Graph Neural Networking challenge 2023:

Building a Network Digital Twin using data from real networks

https://bnn.upc.edu/challenge/gnnet2023

Carlos Güemes, Prof. Albert Cabellos

{carlos.guemes, alberto.cabellos}@upc.edu

Barcelona Neural Networking center

Universitat Politècnica de Catalunya

July 3rd 2023

What is the Graph Neural Networking challenge?

Graph Neural Networking challenge

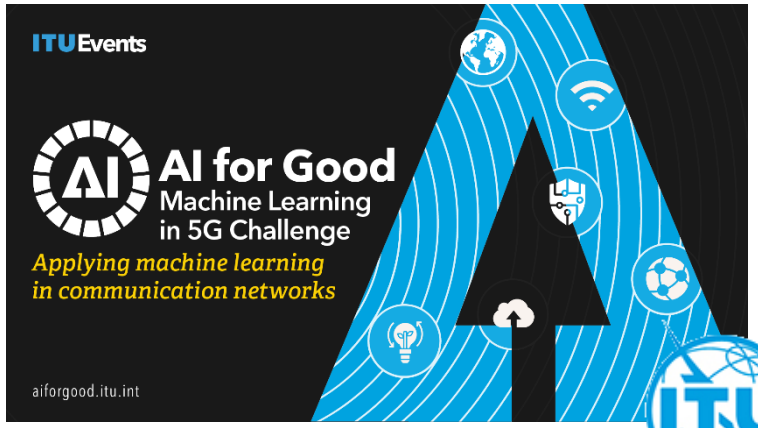
https://bnn.upc.edu/challenge/gnnet2023

- **Series of annual competitions on Graph Neural Networks applied to Networking**
- Each edition brings a fundamental challenge on GNNs applied to Computer Networks:
  - Graph Neural Networking challenge 2020: Modeling QoS-aware queue scheduling policies at networks.
  - Graph Neural Networking challenge 2021: Creating a Scalable Network Digital Twin
  - Graph Neural Networking challenge 2022: Improving Network Digital Twins through Data-centric AI

Graph Neural Networks are becoming a hot topic in networking!

It is the first (and the only) competition on GNNs applied to computer networks

ITU AI/ML in 5G challenge



https://aiforgood.itu.int/about/aiml-in-5g-challenge

- Organized as part of the **ITU AI/ML in 5G challenge**

- Several problem statements on AI/ML applied to networks, one of them is the Graph Neural Networking challenge

Graph Neural Networking challenge 2023

Problem statement:

Creating a Network Digital Twin with Real Network Data

Cash Prizes:

1st Prize: 2000 EUR

2nd Prize: 500 EUR

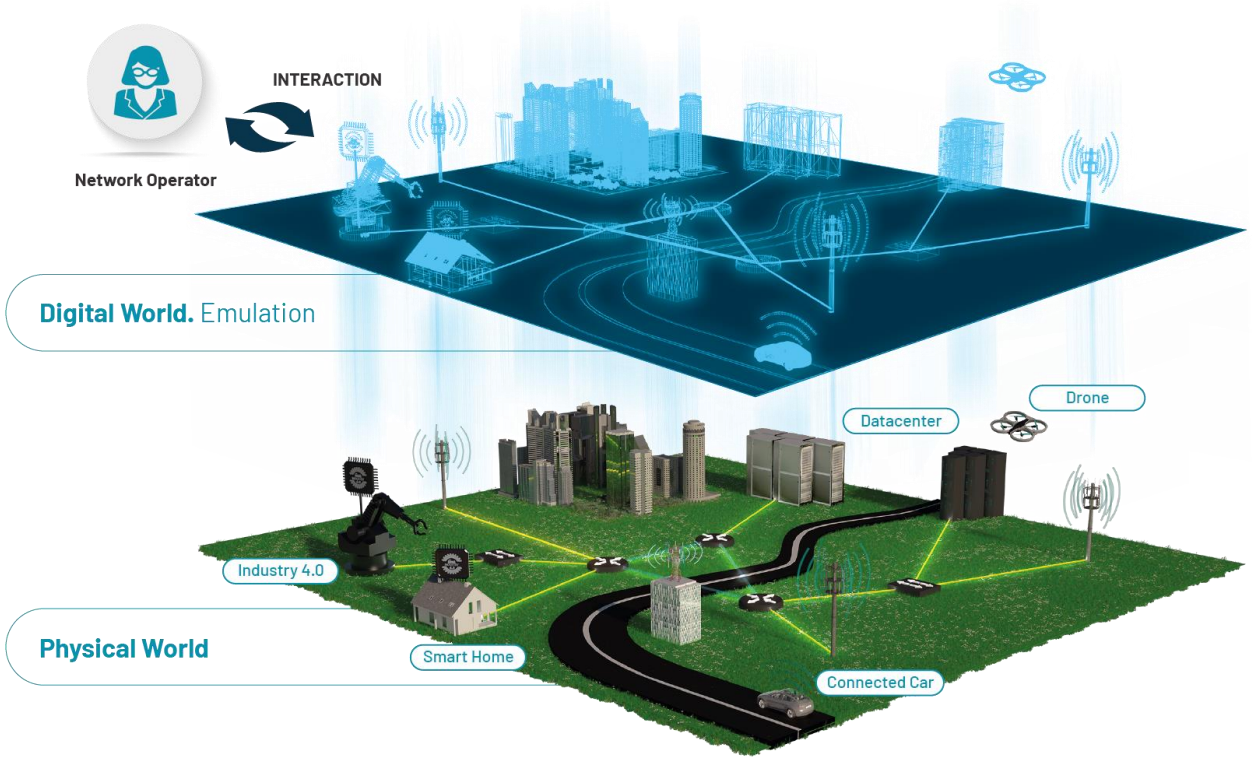
What is a Network Digital Twin?

What is a Digital Twin?

- A digital twin is a virtual replica of a physical object or process

- It permits to simulate the behavior of a physical system under certain input conditions:
  - What will happen if there is a specific failure? (e.g., in the electrical system)
  - What happen if I make a change in the object? (e.g., new wing design)

What is a Network Digital Twin?



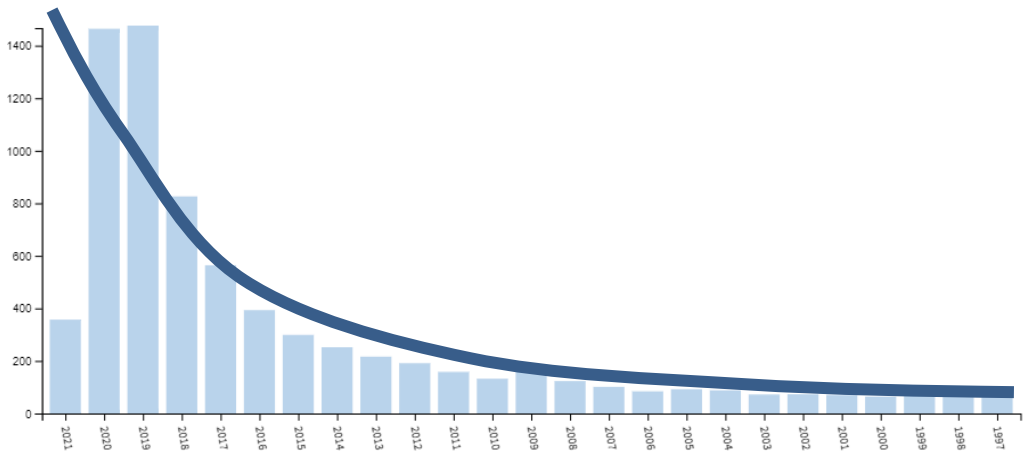
- A Network Digital Twin is a virtual replica of a physical network
- It enables to reproduce the behavior of the network under certain what-if scenarios:
  - What happens if I change the configuration?
  - What happens if there is a random failure?

Building a Network Digital Twin

What are Graph Neural Networks?

GNN are the next big thing in AI

- A Graph Neural Network (GNN) is a class of artificial neural networks for processing graph-data [1]
- Top trends in Graph Machine Learning in 2020 “New cool applications of GNN” [2]
- GNN related publications in networking (e.g,. SIGCOMM 2021 [4])



Exponential growth of papers
Source: Web of Science

time

[1] Battaglia, Peter W., et al. "Relational inductive biases, deep learning, and graph networks." arXiv preprint arXiv:1806.01261 (2018).

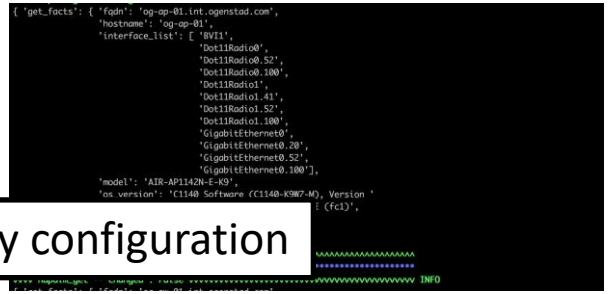
[2] https://towardsdatascience.com/top-trends-of-graph-machine-learning-in-2020-1194175351a3

[3] https://medium.com/mlreview/machine-learning-on-graphs-neurips-2019-875eecd41069

[4] Wang, Xiaojian, Jingyuan Wang, and Ke Tang. "Interpreting Deep Learning Model Using Rule-based Method." *arXiv preprint arXiv:2010.07824* (2020).

State-of-the-Art Network Digital Twins?

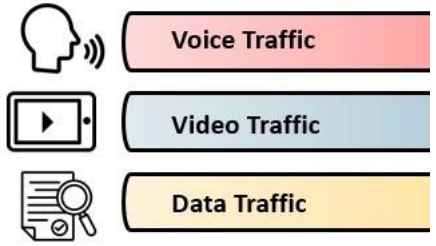
Network Digital Twin

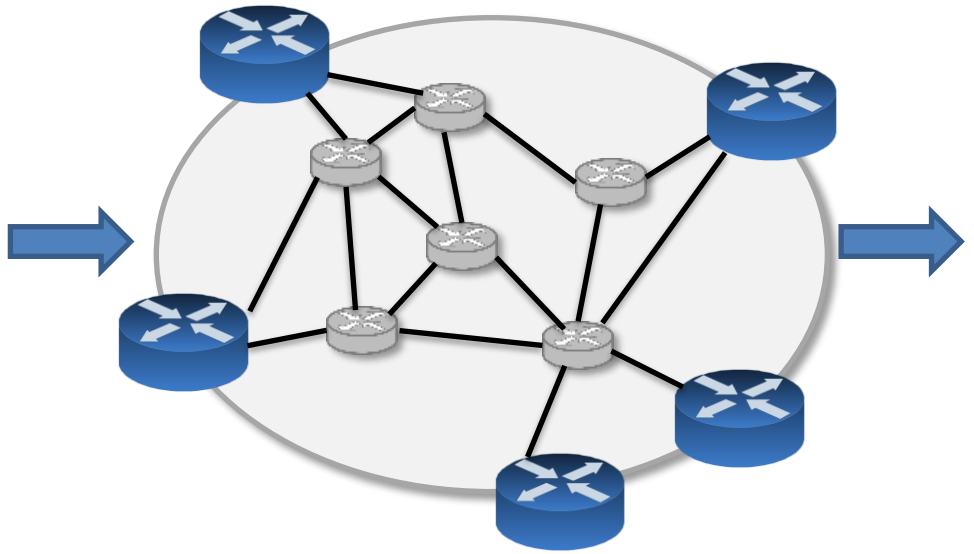


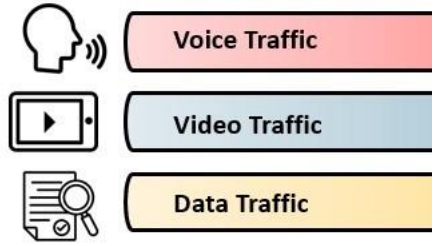
Apply configuration

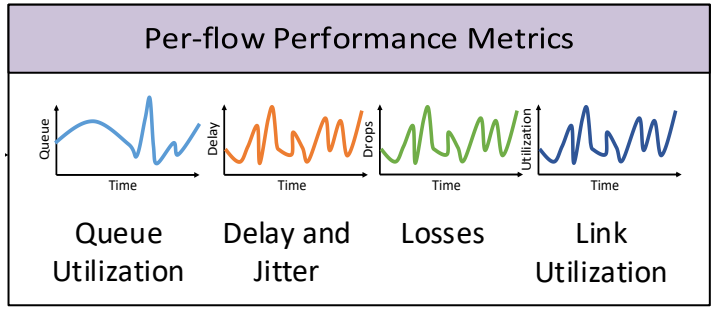
Real Network

Traffic







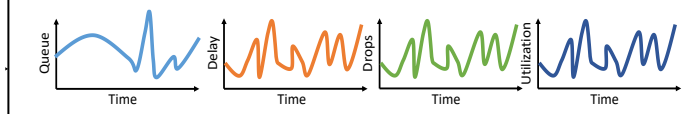


Network Digital Twin

A description of:

Network Digital Twin

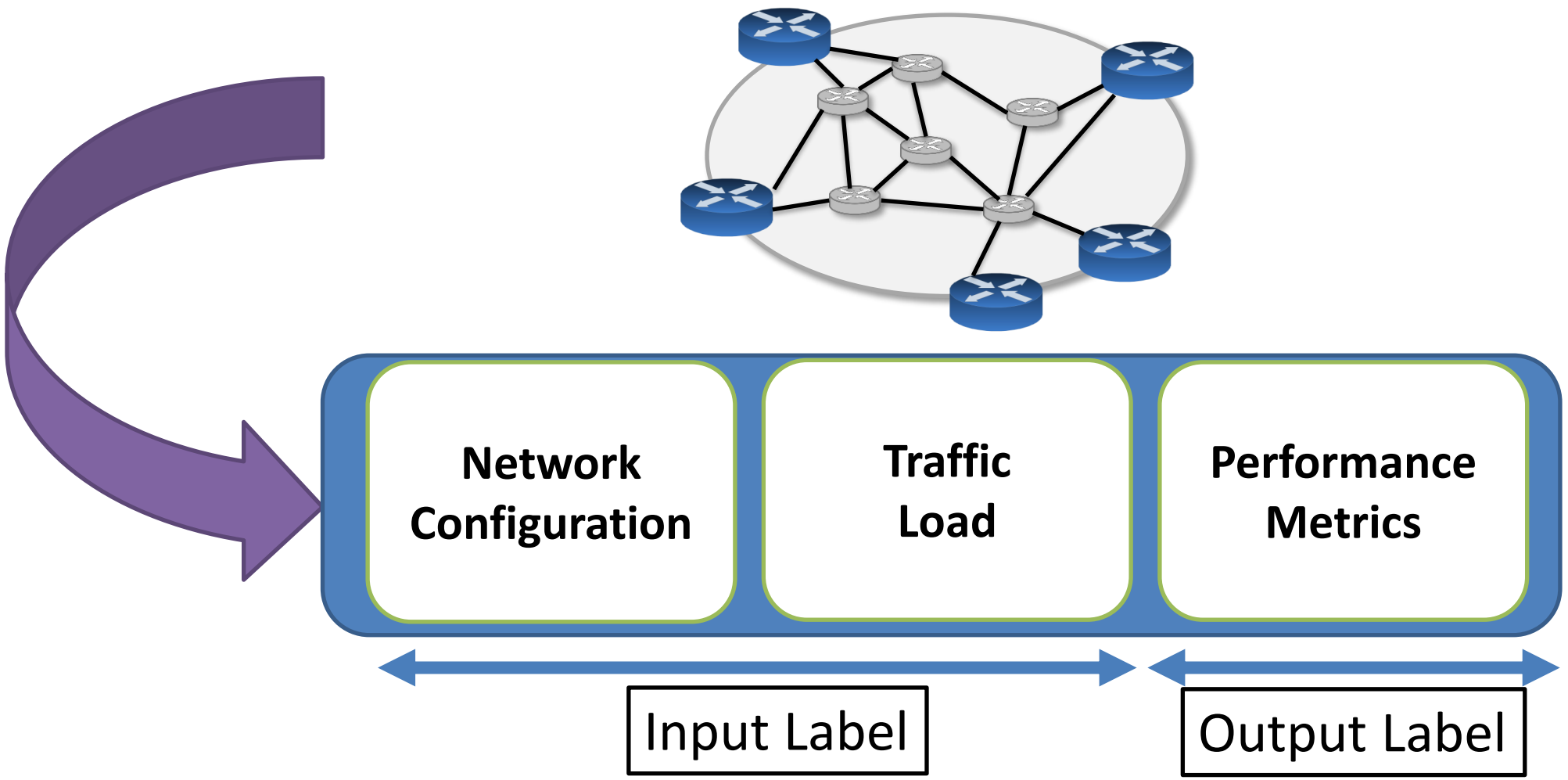
Per-flow Performance Metrics



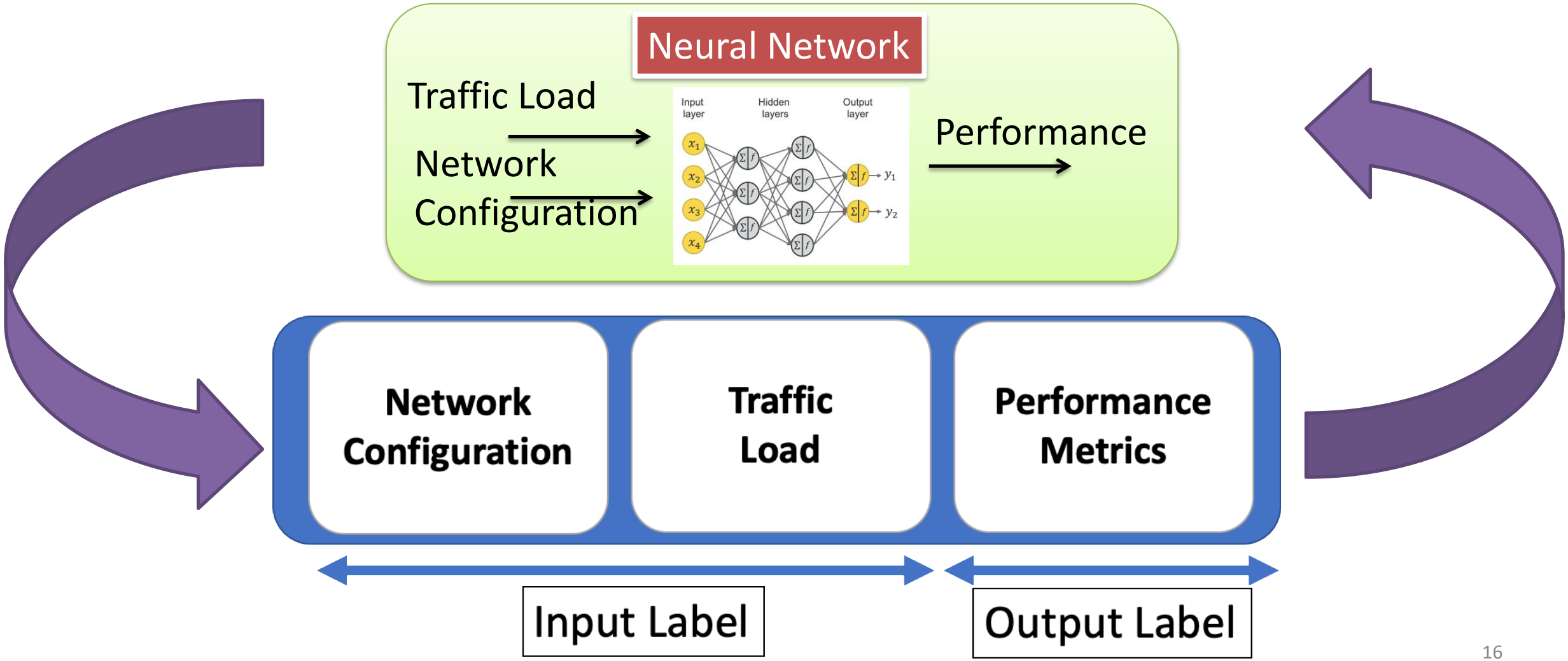
Queue Utilization | Delay and Jitter | Losses | Link Utilization

Note that configuration and traffic con be obtained via a standard Telemetry and Management platforms

Building a Network Digital Twin



Building a Network Digital Twin



Building a Network Digital Twin using GNNs

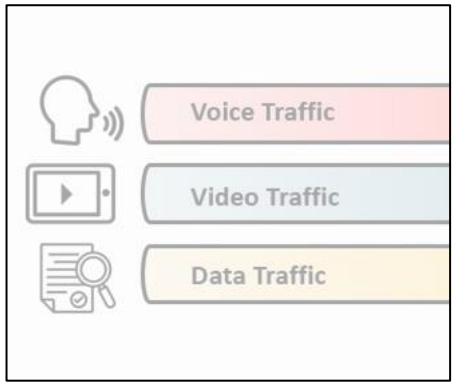
Configuration

Topology, Link Capacity
Routing
- Overlay: SRv6, MPLS…
- Underlay: OSPF, BGP…

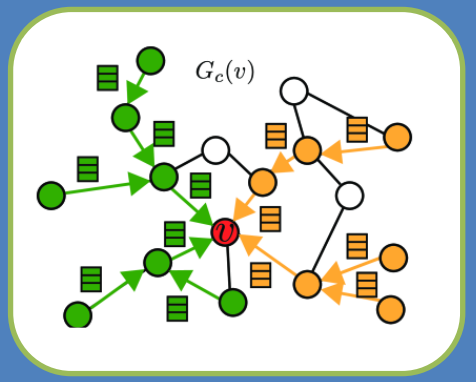
Scheduling Policy (arbitrary)
- Queue Length
- Policy
- Hierarchy of policies

ECMP, LAG, etc

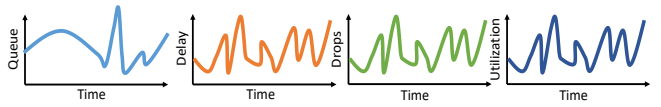
A description of traffic:



Network Digital Twin



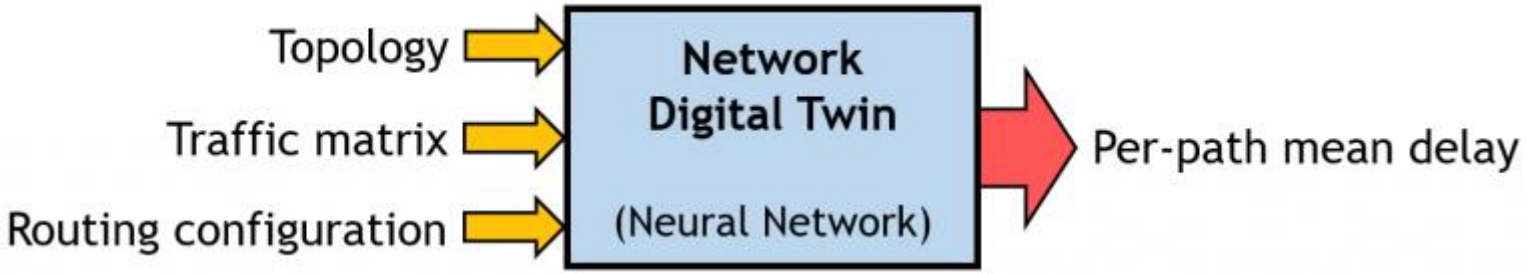
Per-flow Performance Metrics



Queue Utilization | Delay and Jitter | Losses | Link Utilization

State of the Art in Network Digital Twins

- RouteNet-Fermi* is the outcome of the knowledge gathered after three editions of the GNN challenge
- Capabilities:
  - Scale to topologies 100x larger than these seen in training
  - Supports arbitrary scheduling policies
  - Supports traffic models as characteristics of the underlying inter-arrival distribution
    - E.g, Poisson traffic is described by its λ

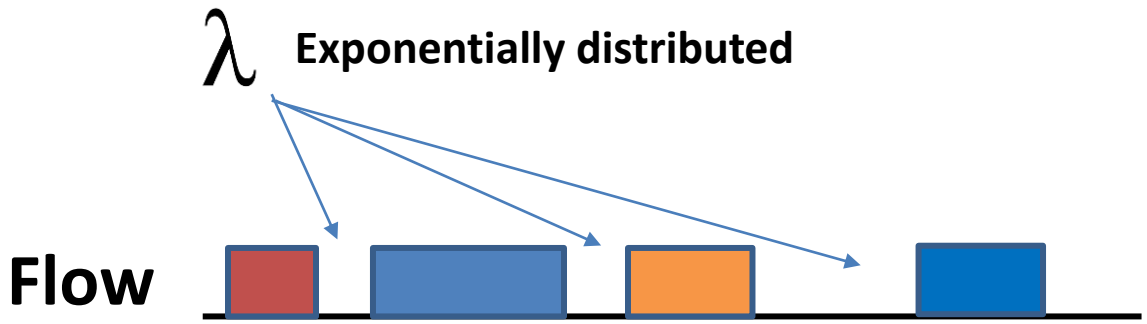


(*) Ferriol-Galmés, Miquel, Jordi Paillisse, José Suárez-Varela, Krzysztof Rusek, Shihan Xiao, Xiang Shi, Xiangle Cheng, Pere Barlet-Ros, and Albert Cabellos-Aparicio. "RouteNet-Fermi: Network Modeling With Graph Neural Networks." *IEEE/ACM Transactions on Networking* (2023).

But…

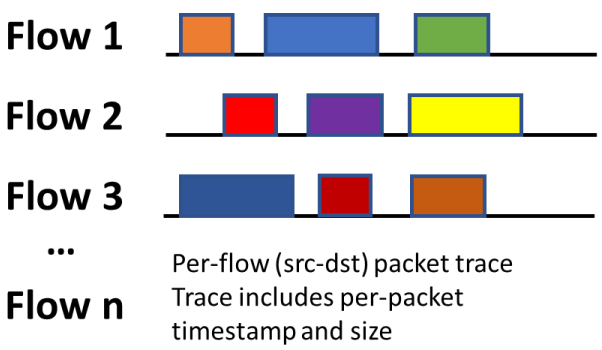
Limitations of the current Datasets

- Datasets to train RouteNet-Fermi have been obtained from **simulation**
  - Omnet++
- Traffic is generated using **stationary traffic distributions**
  - Poisson, On-off, etc
- The moments of such distributions where used as input to RouteNet
  - Example: λ in the case of Poisson



Datasets from real-networks and packet-traces

- We have built a dataset from a real network
  - 2x48-port switches
  - 8xHuawei NetEngine 8000 M1A
  - Traffic generated with T-Rex
  - Delay measured with Mellanox ConnectX-5 cards
- Traffic is measured at the packet-level



Graph Neural Networking challenge 2023

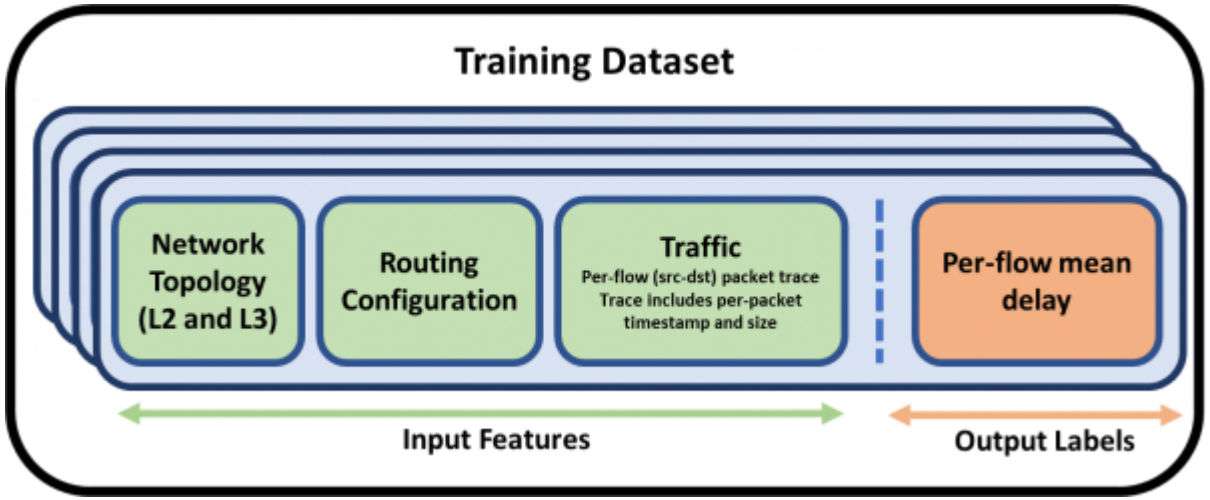
Problem statement:

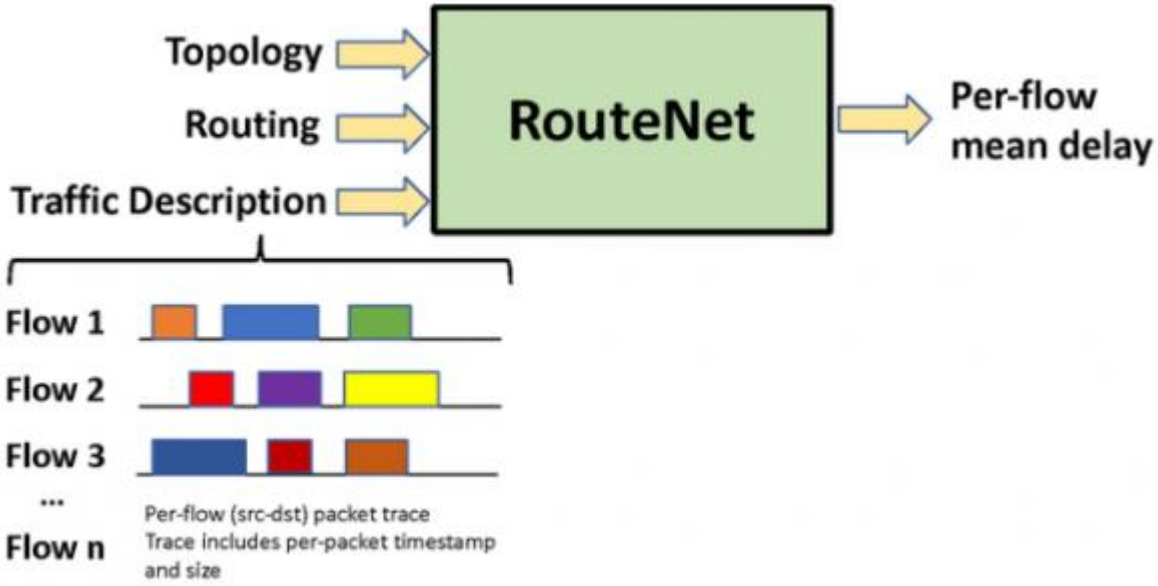
Creating a Network Digital Twin with Real Network Data

Cash Prizes:

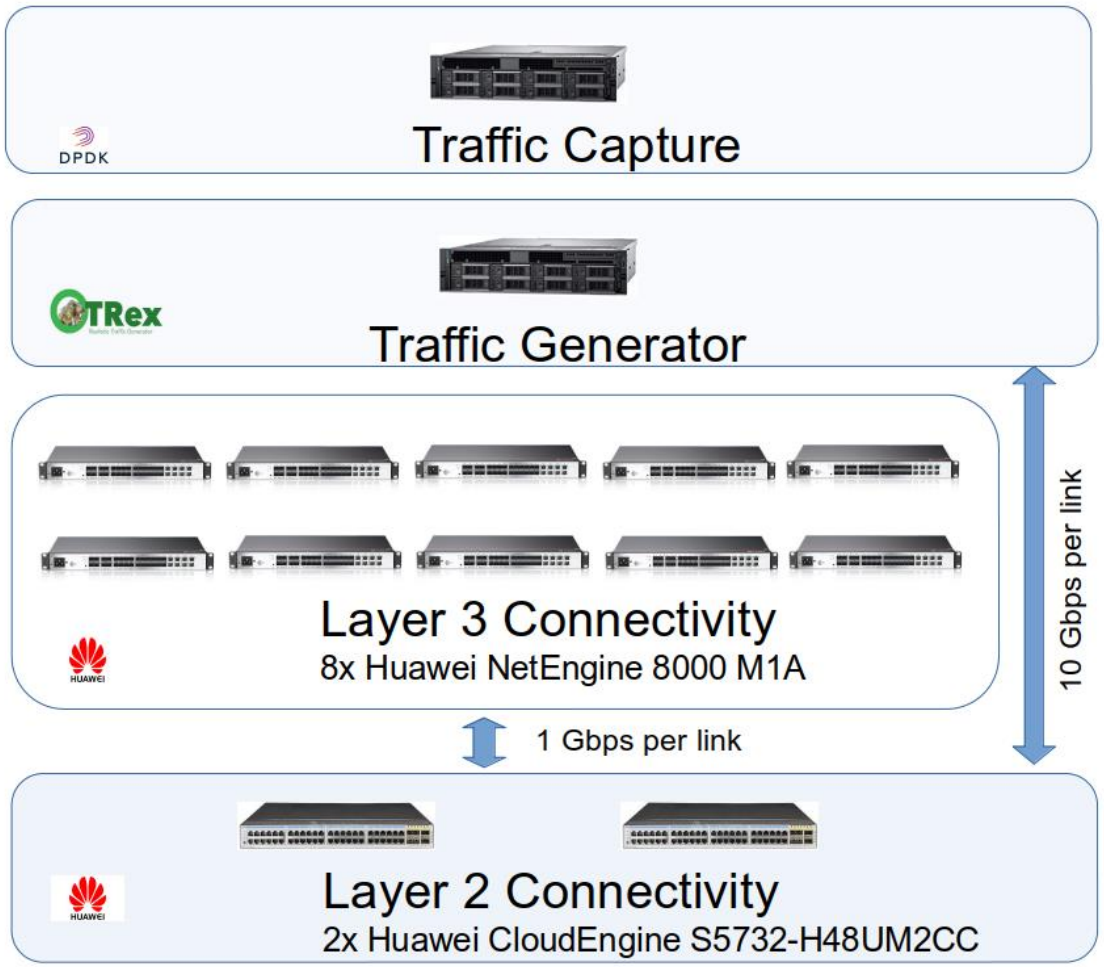
1st Prize: 2000 EUR

2nd Prize: 500 EUR



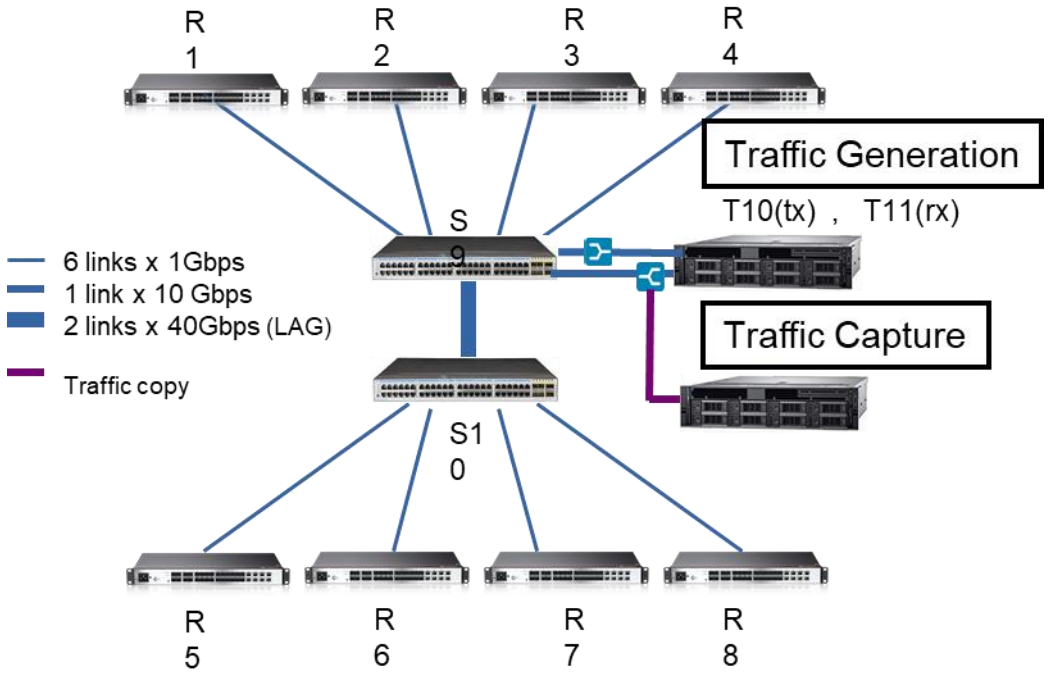


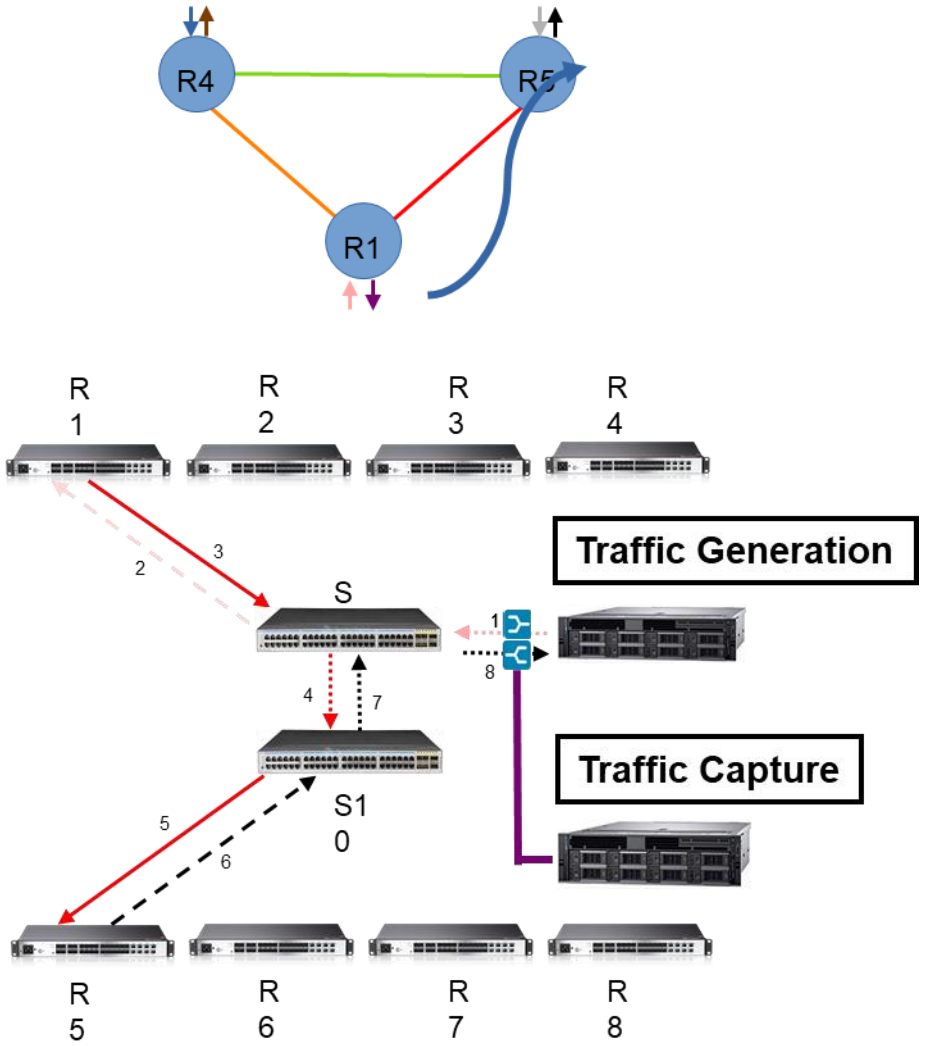
- The challenge is to build a Network Digital Twin that can estimate per-flow mean delay based on the **input per-flow packet-trace**.
- Dataset is obtained from a **real-network**.



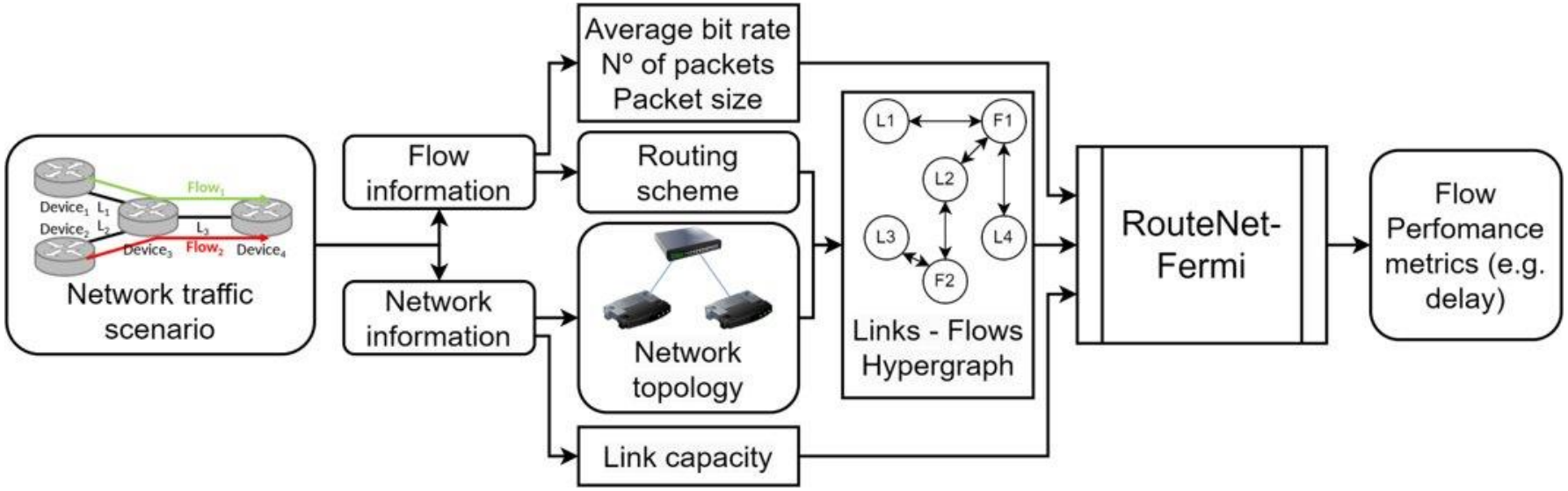
- Testbed includes:
  - Traffic generator (T-Rex)…
  - sends traffic through a set of Routers…
  - that are connected via 2 switches…
  - Traffic is addressed towards the final destination router…
  - …and captured by the Capture server. Then the per-flow delay is computed.

Topology and physical path traversed by the flows





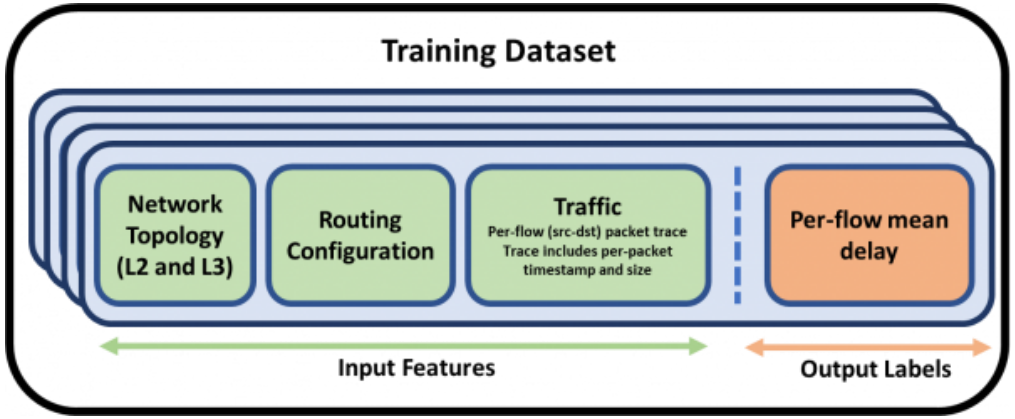
Baseline: RouteNet-Fermi

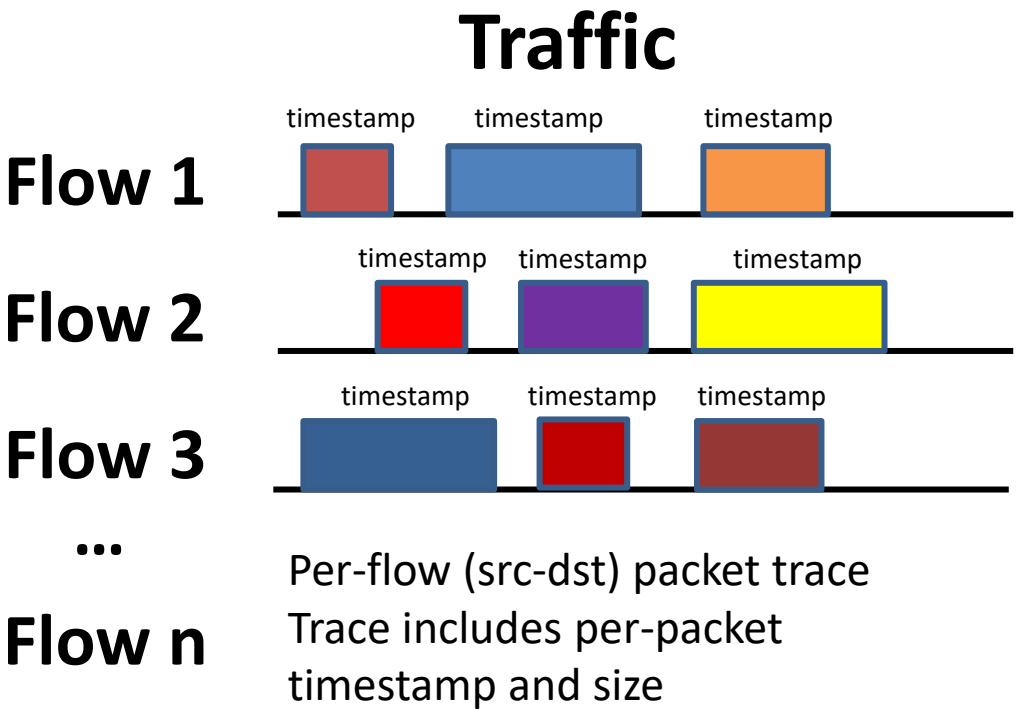


- Open-source implementation in tensorflow found a: https://github.com/BNN-UPC/GNNetworkingChallenge/tree/2023_RealNetworkDT.
- Modify the baseline or start from scratch
  - **Already supports the physical topology (L2 and L3) of the testbed**
- RouteNet-Fermi **does not:**
  - Support packet traces as input
  - Real-world networks characteristics

Baseline has 35% MAPE with the datasets

Dataset

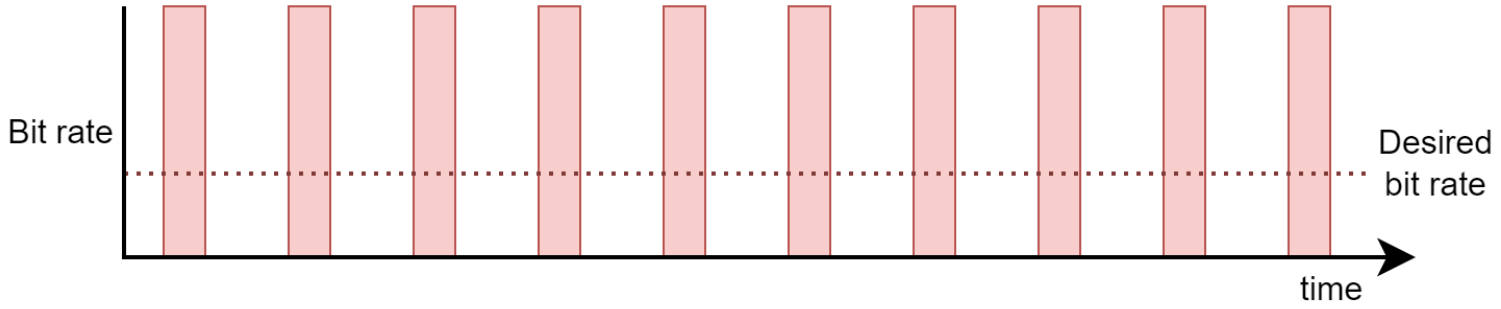




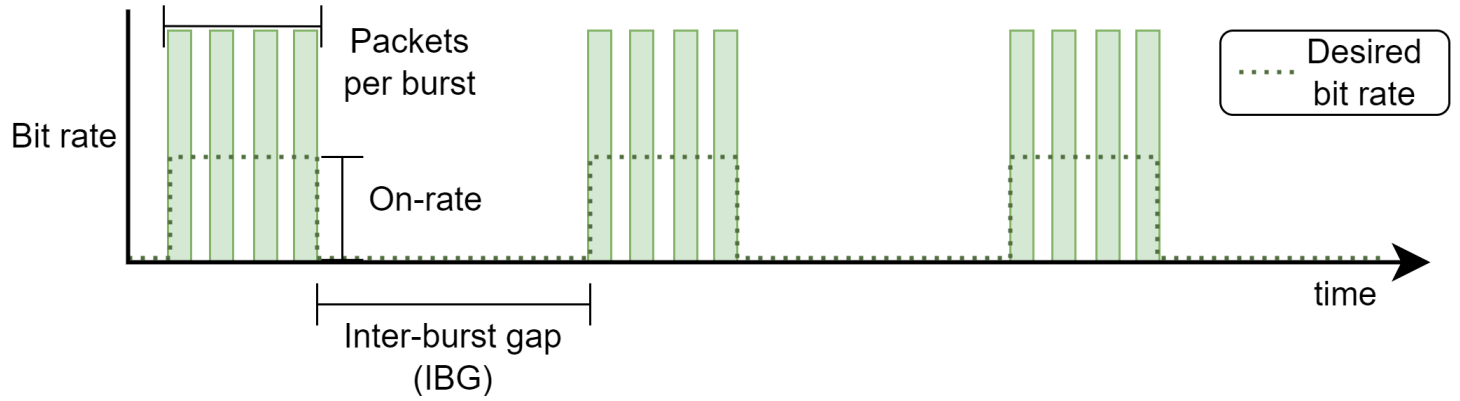
- Dataset includes ≈10.000 samples
  - 4227 samples Constant Bit-Rate samples and MultiBurst
  - 4388 samples MultiBurst
- Each sample contains:
  - Flow information: distribution parameters, number of packets sent, packet size
  - Path information: network routing and physical path
  - Topology information (router connections, link capacity)
  - Packet-level traces
  - Performance information **(only in training dataset)**: average flow and path delay, jitter, packet loss rates

Dataset

- CBR: traffic is sent in short and intense bursts so that the average bit rate is equal to the desired bit rate



- Multiburst: burst traffic defined by three parameters:



- **In both cases the packet size is constant**

Dataset

- Both dataset combined are **427.43GB** in size
  - Because we are publishing packet-traces
- Both datasets can be downloaded at: https://bnn.upc.edu/challenge/gnnet2023/dataset/
- Dataset is split onto (aprox.) 5GB files

You can start participating with just one file

Evaluation

- At the end of the challenge (Oct 2nd), we will evaluate participants' solutions on a **test dataset.** We will only release input labels, output labels will be hidden.

- The test dataset will follow similar distributions to the training dataset (released at the beginning)

- **The evaluation phase lasts 15 days**, and it is made automatically in our evaluation platform

- Participants will be ranked based on the MAPE (accuracy) of their solutions in the test dataset.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

- **Participants will see the ranking in real-time**

Evaluation

After the evaluation…

- **Provisional ranking** with the scores of all teams

- We will ask top-5 teams for:
  - Source code of their proposed solution
  - **A short report** describing their solution (1-3 pages)

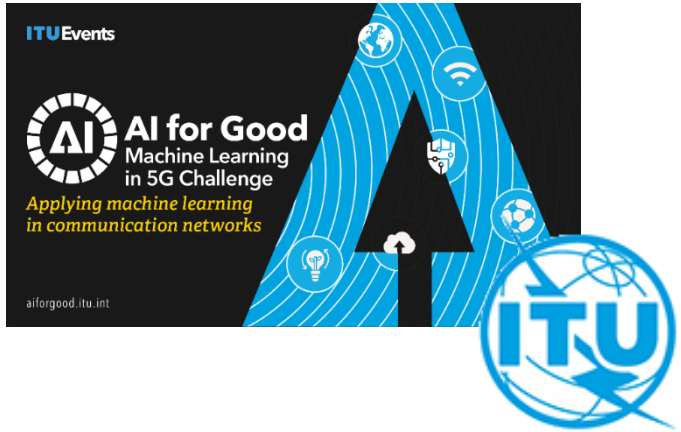
- We will **validate the top-5 solutions** to check that they comply with all the rules

- **We will not publish the solutions and if asked, we´ll be happy to sign an NDA**

Quick summary

Main resources:

- RouteNet-Fermi as baseline
  - You can extend it, you can came up with your own solution from scratch
- Dataset
- Quick-start tutorial
- Mailing list for Q&A from participants (support from organizers)



Final outputs from participants:

- Source code of their proposed solution
- **A short report** describing their solution (1-3 pages)

Barcelona Neural Networking Center

Expected outcomes:

- First Network Digital Twin trained with a dataset from a real-world network

Graph Neural Networking challenge 2023:

Building a Network Digital Twin using data from real networks

https://bnn.upc.edu/challenge/gnnet2022

Timeline

Registration is mandatory

- **Challenge duration:** June-Nov 2023
- **Open registration:** June 7th-Sep 30th
- **Release of tools and validation dataset:** June 30th
- **Score-based evaluation phase:** Oct 2nd-Oct 17th 2023
- **Provisional ranking of all the teams:** Oct 18th 2023
- **Top-5 teams submit the dataset, code and documentation:** Nov 1st 2023
- **Final ranking and official announcement of top-3 teams:** Nov 2023
- **Best solutions pitch in a 3-day event end of to determine the finalists:** 28 – 30 November 2023
  Award ceremony and presentations: December 13th, 2023

Graph Neural Networking challenge 2023

https://bnn.upc.edu/challenge/gnnet2023/

Problem statement:

Creating a Network Digital Twin with Real Network Data

Cash Prizes:

1st Prize: 2000 EUR

2nd Prize: 500 EUR

1.- Register at the challenge website
2.- Download the dataset (just the first two files)
3.- Download the baseline (RouteNet-Fermi)
and **start coding**

Graph Neural Networking challenge 2023:

Building a Network Digital Twin using data from real networks

- We recommend all participants that they start from the open-source baseline at https://github.com/BNN-UPC/GNNetworkingChallenge/tree/2023_RealNetworkDT.
- The repository also includes a guide in how to modify different aspects of the baseline.
- We will proceed now with a demonstration that will cover the following topics:
  - Downloading the dataset
  - How process dataset samples from its raw files into a Tensorflow compatible format
  - **How to extract new features (through an example: extracting the Inter-Packet Gap)**

Graph Neural Networking challenge 2023:

Building a Network Digital Twin using data from real networks

https://bnn.upc.edu/challenge/gnnet2023

Carlos Güemes, Prof. Albert Cabellos

{carlos.guemes, alberto.cabellos}@upc.edu

Barcelona Neural Networking center

Universitat Politècnica de Catalunya

July 3rd 2023

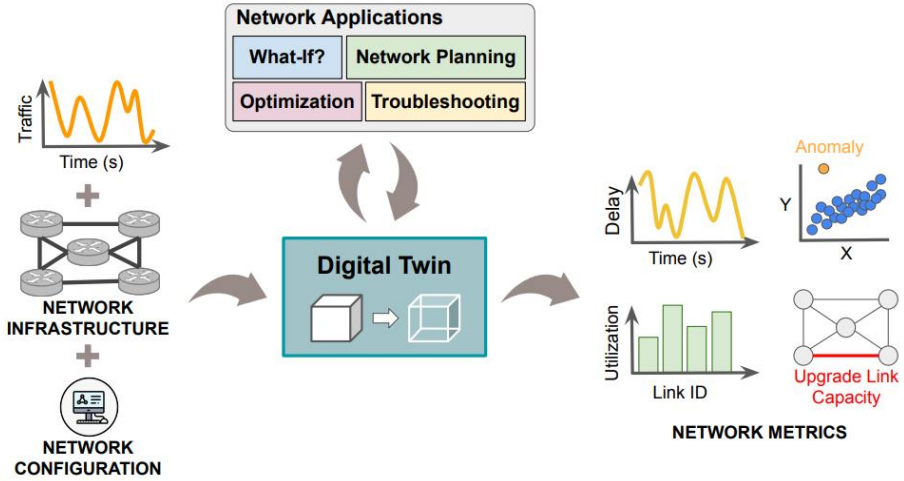
Backup slides

Network Digital Twin: Use case examples

Digital Twins can be applied to many fundamental networking applications*

Network Optimization and What-if analysis

- What happens if we re-route traffic on another path? (Traffic Engineering)
- Can I support new user SLAs with the current resources?



Network Planning:

- Which is the best network upgrade within a limited budget?

Troubleshooting:

There was a temporary service disruption that affected some SLAs:

- What was the root cause?
- Can we find a way to prevent this in the future? (*e.g.*, add link redundancy)

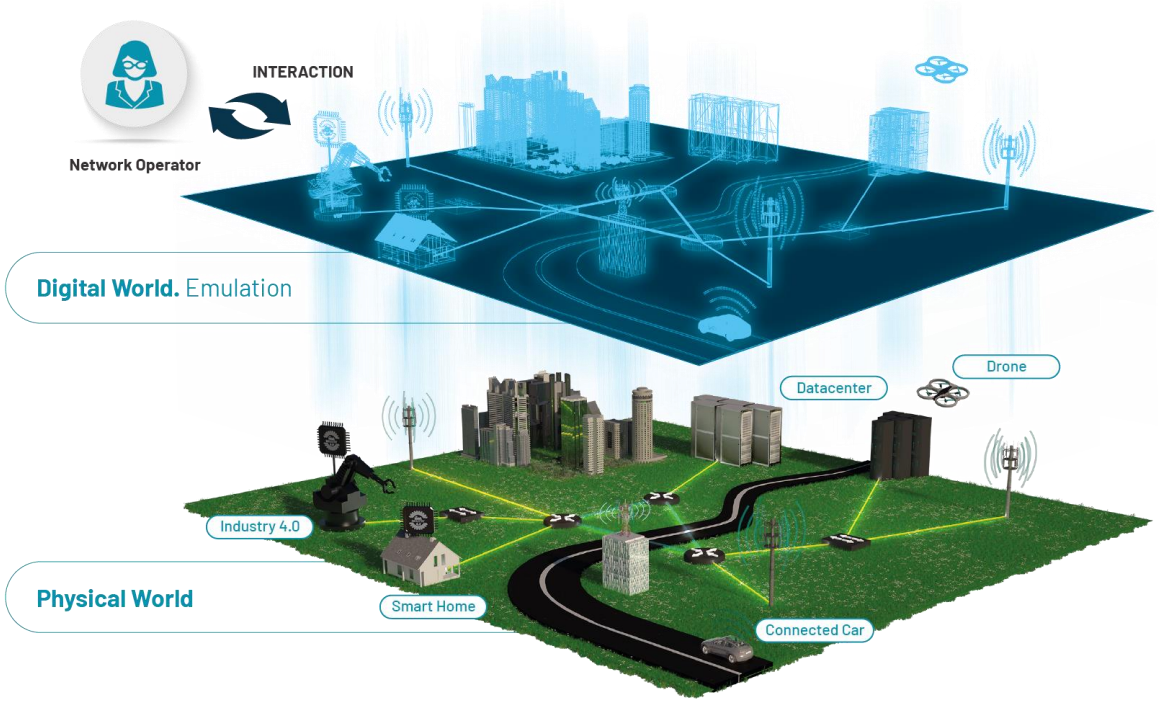
* P. Almasan, M. Ferriol-Galmés, J. Paillisse, J. Suárez-Varela, D. Perino, D. López, A. A. Pastor Perales, P. Harvey, L. Ciavaglia, L. Wong, V. Ram, S. Xiao, X. Shi, X. Cheng, A. Cabellos-Aparicio, P. Barlet-Ros, "Network Digital Twin: Context, Enabling Technologies and Opportunities," arXiv preprint arXiv:2201.01144, 2022

Network Digital Twin: Context

Is this a new concept?

What about the existing literature on network modeling?

(e.g., network simulators, analytical models)



Network Digital Twin: Context

Is this a new concept?

What about the existing literature on network modeling?

(e.g., network simulators, analytical models)

- **Network Digital Twins**:
  - Renovated concept of classical network modeling with the ambition of achieving **accurate real-time digital replicas** of the network
  - **Machine learning (ML)** is promising for building **accurate and lightweight** data-driven network models