TOWARDS MORE FLEXIBLE NETWORK INFRASTRUCTURES WITH PROGRAMMABLE DATA PLANES

AUGUST 2022 – JANUARY 2023

SÁNDOR LAKI

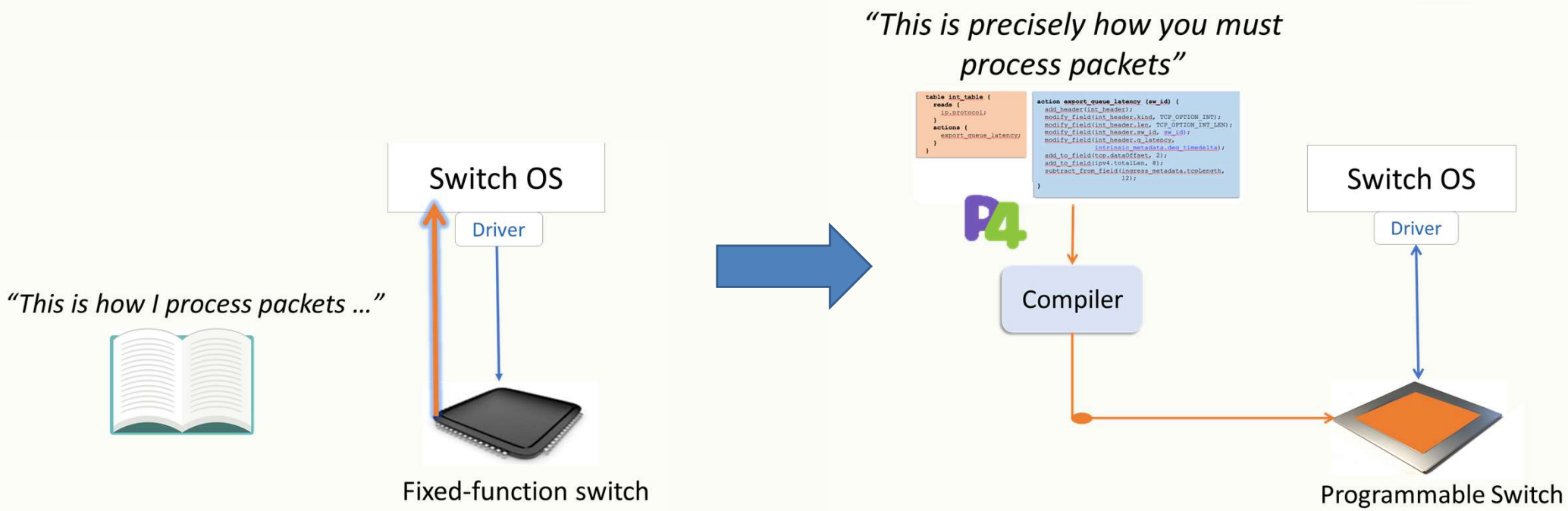
DEPT. OF INFORMATION SYSTEMS
FACULTY OF INFORMATICS
ELTE EÖTVÖS LORÁND UNIVERSITY

EMAIL: LAKIS@INF.ELTE.HU
WEB: HTTP://LAKIS.WEB.ELTE.HU

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE
HUNGARY

PROGRAM FINANCED FROM THE NRDI FUND

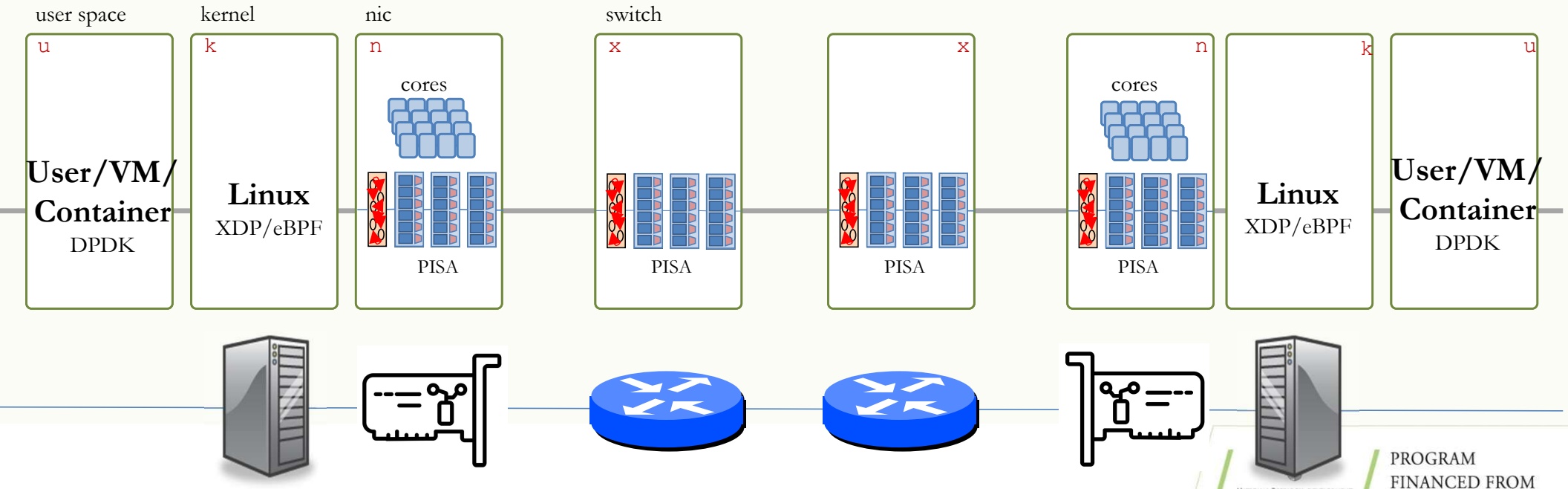
Paradigm shift in networking



National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND

Network as a programmable platform

- Programmable elements (e.g., in P4) at any point of the e2e path



National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND

Where can we use programmable data planes?

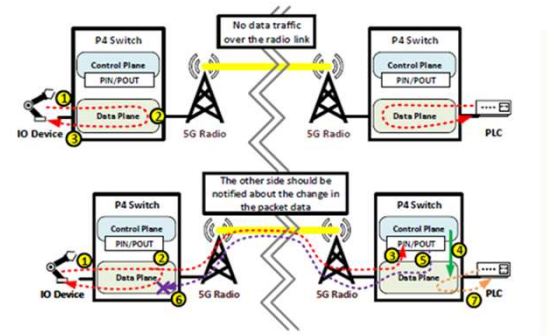
- In-network computing
- Programmable Resource Sharing
- Network programmability

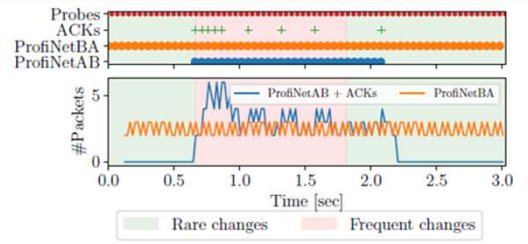
National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND

INC:
In-network computing

1 INC for accelerating Industry 4.0 applications

- **Adaptive traffic reduction on wireless links with in-network caching and filtering**
  - ProfiNet-RT traffic is redundant
  - Traffic reduction on a critical link
  - Loss and delay sensitive
  - Filtering and caching can help in reducing traffic
    - Presented at IEEE NetSoft'21
  - Adaptivity is needed
    - Automatic recognition of traffic patterns
    - Turning traffic filtering on and off accordingly
    - Extended paper is currently under minor revision in IEEE Access (Q1 journal)





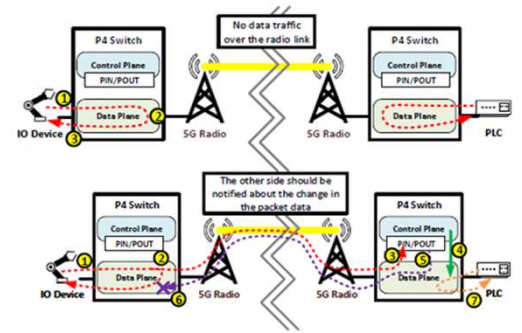
	30 relax 20 change		10 relax 20 change		random intervals		20% prob. of change	
	1 ms	*5 ms*	*1 ms*	*5 ms*	*1 ms*	*5 ms*	*1 ms*	*5 ms*
optimised	84%	88%	140%	147%	139%	144%	40%	70%
adaptive	64%	66%	93%	101%	88%	92%	41%	72%

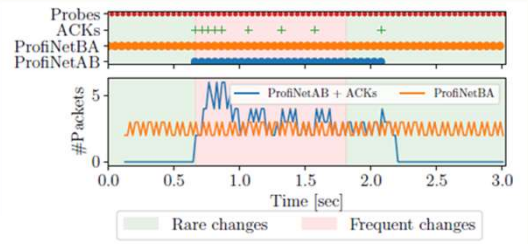
NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROGRAM FINANCED FROM THE NRDI FUND

1 INC for accelerating Industry 4.0 applications

- **Adaptive traffic reduction on wireless links with in-network caching a[illegible] filtering**
  - ProfiNet-RT traffic is redu[illegible]
  - Traffic reduction on a c[illegible]
  - Loss and delay sensitive
  - Filtering and caching can help in [illegible]
    - Presented at IEEE NetSoft'21
  - Adaptivity is needed
    - Automatic recognition of traffic patterns
    - Turning traffic filtering on and off accordingly
    - Extended paper is currently under minor revision in IEEE Access (Q1 journal)

More details in the talk of Péter Vörös



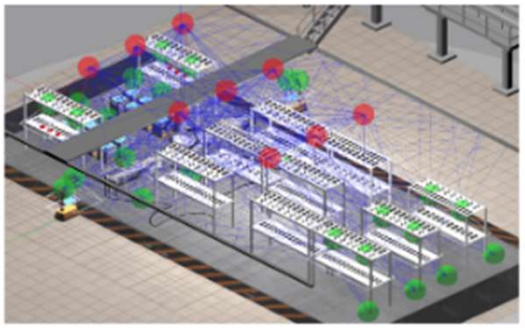


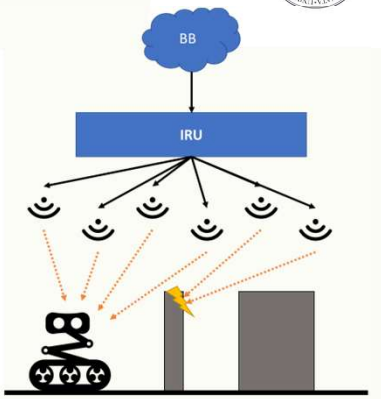
	10 relax 20 change		random intervals		20% prob. of change	
	1 ms	5 ms	1 ms	5 ms	1 ms	5 ms
optimis[illegible]	140%	147%	139%	144%	40%	70%
adaptive	93%	101%	88%	92%	41%	72%

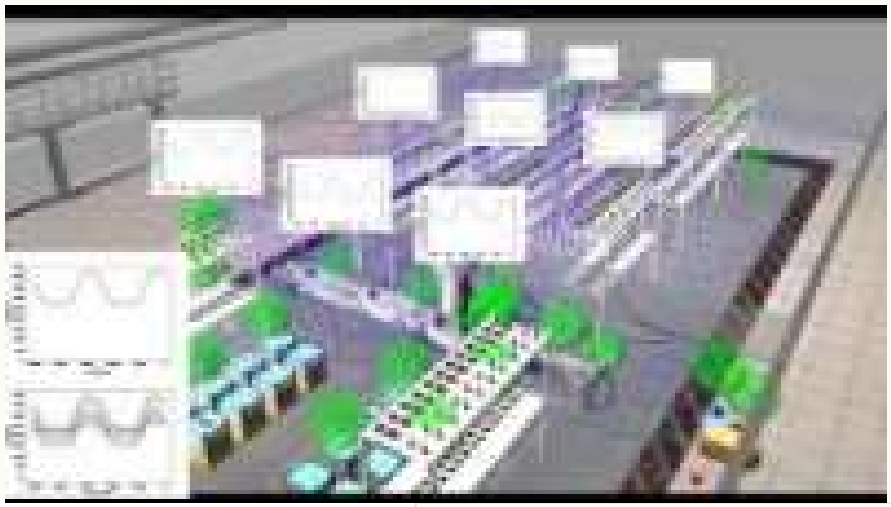
NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROGRAM FINANCED FROM THE NRDI FUND

2 INC for accelerating Industry 4.0 applications

- **Dynamic on/off switching between indoor radio antennas**
  - Cooperative Multi-Point Transmission (CoMP)
    - Forwarding signal through all radio antennas
    - High energy consumption
  - Reducing the number of active radio antennas
    - Save energy
    - While keeping redundancy and high availability
  - Out method incorporates
    - Radio Propagation Digital Twin
      - Real-time emulation of radio signal properties
    - Real-time radio antenna load monitoring in the data plane
    - Pair-wise load balancing data plane algorithm
  - 69% energy saving compared to the traditional CoMP
  - Submitted to IEEE JSAC (D1 journal)



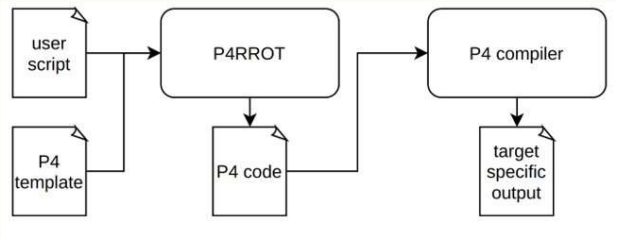




National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND

3a General purpose INC

- **P4RROT: An open-source P4 code generator**
  - Helping the development of INC applications
    - Implenting L7 logic in P4 is difficult
  - Simple and familiar interface implemented in a high-level and well-known language (Python 3)
    - Easy to adopt
  - Modularity, easy to extend.
    - Flexibility.
    - (Possibly provided as a service.)
  - Joint work with Stefan Schmid (TU Berlin)
    - Accepted paper in ACM SIGCOMM CCR (Q2 journal) – with OTKA ack.



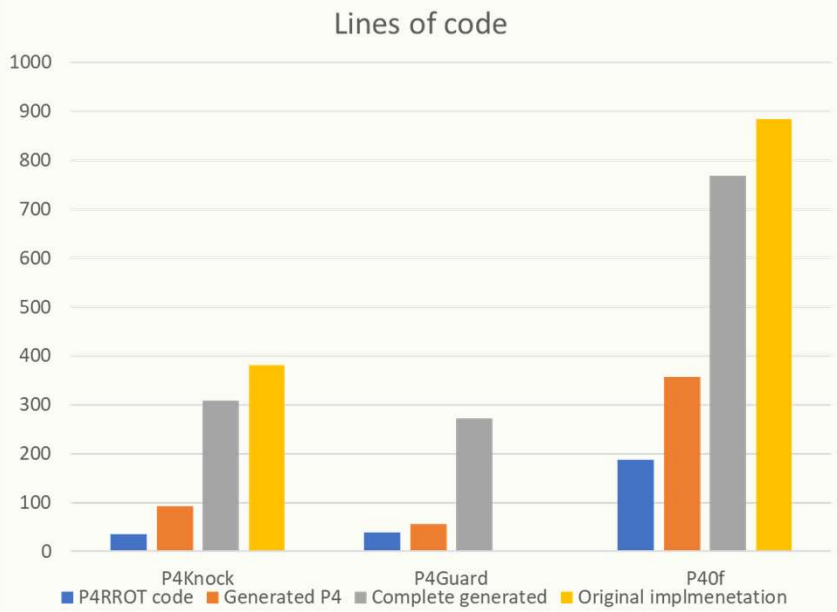

```
(
fp
.add(SwitchTable(['x']))
    .Case([5])
        .add(AssignConst('y',7))
    .Case([9])
        .add(AssignConst('y',5))
    .Default()
        .add(AssignConst('y',9))
.EndSwitch()
)
```

```
...
action case_uid2(){ y = 7; }
action case_uid3(){ y = 5; }
action default_case_uid4(){ y = 9; }

table switch_uid5 {
    key  = { x: exact; }
    actions  = { case_uid2; case_uid3; default_case_uid4; NoAction;}
    const default_action = default_case_uid4;
    const entries = {
        ( 5 ) : case_uid2();
        ( 9 ) : case_uid3();
    }
}

...

apply{
    ...
    switch_uid5.apply()
    ...
}
```


NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROGRAM FINANCED FROM THE NRDI FUND

3b General purpose INC

- **P4RROT for implemening in-network security applications**
  - Three in-network security applications previously implemented in P4
    - PortKnocking
    - Hotfixes
    - OS Fingerprinting
  - Primitives for supporting security applications
    - Support for Tofino HW
  - Reimplementation of existing business logic in P4RROT
    - Functional equivalence?
    - Performance and resource usage
    - But smaller code complexity in terms of lines of code



- Paper submitted to IEEE ICC 2023

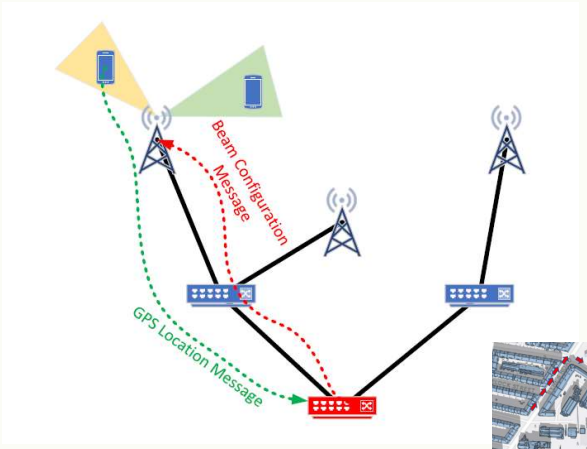
NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROGRAM FINANCED FROM THE NRDI FUND

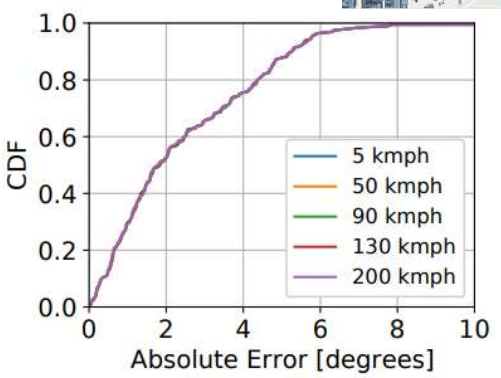
4 INC for supporting beamforming

- **Joint work with Jaspreet Kaur** (Uni. of Glasgow) – EIT-D PhD mobility

- **User-assisted approach**
  - **P4 switch assumption**
    - INC angle computation (or distance, etc.)
    - Generate configuration message to be sent to the BS
  - **BS reconfigures the beam**
    - according to new angle (and other) information
  - **Grid-based approximation**
    - with error bounds
    - Trade-off between TCAM space and accuracy

- Published paper at EUROP4 2022
  - OTKA ack.





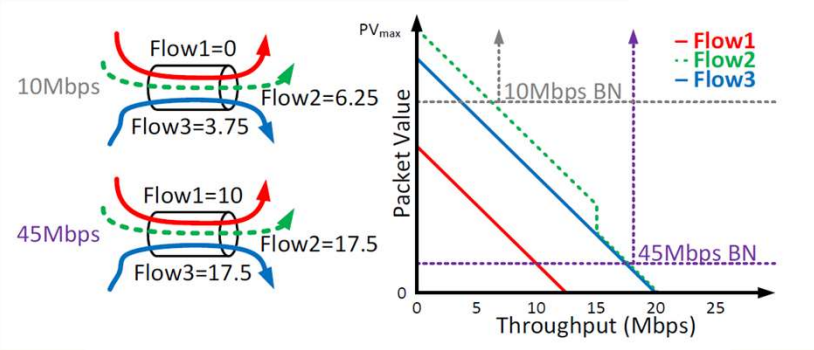
(a) Control Cycle=5ms

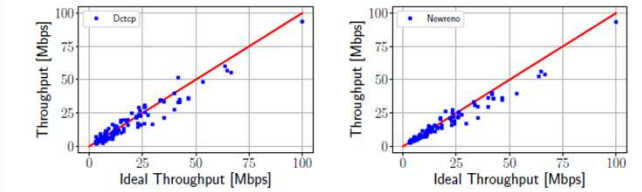
PROGRAM FINANCED FROM THE NRDI FUND

Programmable Resource Sharing

5 Programmable Resource Sharing – old topic…

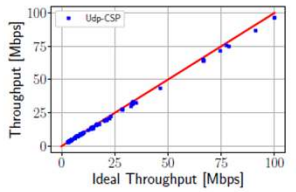
- **Per-Packet Value (PPV) method**
  - End-to-end flows in arbitrary topology
  - Policies given as TVFs
    - TVF ~ inverted bandwidth-function
    - Congestion-level dependent weights
  - Generalized max-min fair allocation
    - Equilibrium exists, provable guarantees
    - Even for different CCA strategies
  - Published paper in IEEE Access (Q1 journal)
  - **BUT, far from Internet economics**
    - Traffic aggregates generated by subscribers
    - Operators sharing the same physical infrastructure
    - Slices, etc.





(a) DCTCP (b) NewReno

(c) Sprint



(d) UDP-CSP

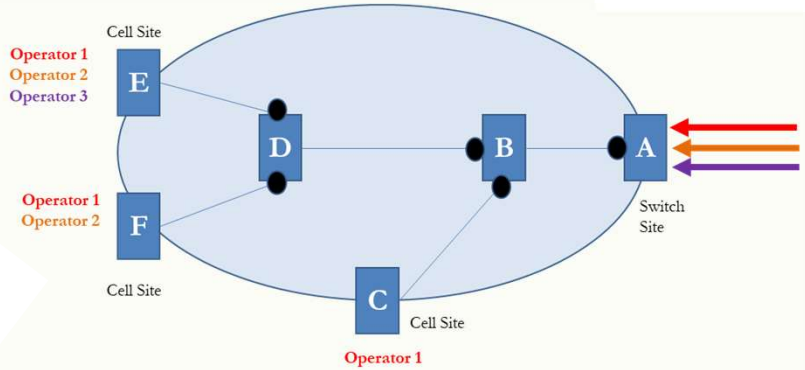
NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

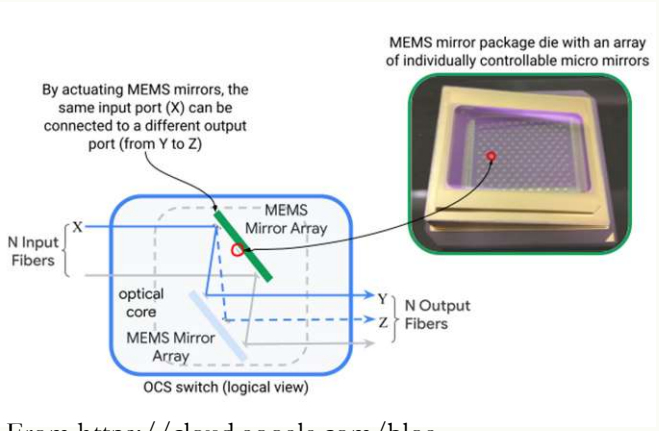
PROGRAM FINANCED FROM THE NRDI FUND

5 Programmable Resource Sharing – new perspectives..

- **Mobile backhaul networks**
  - **Better end-user fairness**
  - **Better fairness between MVNOs**
  - Flexible and dynamic policy control
  - Easier slice management
  - PPV emulation with existing hw
- **Data center networks**
  - **Weighted fairness between services, tenants**
  - **Load balancing between alternative paths**
  - **New DC architectures like Google's Jupiter**
    - Fully optical switching, dynamic connections
    - Links based on traffic matrix
    - Time-scale ~ 10sec
  - 10 sec is too high, load should be handled at smallet time-scale
  - **+ Microbursts**

More details in the talk of Gergő Gombos



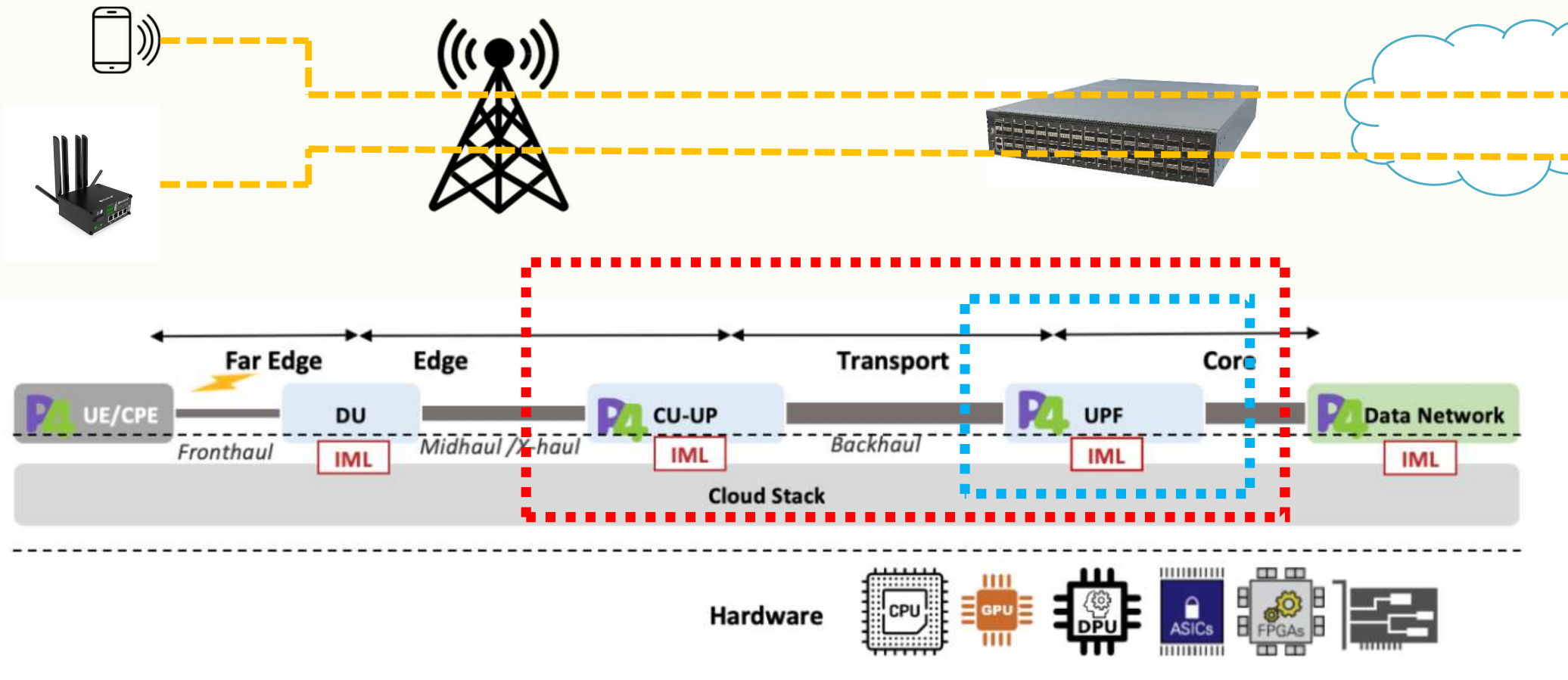


From https://cloud.google.com/blog

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY
PROGRAM FINANCED FROM THE NRDI FUND

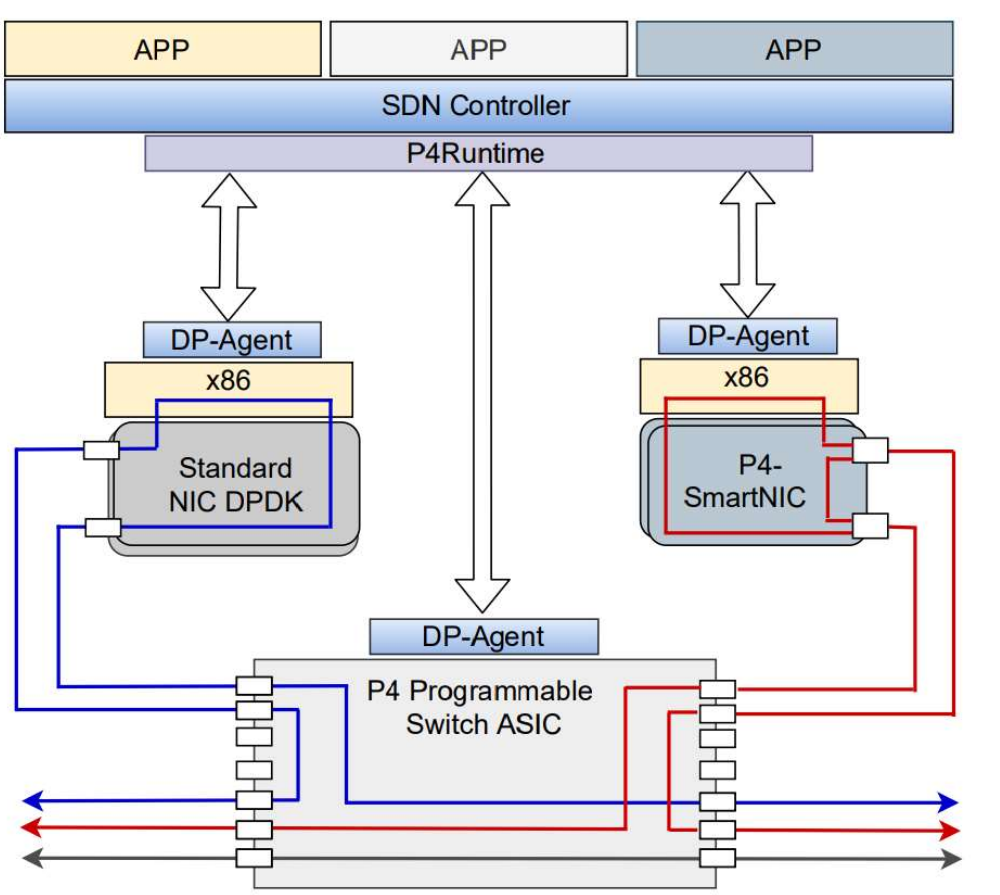
Network Programmability

6 E2E programmability vision



6 UPF design

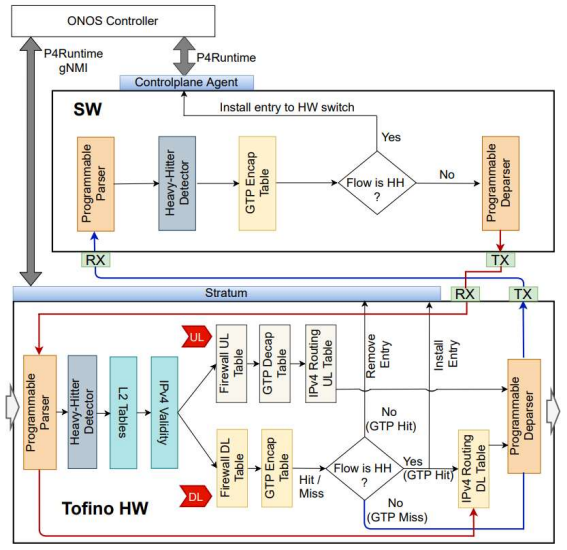
- Key functions
  - L2 switching/virtualization
  - QoS support
  - Firewall
  - GTP decap/encap
  - L3 routing
- Disaggregation of the pipeline
  - Horizontal split
    - Identical logic, but the traffic is split
  - Vertical split
    - Chain of basic functional blocks

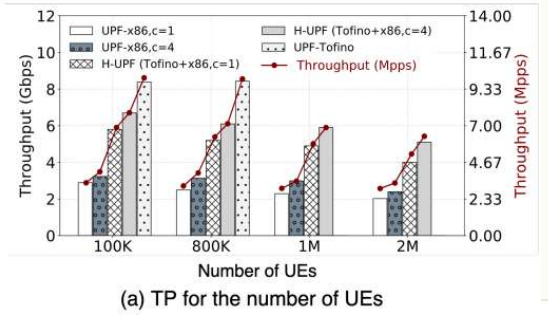


6 An example Hybrid UPF – Intel Tofino + x86

Joint work C. E. Rothenberg and A. Kassler

- Tofino ASIC
  - Guaranteed low and bounded per packet delay
  - >6.5 Tbit/sec forwarding capacity
  - Limited SRAM resources - 1000s of UE matches only
  - Good target for crucial control functions like ACL
- Solutions
  - Option 1 – Scaling out to multiple switches
  - Option 2 – Differentiate between UEs
    - 90-95% of UEs are inactive or non-heavy-hitters
    - Only 5-10% have high throughput demand (heavy-hitters (HH))
    - E.g., 5M UEs: 5-10% smart phones (HH), 10-20% wideband IoT (HH), 70-85% narrowband IoT (non-HH)
    - Deploying HHs on Tofino, while non-HHs on x86
- Published paper in IEEE TMC (D1 journal)





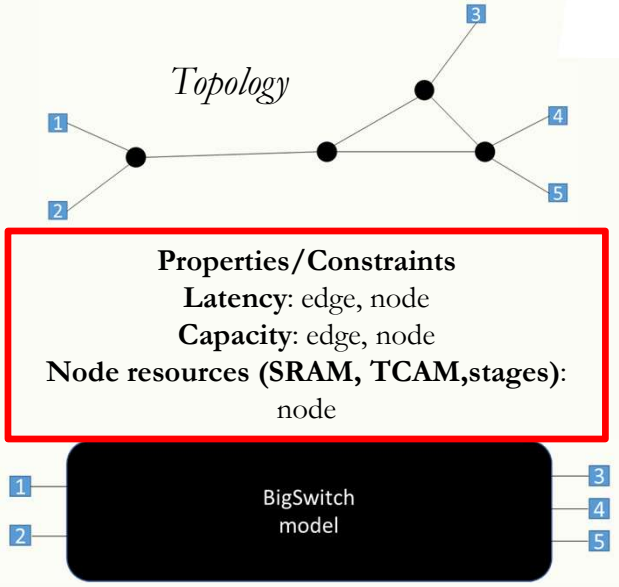
(a) TP for the number of UEs

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE HUNGARY

PROGRAM FINANCED FROM THE NRDI FUND

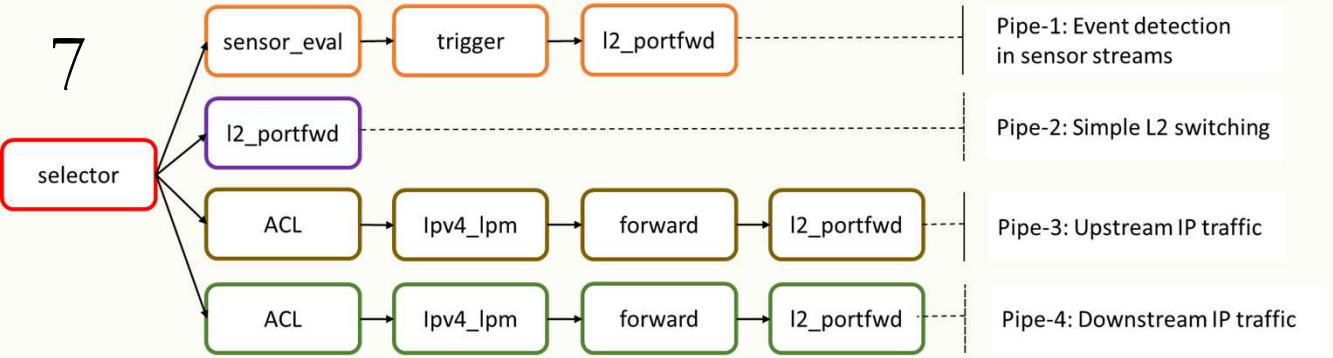
7 Network-wide programmability
with the BigSwitch model

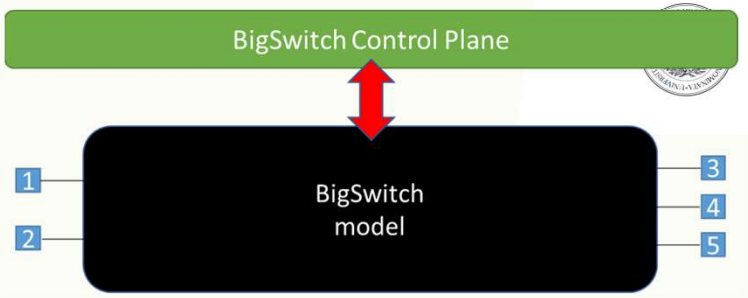
- Givel a network topology
  - With capacity, latency and resource constraints
- Description of network-wide behavior, e.g., in P4
  - Specific architecuture model is needed - BigSwitch
  - Ususally a composition of multiple pipelines
  - Ingress/Egress ports are entry/exit points of the network
- Flow Group
  - relation between a set of entry points and a set of exit points – representing high level ports
  - implements a pipeline of the BigSwitch model
- Flow Group can be split to flows
- Each flow has an exact entry and exit point
- Flow path set can be determined
  - E.g., considering a delay budget, not allowing detours, limit the number of loops in the path or fully denying them



Flow Groups: e.g., f1: p1->p2, f2: p2->p1,
f3: {p1,p2}->{p3,p4,p5},
f4: {p3,p4,p5}->{p1,p2}
Each flow group executes a pipeline
Each pipeline/flow group may have requirements:
latency budget, min. throughput demand, etc.

National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND



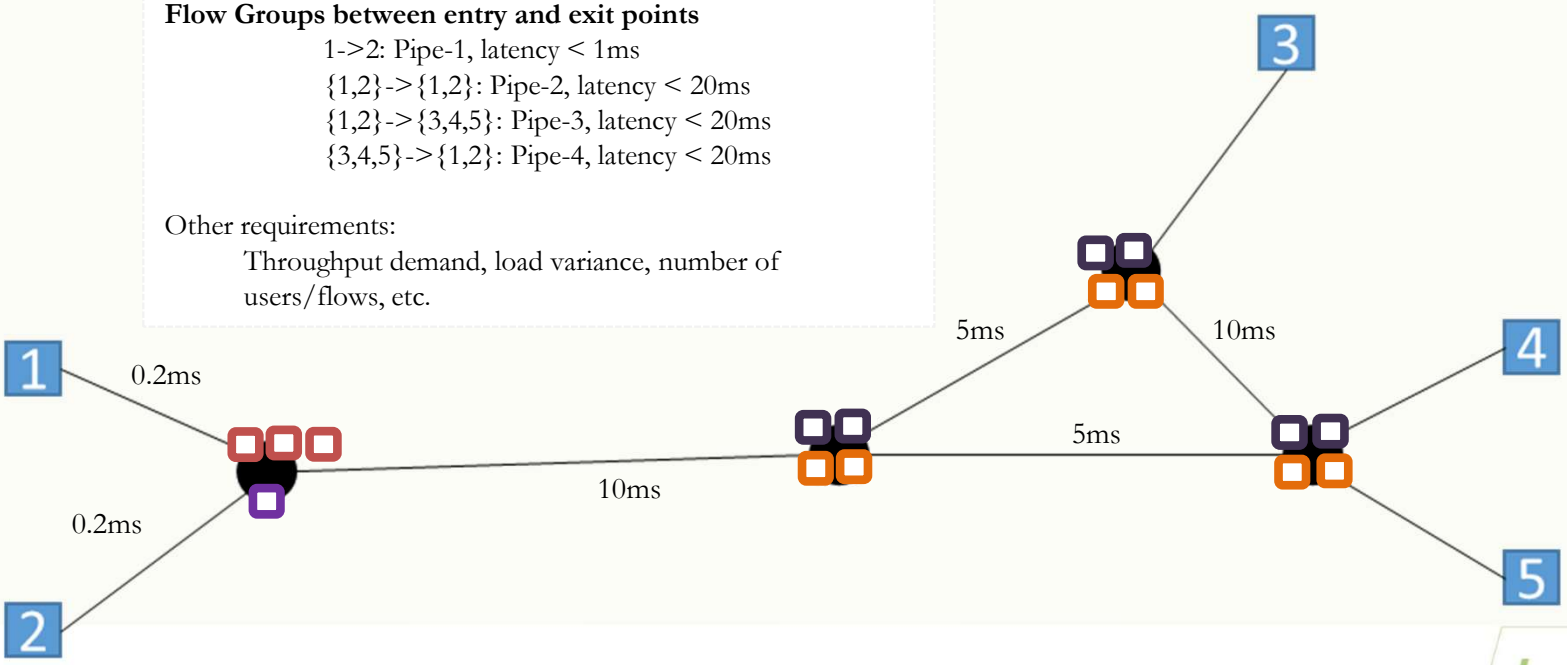


Flow Groups between entry and exit points

1->2: Pipe-1, latency < 1ms
{1,2}->{1,2}: Pipe-2, latency < 20ms
{1,2}->{3,4,5}: Pipe-3, latency < 20ms
{3,4,5}->{1,2}: Pipe-4, latency < 20ms

Other requirements:

Throughput demand, load variance, number of users/flows, etc.



National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND

Publications and other activities
since August 2022

- With TKP ack:
  - G. Gombos, D. Kis, L. Tóthmérész, T. Király, Sz. Nádas, S. Laki: Flow Fairness with Core-Stateless Resource Sharing in Arbitrary Topology [Accepted], IEEE Access journal (Q1, OA, IF: 3.367), vol. 10, pp. 120312-120328, 2022.
  - S. K. Singh, C. E. Rothenberg, J. Langlet, A. Kassler, P. Vörös, S. Laki, G. Pongrácz: Hybrid P4 Programmable Pipelines for 5G gNodeB and User Plane Functions [Accepted], IEEE Transactions on Mobile Computing (IEEE TMC) journal (Q1, IF: 6.07), Volume: tba, Issue: tba, Page(s): 1-18, 2022
  - D. Kis, G. Gombos, S. Laki, Sz. Nádas: Resource Sharing Beyond FQ: 35K Users at 100 Gbps [Accepted], Proceedings of ACM SIGCOMM 2022 (demo paper), 22-26 August, 2022 - Amsterdam, The Netherlands
  - Cs. Györgyi, K. Kecskeméti, P. Vörös, S. Laki: Simplifying the development of in-network security applications with P4RROT, Submitted to IEEE ICC 2023
  - Cs. Györgyi, P. Vörös, J. Pető, G. Szabó, S. Laki: Radio Propagation Digital Twin Aided Multi-Point Transmission with In-network Dynamic On-Off Switching, Submitted to IEEE JSAC
  - Cs. Györgyi, K. Kecskeméti, P. Vörös, G. Szabó, S. Laki: Adaptive Network Traffic Reduction on the Fly with Programmable Data Planes, Submitted to IEEE Access

- Without TKP ack:
  - Cs. Györgyi, S. Laki, S. Schmid: P4RROT: Generating P4 Code for the Application Layer [Accepted], SIGCOMM CCR journal (Q2, IF: 2.322), January 2023.
  - H. Mallouhi, S. Laki, Towards Disaggregated P4 Pipelines with Information Exchange Minimization [Accepted], Proceedings of ACM CoNext 2022 - Student Workshop (Poster paper), 6-9 December, 2022 - Rome, Italy
  - H. Mallouhi, J. Kaur, H. Abbas, S. Laki, In-network Angle Approximation for Supporting Adaptive Beamforming [Accepted], Proceedings of ACM CoNext 2022 - EUROP4 Workshop (Full workshop paper), 6-9 December, 2022 - Rome, Italy

National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND

Publications and other activities
since August 2022

- Others
  - Organization of the 2nd P4Pi Hackathon at ACM SIGCOMM 2022, Amstertam, NL
    - Jointly with Noa Zilberman (Oxford), Robert Soulé (Yale), Fernando Ramos (Uni. Lisbon)
  - Participation in a Dagstuhl Seminar on Network Automation, November 2022
    - Towards More Flexible and Automated Communication Networks
  - Nomination to the P4 Technical Steering Team of ONF's P4 project

Cooperation with the mathematical optimization RG

- Lilla Tothmérész, Tamás Király

- External collaborators
  - Andreas Kassler,Karlstad University (Sweden)
  - Christian E. Rothenberg, University of Campinas (BR)
  - Chrysa Papagianni, University of Amsterdam (NL)
  - Noa Zilberman, University of Oxford (UK)
  - Robert Soulé, Yale University (USA)
  - Koen de Schepper, Nokia Bell Labs Anwerp (BE)
  - Stefan Schmid, TU Berlin (DE)
  - Luis M. Contreras, Telefonica (S)
  - Fernando Ramos, Uni. Lisbon, (PT)
  - Szilveszter Nádas, Gergely Pongrácz, Géza Szabó, Ericsson Research (HU)

- TPC memberships:
  - USENIX ATC 2022-2023 (Rank A), IEEE CCNC 2020-2023 (Rank B), IEEE NetSoft 2021, 2023 (Rank B), IEEE CloudNet 2021-2022, EUROP4 WS 2020-2022

Open Networking Foundation

National Research, Development and Innovation Office Hungary
PROGRAM FINANCED FROM THE NRDI FUND

Special thanks to

My closer colleagues: Gergő Gombos, Péter Vörös, János Szalai-Gindl, Dhulfiqar A. AlWahab and Máté Tejfel

My PhD students: Dániel Varga, Csaba Györgyi, Ferenc Fejes, Dávid Kis, Hiba Mallouhi and Ákos Rudas

Q&A

WEB: [HTTP://LAKIS.WEB.ELTE.HU](http://lakis.web.elte.hu)

Az Alkalmazásiterület-specifikus nagy megbízhatóságú informatikai megoldások című projekt a Nemzeti Kutatási Fejlesztési és Innovációs Alapból biztosított támogatással, a Tématerületi kiválósági program (TKP2020-NKA-06, Nemzeti Kihívások Alprogram) finanszírozásában valósult meg.

NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE
HUNGARY

PROGRAM FINANCED FROM THE NRDI FUND