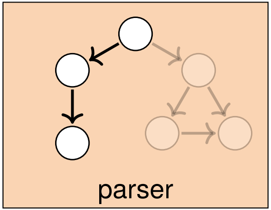
An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**



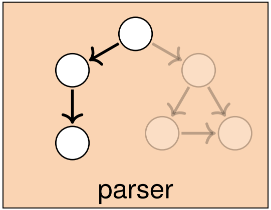
An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**
- **P4 compiler tracks switch occupation**



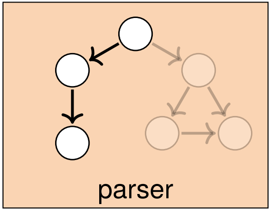
An extension to current switching ASICs

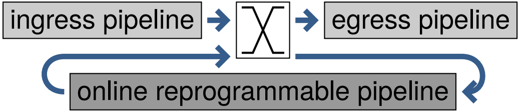
- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**
- **P4 compiler tracks switch occupation**
- **Isolation same as with multiple switching ASICs**



An extension to current switching ASICs

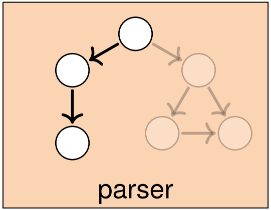
- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**
- **P4 compiler tracks switch occupation**
- **Isolation same as with multiple switching ASICs**
- **Additional Pipeline for custom parser after function selection**
  - Some switches already have an additional pipeline
  - Pipelines can share resources [RMT 2013]

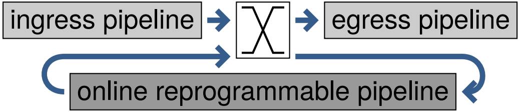




An extension to current switching ASICs

- **Parser, matching, actions, and deparser are stored in SRAM and TCAM**
  - ▶ Use per entry validity bit for atomic updating [CoPTUA 2004]
- **Insertion/Removal order to avoid inconsistent states**
- **P4 compiler tracks switch occupation**
- **Isolation same as with multiple switching ASICs**
- **Additional Pipeline for custom parser after function selection**
  - ▶ Some switches already have an additional pipeline
  - ▶ Pipelines can share resources [RMT 2013]




ingress pipeline

egress pipeline

online reprogrammable pipeline

Limitations

✗ Needs to be done by ASIC vendors

Advantages

✓ Same performace as current ASICs

Conclusion

- **Online reprogrammibility is needed for ”Programmable Switches as a Service“**
- **We propose an architecture for online reprogrammibility**
  - ▶ No implementation yet

Conclusion

- **Online reprogrammibility is needed for ”Programmable Switches as a Service“**
- **We propose an architecture for online reprogrammibility**
  - ▶ No implementation yet

New interesting resource management questions

- Measuring & accounting resource usage
- Resource allocation
- Avoiding resource fragmentation
- …