ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

XXXIII Cycle

Flow scheduling for commodity switches in datacenter networks

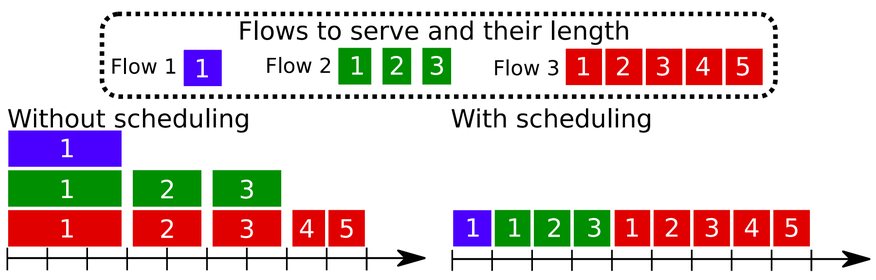
German Sviridov
Supervisors: Prof. Paolo Giaccone
Prof. Andrea Bianco

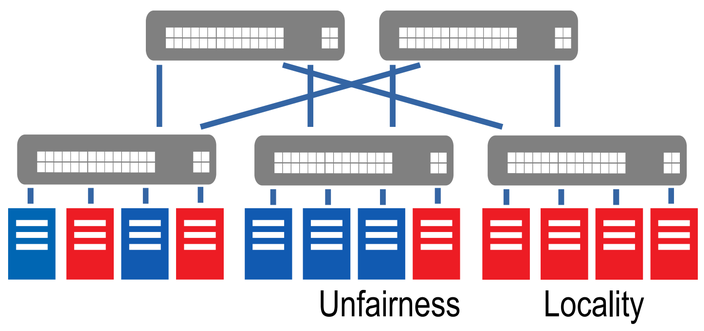
Research context and motivation

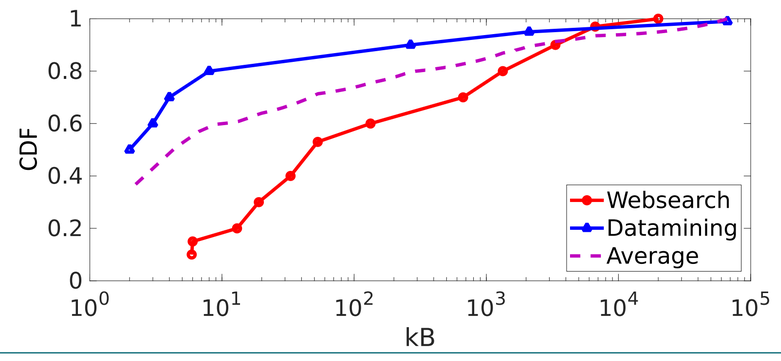
Flow scheduling in datacenter networks:
- Flow completion time (FCT) is used as the primary performance metric in datacenter communication
- Reduction of FCT can be achieved by acting on **Packet / Flow scheduling**

State of the art approaches:
- Require **complex architectures** at switches
- Do not consider **traffic locality** and lead to **unfairness**
- Require knowledge of the **length of individual flows**

In realistic scenarios:
- Simple architectures in switches
- Workload patterns are not uniform in the network
- Only the **flow length distribution** is known





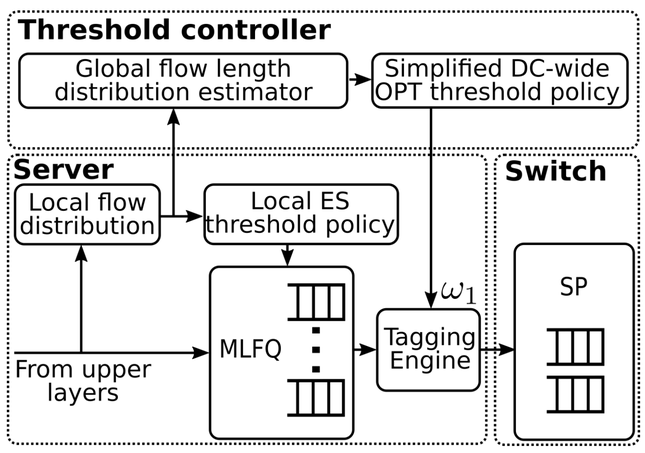


Addressed research questions/problems

- Is it possible to achieve a better trade-off between the performance (delay) and the complexity?
- Key idea: exploit both the local and the global knowledge of the datacenter flow length distribution to perform scheduling.

Novel contributions

- A scheduling mechanisms, namely **Network Optimal Split** with 2 queues (NOS2)
- **Centralized controller**
- Most of complexity moved to servers
- No dedicated hardware
- **Decoupled scheduling**:
  - Simple and fast at servers
  - Optimized but inexpensive at switches

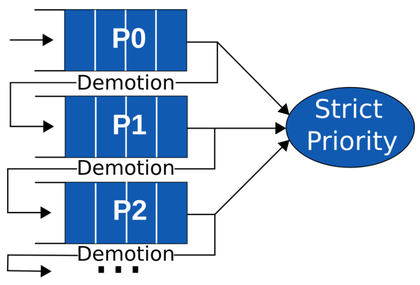


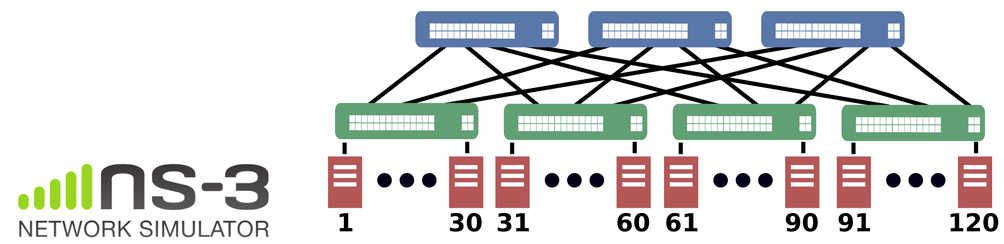
Future work

- Use **of programmable switches** to capture fine-grained flow length distribution locality
- Implementation of NOS2 inside **Linux network stack** and evaluation in a realistic setup
- Server scheduling based on more **advanced knowledge** about the global workload

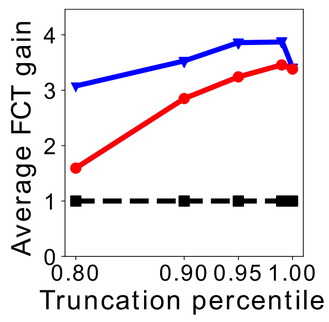
Adopted methodologies

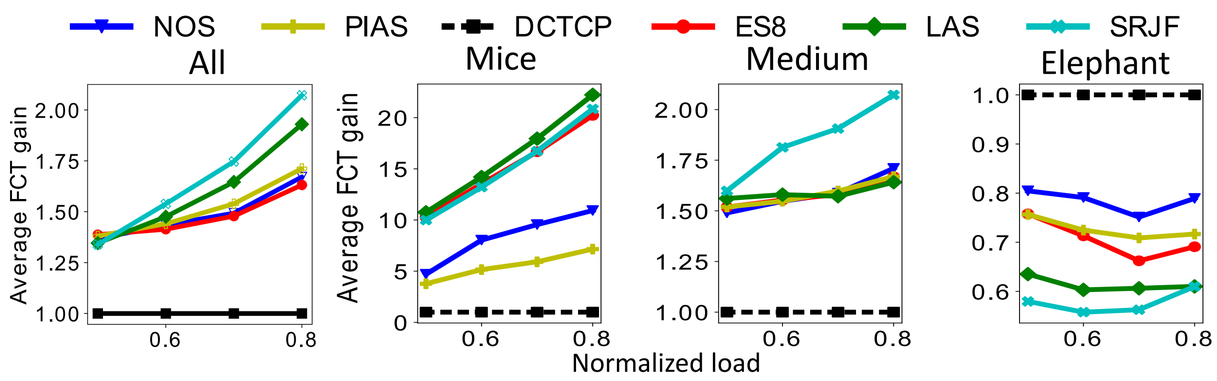
Host scheduling - MLFQ
- Flows are demoted based on the amount of transmitted bytes (thresholds):
  - Short flows finish quickly
  - Long flows compete against each other
- **Demotion thresholds** are assigned using:
  - Equal split (ES-N): Splits the flow length distribution in N equal parts
  - Optimized (OPT): Performs optimization using a Markovian model of the scheduler





Evaluation in NS3
- Leaf and spine topology with 120 servers
- Accurate network representation with realistic protocol stack
- NOS2 achieves:
  - Better resilience to flow length distribution misestimation
  - Performance close to state of the art approaches





Submitted and published works

- German Sviridov, Andrea Bianco, Paolo Giaccone, To sync or not to sync: why asynchronous traffic control is good enough for your data center, IEEE Globecom, Abu Dhabi, UAE, Dec. 2018
- German Sviridov, Marco Bonola, Angelo Tulumello, Paolo Giaccone, Andrea Bianco, Giuseppe Bianchi, LODGE: LOcal Decisions on Global statEs in programmable data planes, IEEE Netsoft, Montreal, Canada, June 2018
- German Sviridov, Andrea Bianco, Paolo Giaccone, Low complexity flow scheduling for commodity switches in DCN, IEEE Globecom, Waikoloa, HI, USA, Dec. 2019
- German Sviridov, Paolo Giaccone, Andrea Bianco, LOcAl DEcisions on Replicated States (LOADER) in programmable data planes: programming abstraction and experimental evaluation, in: IEEE TNSM, Under submission
- Abubakar Sidique Muqaddas, German Sviridov, Paolo Giaccone, Andrea Bianco, Optimal state replication in stateful dataplanes, in: IEEE JSAC, Under submission

List of attended classes

- 01QSAIU – Heuristics and metaheuristics for problem solving: new trends and software tools (13/07/18, 4)
- 01QSCIU – Reconfigurable computing (15/06/18, 4)
- 01QORRV – Writing Scientific Papers in English (21/03/18, 3)
- 01SGURP – Intellectual Property Rights, Technology Transfer and Hi-Tech Entrepreneurship (22/03/18, 6)
- 01TEVRV – Deep learning (04/06/2019, 6)
- 02LWHRV – Data Science for Networks (15/02/2019, 6)
- 01QFFRV – Advanced Techniques for optimization (08/03/2019, 4)
- 01RONKG – Python in the Lab (03/09/2019, 4)
- 08IXTRV – Project management (12/7/2018, 1)
- 02LWHRV – Communication (21/06/2018, 1)

POLITECNICO DI TORINO

PhD program in
Electrical, Electronics and Communications Engineering