Fundamentos, investigación y casos reales de la IA/ML en la gestión de redes

Christian Rothenberg
FEEC/UNICAMP
9 de Mayo 2025

SMARTNESS

lacnic 43
5-9 de Mayo 2025 / São Paulo, Brasil

IA/ML & Redes

lacnic43
5-9 de Mayo 2025 / São Paulo, Brasil

Agenda

SMARTNESS

- **Motivación**
- **Fundamentos de IA/ML**
- **Investigación**
  - Automatización, QoE, detección de problemas , etc. etc. etc.
- **Casos reales**
  - YouTube QoE en redes 4G/5G
  - Identificación de fallos suaves en redes ópticas usando Network Digital Twin
- **Conclusiones y recomendaciones**

Sobre mí

Christian Esteve Rothenberg

- Associate Professor (tenure track) at FEEC/UNICAMP (since 2013)
  - Leading the INTRIG lab at DCA/FEEC/UNICAMP

    INTRIG: Information & Networking Technologies Research & Innovation Group
  - Currently, supervising 7 PhD, 7 MSc candidates, and 4 undergrad students
    - Alumni: 5+ Postdocs, 9 PhDs, 25+ MSc, 50+ undergrad projects.
- PhD in Electrical and Computer Engineering (FEEC/UNICAMP, 2010)
- MSc in Electrical Eng and Information Technology (Darmstadt University, 2006)
  - Research internet at Deutsche Telekom T-Systems
- Telecommunication Eng (Universidad Politécnica de Madrid, 2004)
- Visiting researcher at Ericsson Research Nomadic Lab, Jorvas, Finland, 2008,
- Research Scientist at CPqD R&D Center in Telecommunication (2010-2013)
  - Technical Lead of SDN activities in the Converged Networking Division
- Currently. PI / Director of the SMARTNESS Engineering Research Center (ERC)

vmware
nicira
SmartNIC
vs.
Standard NIC
Professor Scott Shenker
UC Berkeley
An attempt to motivate and clarify Software-Defined Networking (S…
BAREFOOT NETWORKS
LF NETWORKING
OPEN NETWORKING FOUNDATION
Maturity, Integration, Convergence of SDN/NFV/IBN, Emergence of P4 ("OF2.0")
SDN Adoption; NFV Emergence
SDN in 5G, Edge Computing, and Cloud-Native Networking
SDN Evolved: Automation, AI, and the Future
SDN@6G
Genesis of SDN and OpenFlow
Commercialization., Ecosystem Growth
2009
2014
2019
2024
Martin Casado <casado@nicira.com>
to Christian
Hey Christian,
Attached is the current draft. Onix is a com
than Nox and there are no plans to release
cheers,
.martin
One attachment • Scanned by Gmail
paper.pdf
CPqD
UNICAMP
RouteFlow
libfluid
The ONF OpenFlow driver
Software-Defined Networking: A Comprehensive Survey
Cited by 6784
Ramon Fontes
Christian Rothenberg
Wireless Network Emulation with Mininet-WiFi
Management plane
Control plane
Data plane
Net App
Network Applications
Programming Languages
Language-based Virtualization
Northbound Interface
Network Operating System
Network Hypervisor
Southbound Interface
Network Infrastructure
Debugging, Testing & Simulation
Network Operating System (NOS) and Network Hypervisors
SMARTNESS
FAPESP
2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
Ofsoftswitch 1.2. 1.3
ERICSSON
GitHub
fibre
FUTURE INTERNET BRAZILIAN ENVIRONMENT FOR EXPERIMENTATION
RNP
necos
Padtec
openRAN BRASIL

What is SMARTNESS 2030?

CPE: FAPESP Engineering Research Center (ERC)

- Co-Financed by the São Paulo Research Foundation (**FAPESP**).
- FAPESP is a **solid and stable** foundation, with budget of 1% of all state taxes collection (3.5 Billion SEK in 2021).
- ERC is FAPESP's **top program** for collaborative **research with Industries.**
- ERC premise: the execution of internationally competitive research in accordance with **global excellence** benchmarks.
- There are currently more than 15 ERC in different technological areas, e.g., oil and gas, biotechnology, agribusiness, energy, artificial intelligence, etc.

- **SMARTNESS** is the **first ERC** in the **Telecom area**

https://fapesp.br/cpe

SMARTNESS 2030

A networking-centric Engineering Research Center

Mission

Cutting-edge research in communication networks and advanced digital application services.

Towards 6G.

Founders

Ericsson, UNICAMP, USP and UFSCar.

Hub center at UNICAMP.

Long-term investment

10 years.

56 MBRL (~120 MSEK)
1:1:2 – Ericsson: FAPESP: UNICAMP.

50+ associated researchers
15+ university partners
120+ scholarships

History/ Status

2018-20 – Work on the Proposal

Feb/ 2021 – Prop. submission

May / 2022 – FAPESP approval

Dec / 2022 – Kick-off ceremony

April 2023 - Official start

Scientific & Technology **Advancements**

STAs

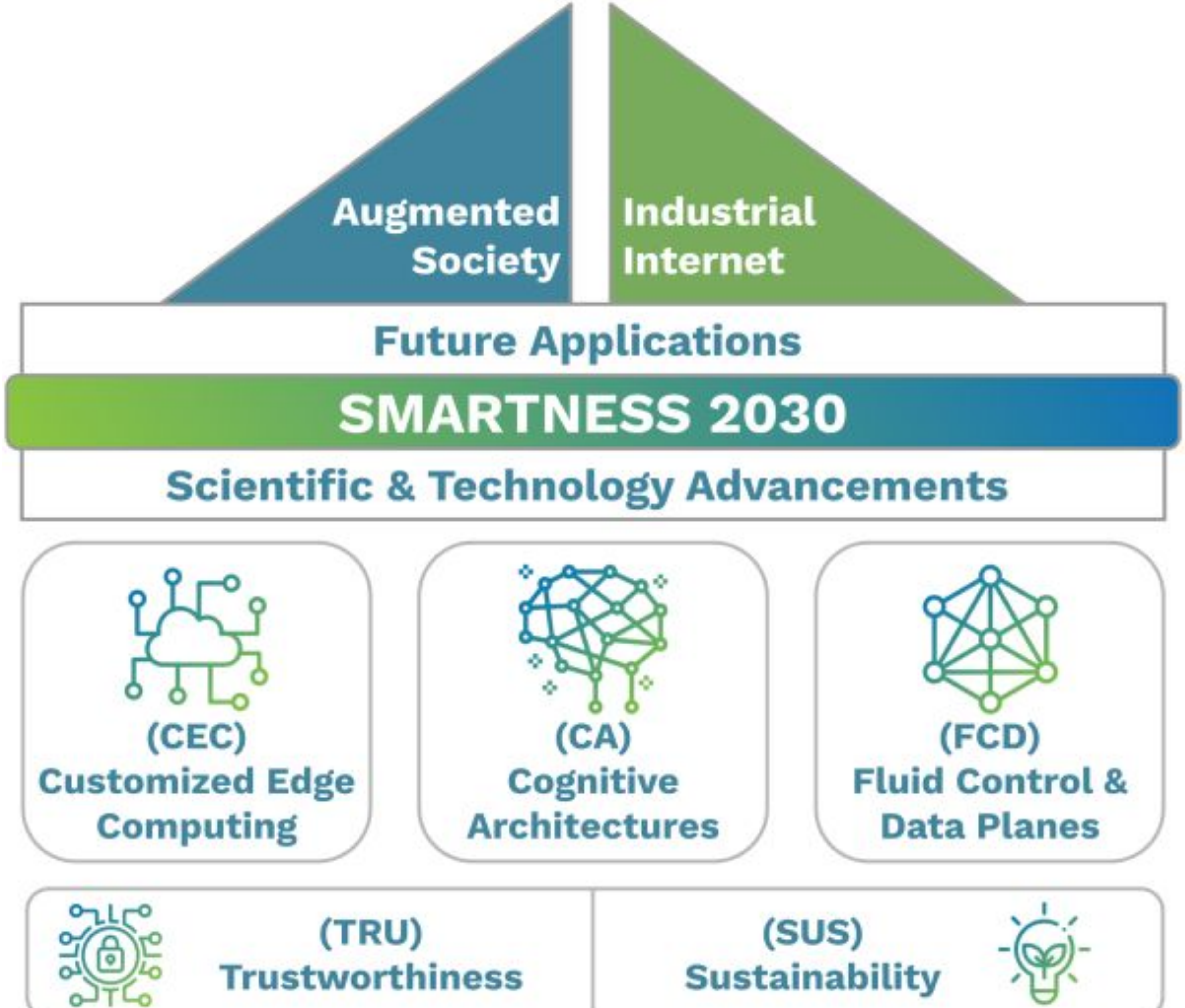
CEC: Customized Edge Computing

CA: Cognitive Architectures & Machine Intelligence

FCD: Fluid Control & Data planes

TRU: Trustworthiness

SUS: Sustainability



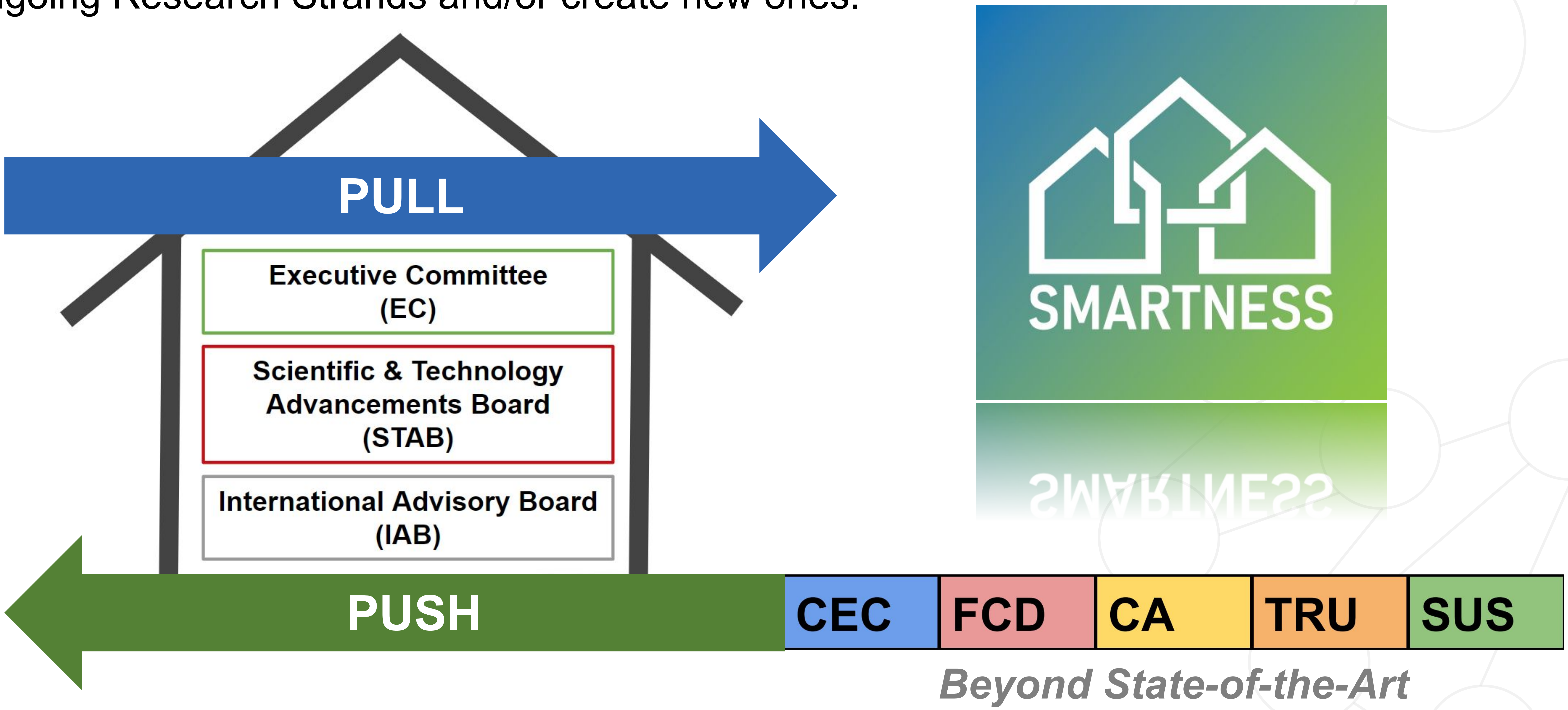
PUSH/PULL modes of operation

"Technology Push & Market Pull" like workflows

- **Academic Research Push:** New research findings, ideas, trends, etc. from SMARTNESS pushed to Ericsson Research
- **Ericsson Research Pull:** New research contribution demands/opportunities from projects / standards brought to SMARTNESS 2030 to shape ongoing Research Strands and/or create new ones.

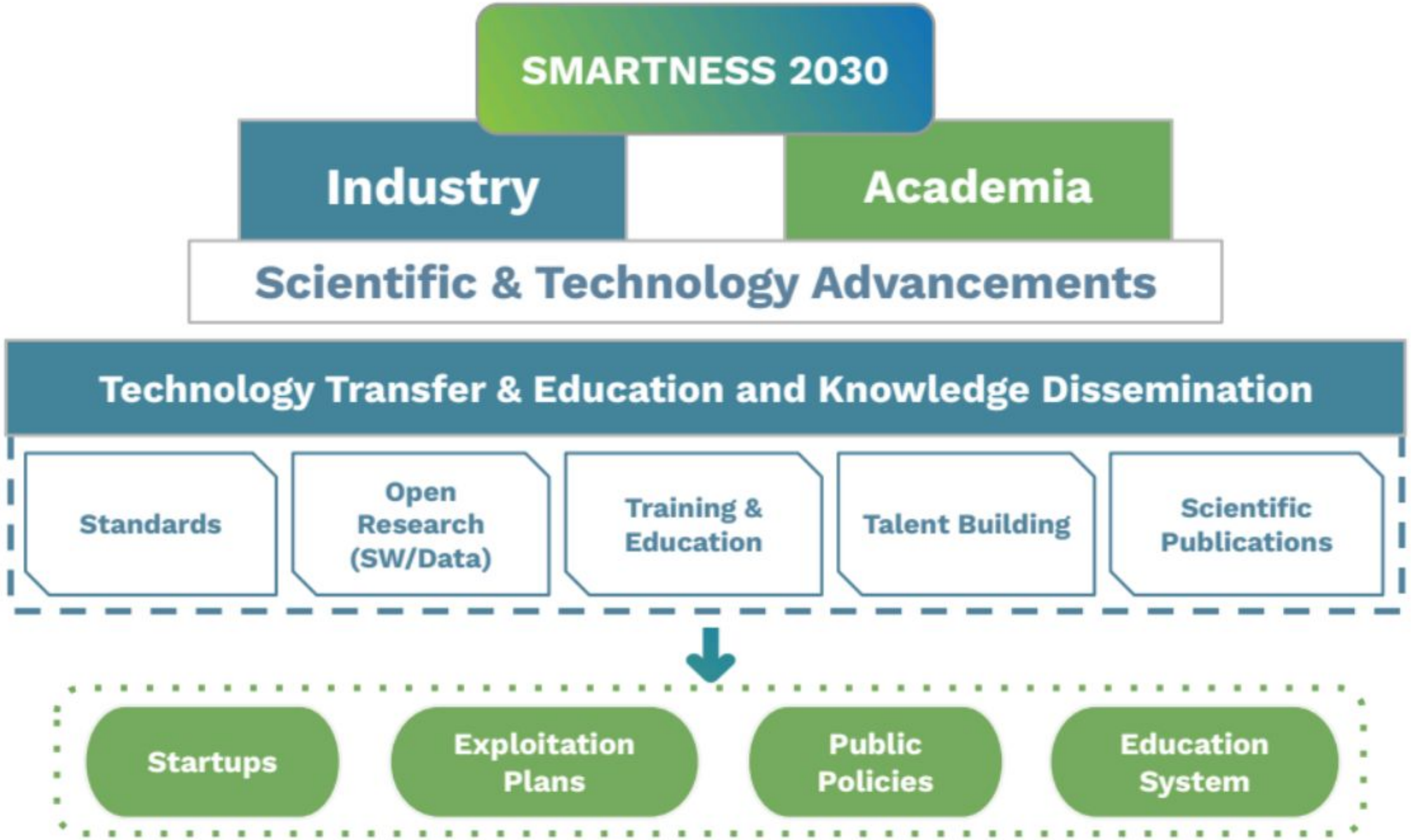
ERICSSON

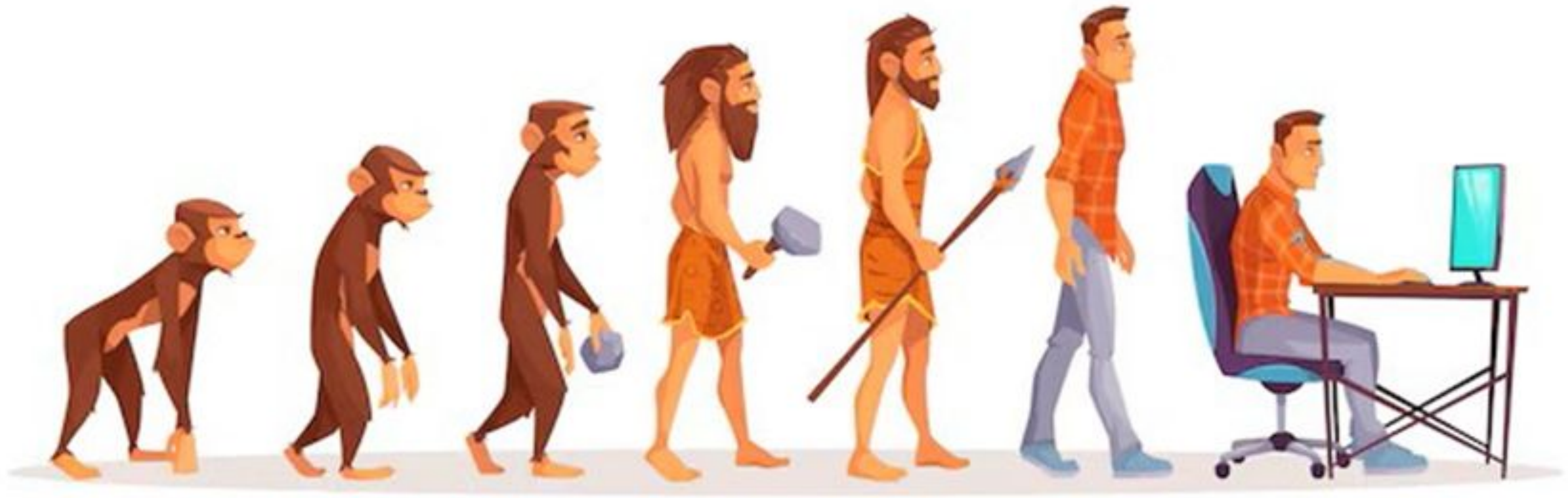
- Technology Journeys
- Future Network Programs
- EU SNS JU Projects
- Standardization
- Open Source
- Etc.



Sobre SMARTNESS (2023-2033)

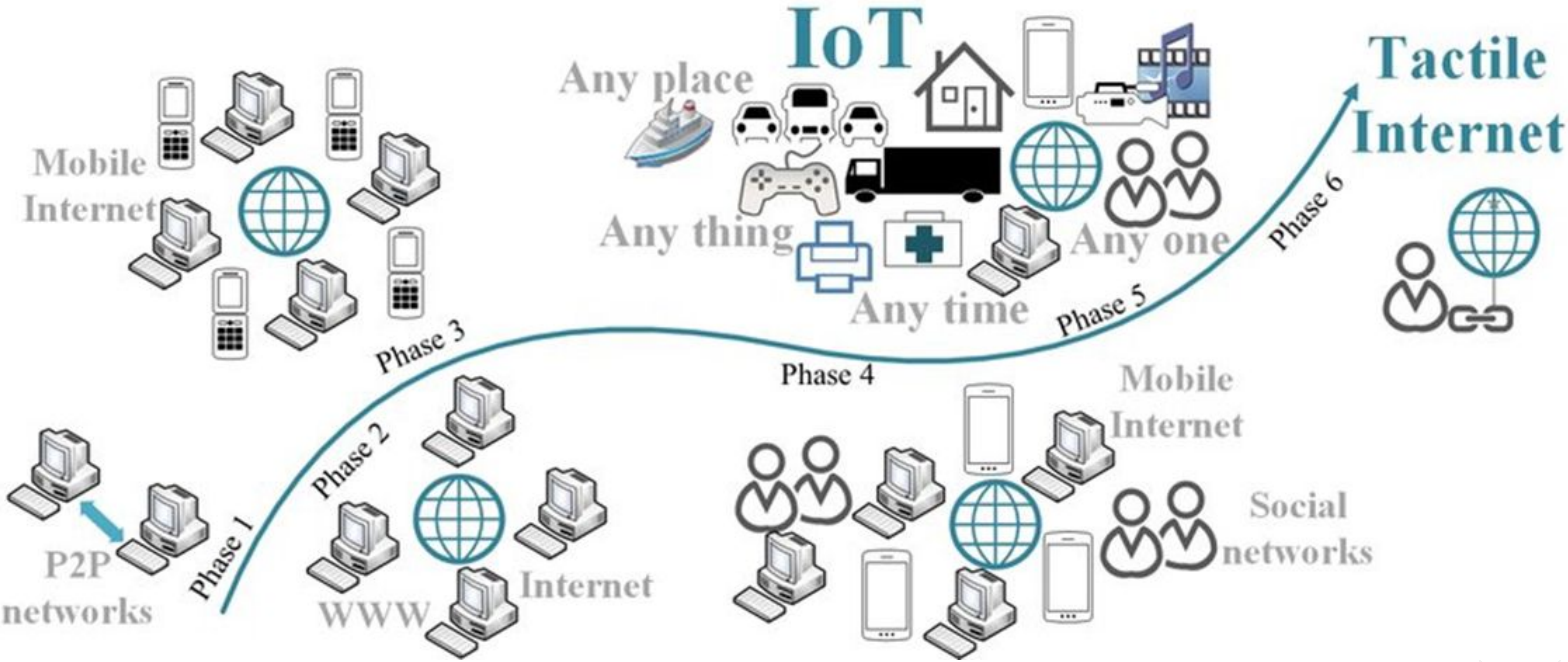
CPE: FAPESP Engineering Research Center (ERC)



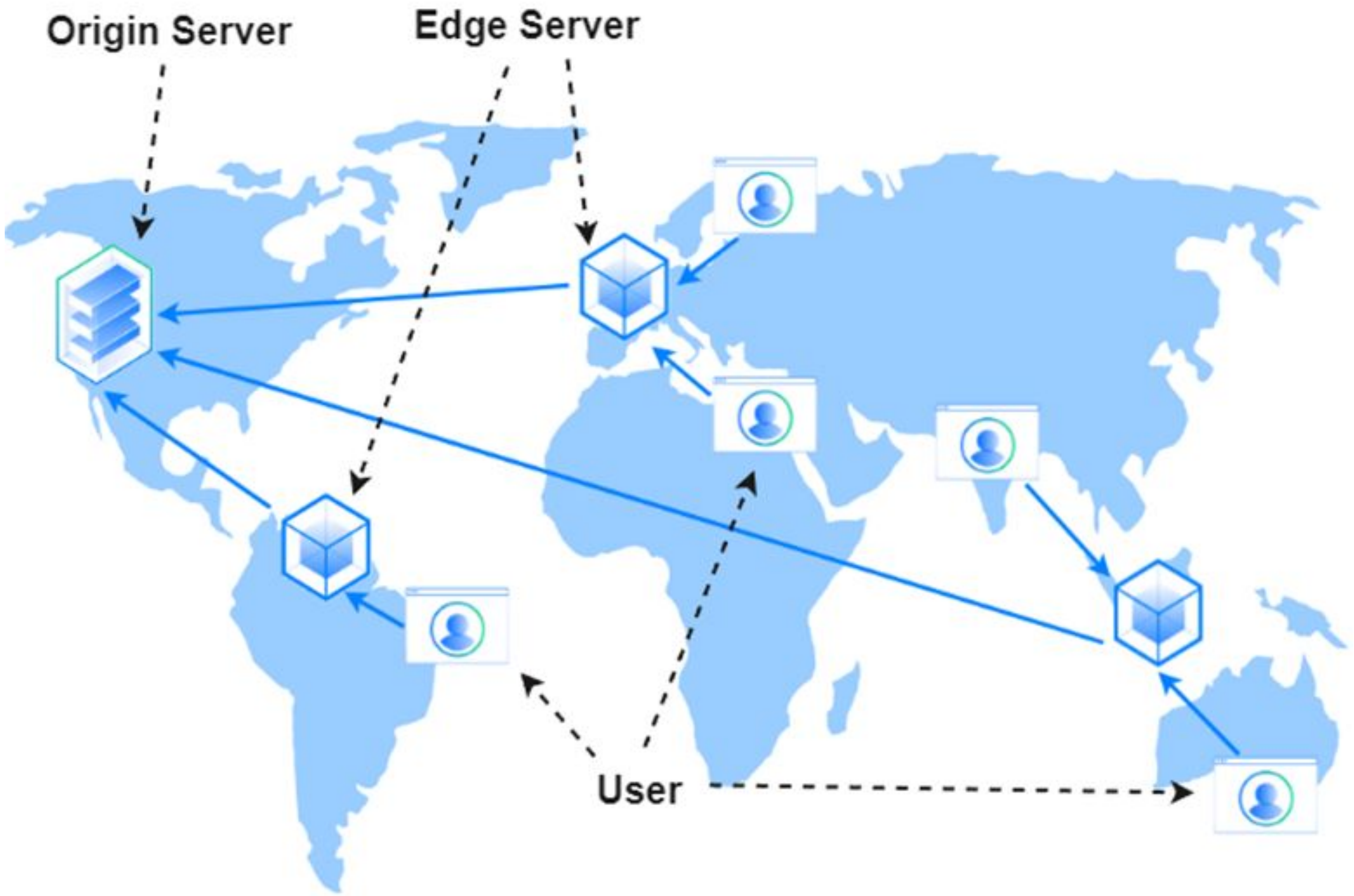


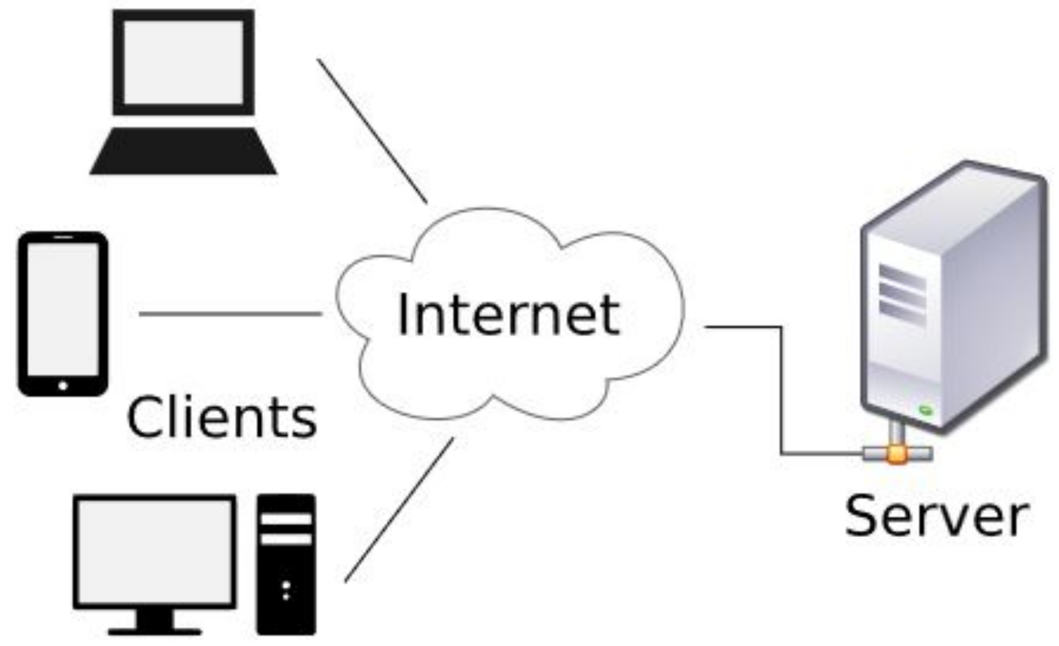
MOTIVATION

Where are we heading to?

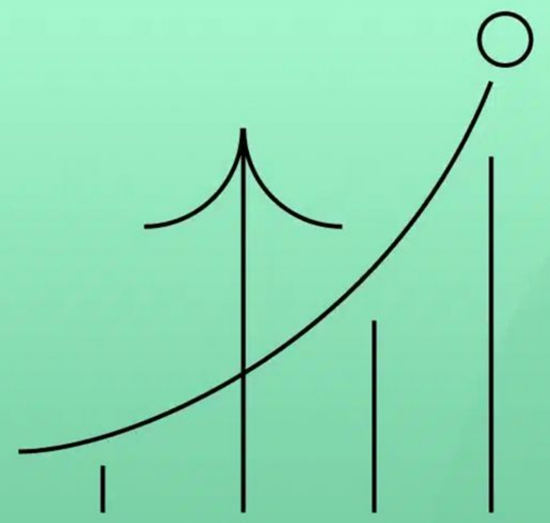


Content Delivery Network (CDN)

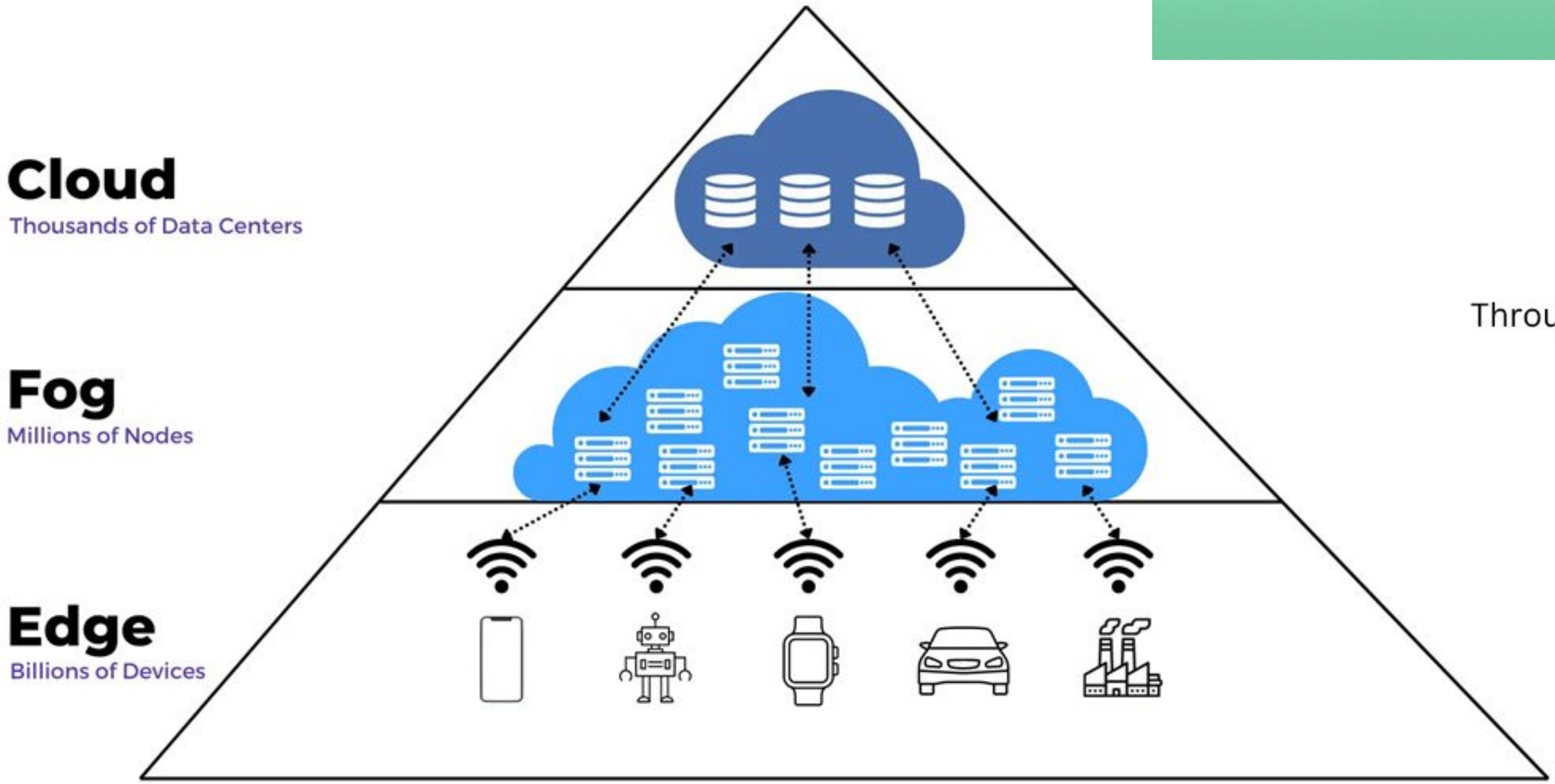


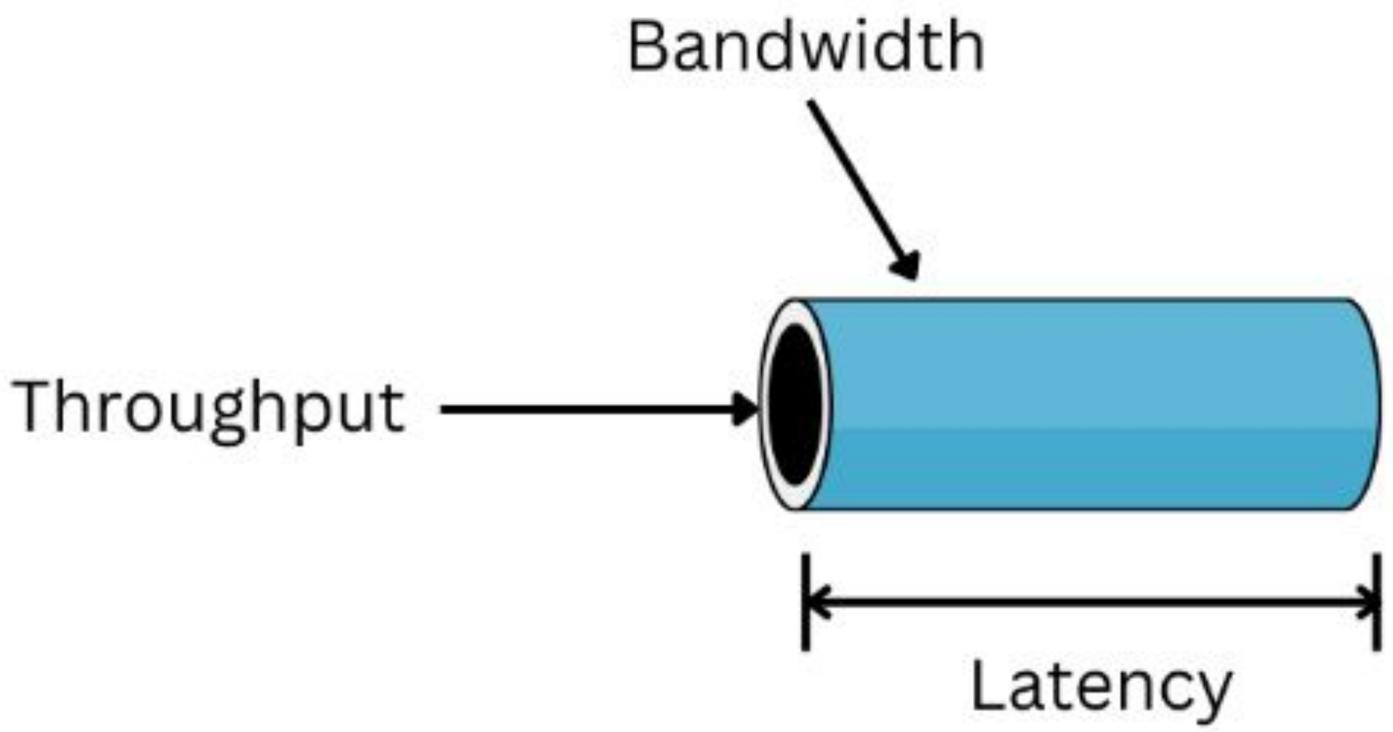


Many layers!

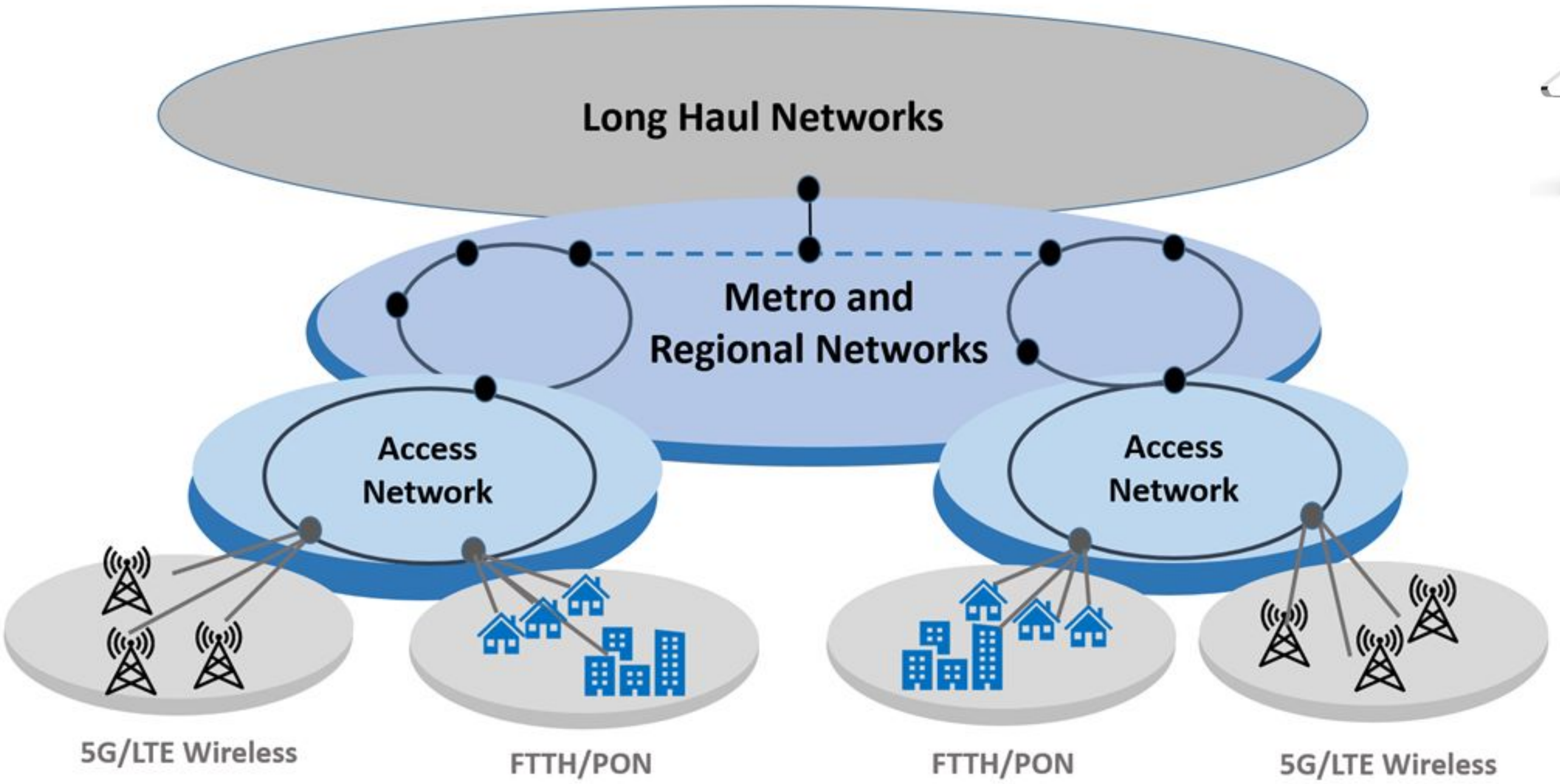


Application scalability

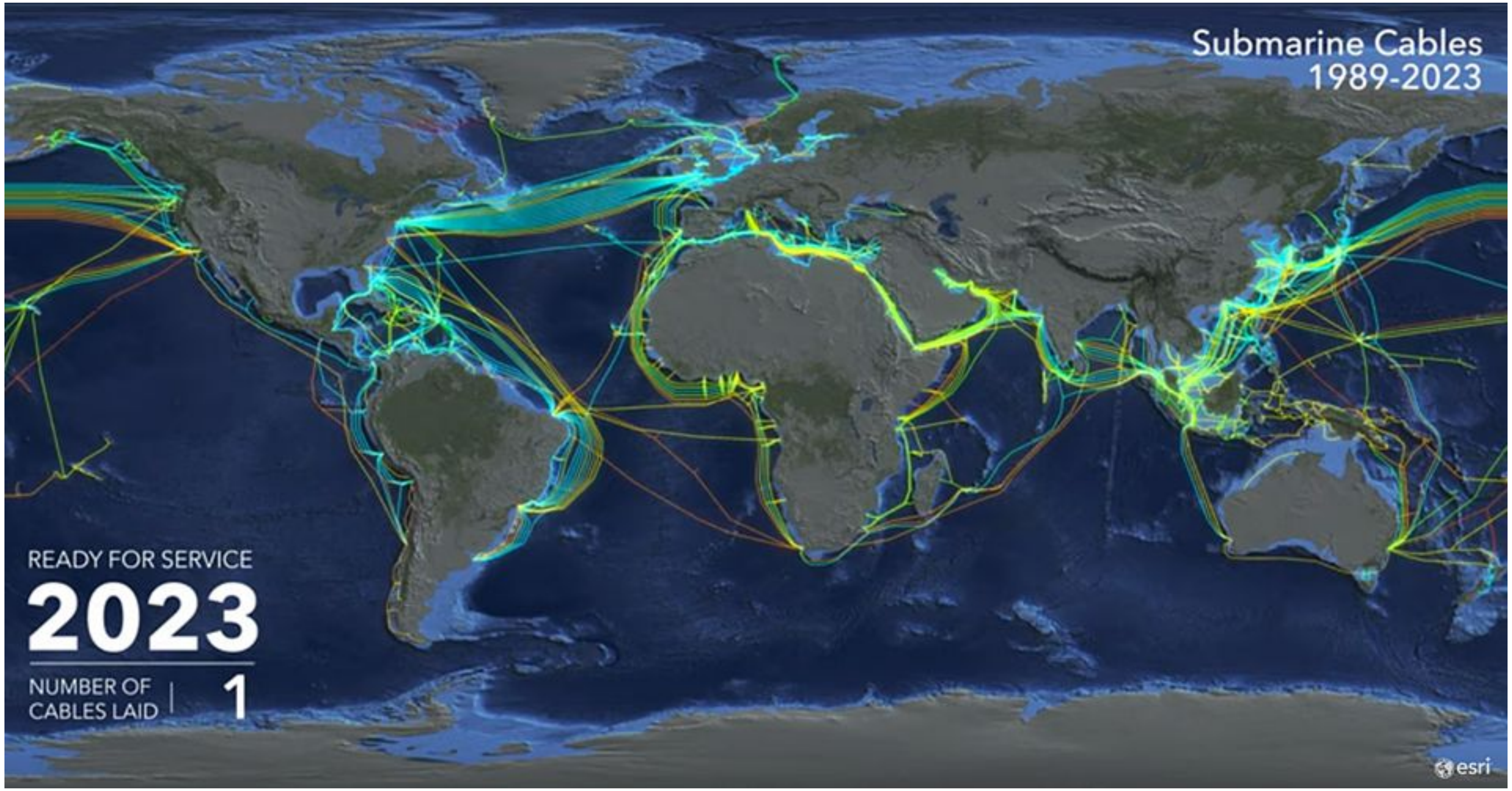


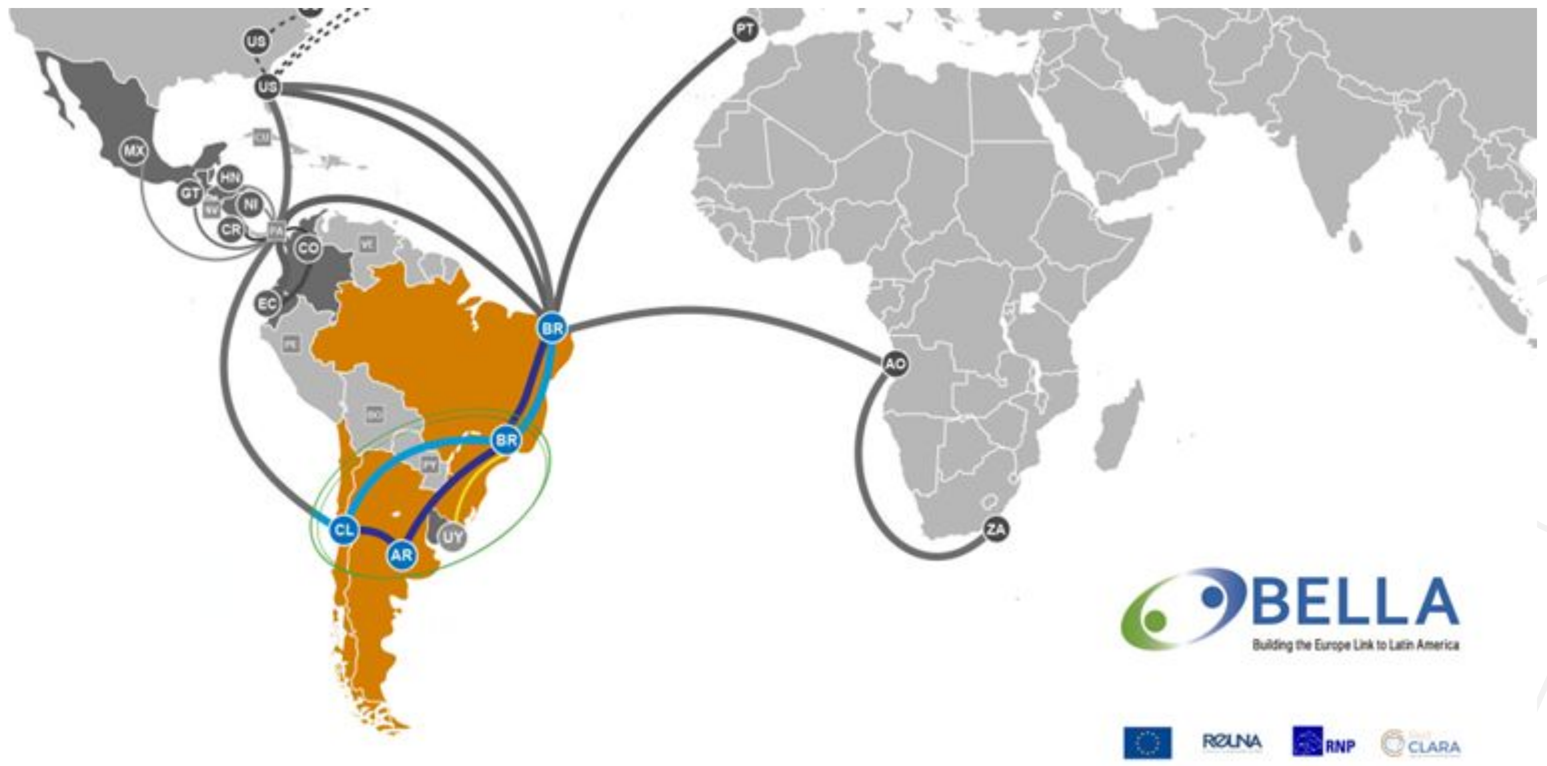


Wide Area Networks (WAN)

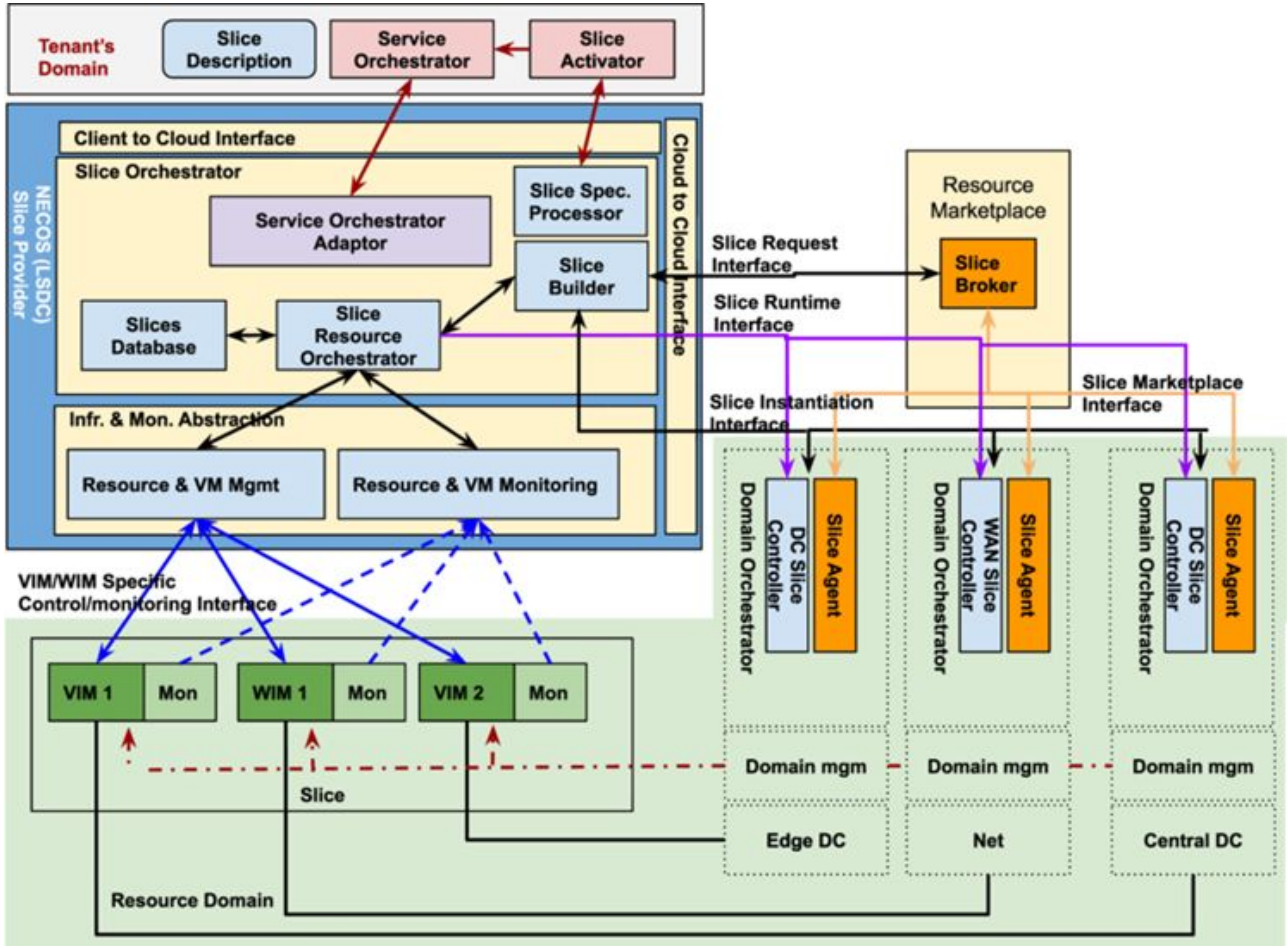


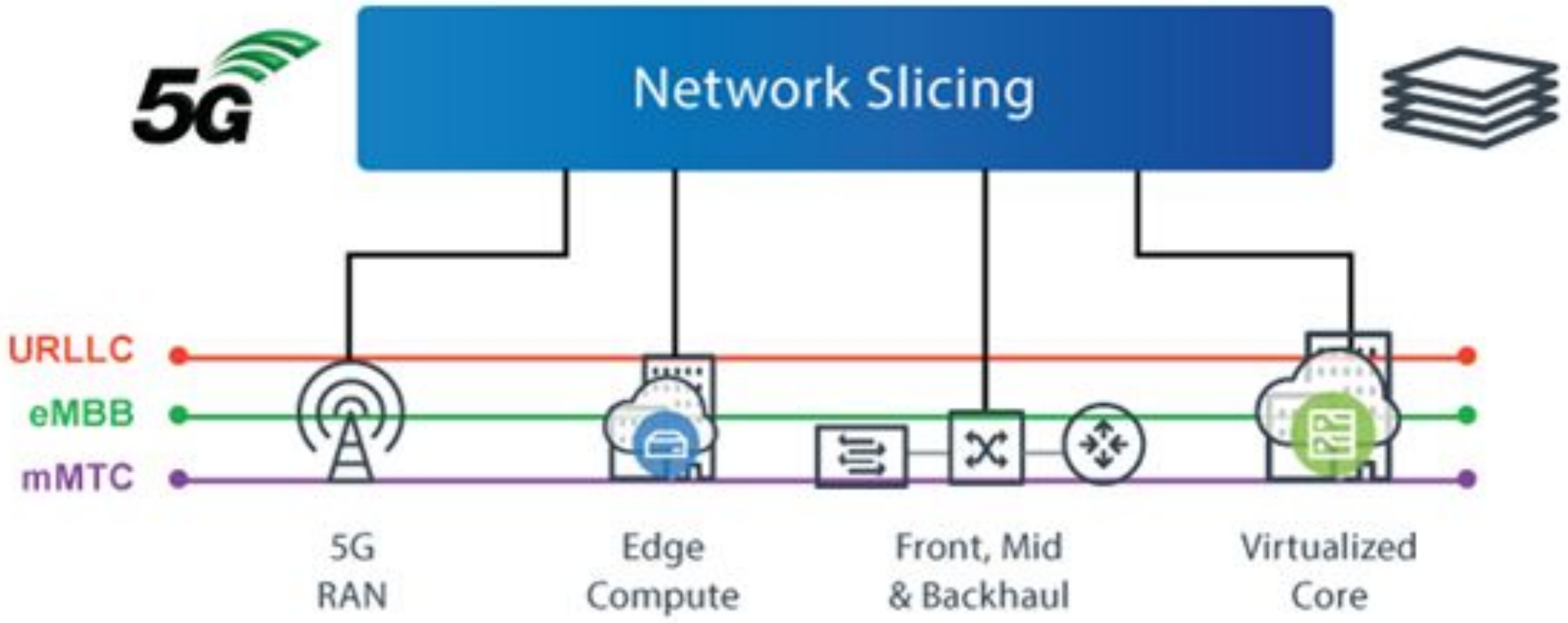
Submarine Cables





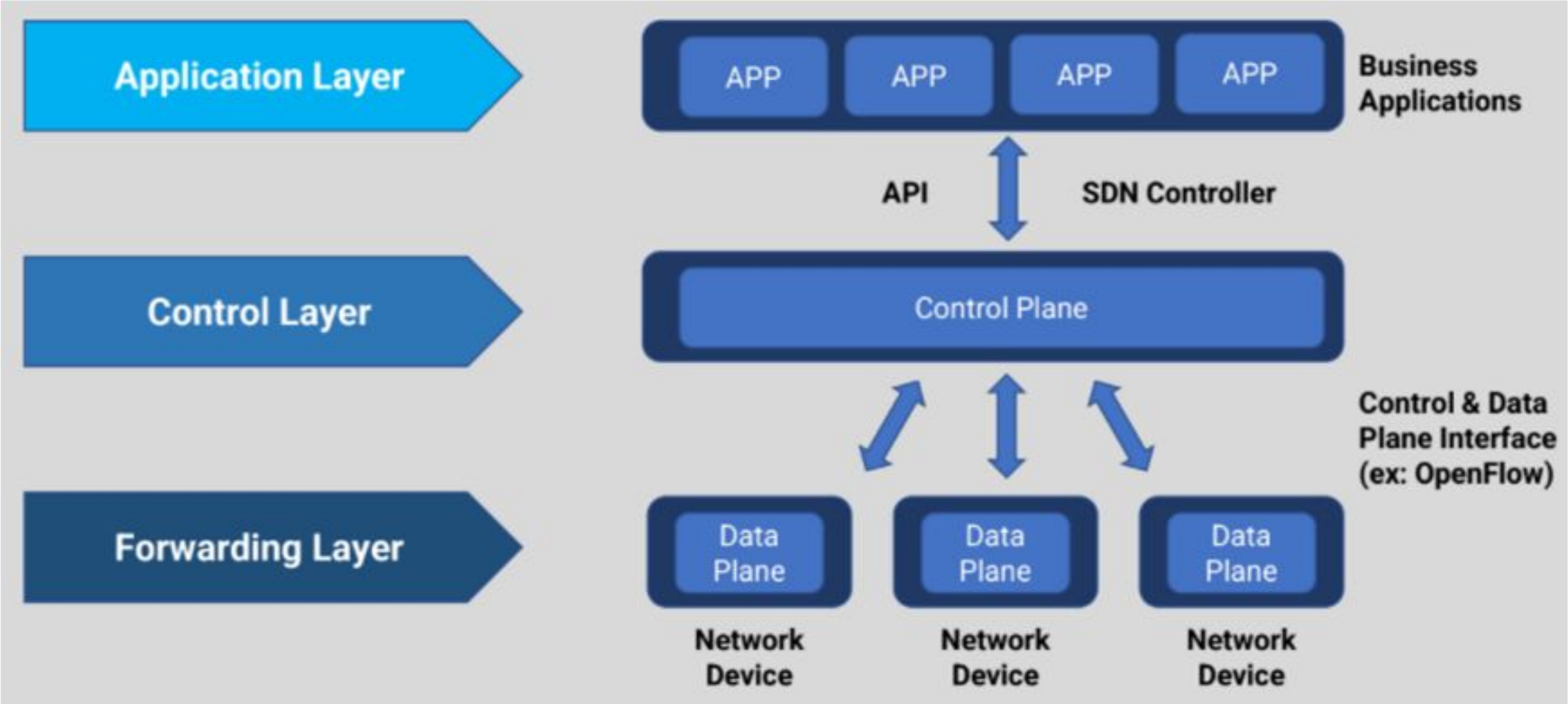
Cloud-Network Slicing



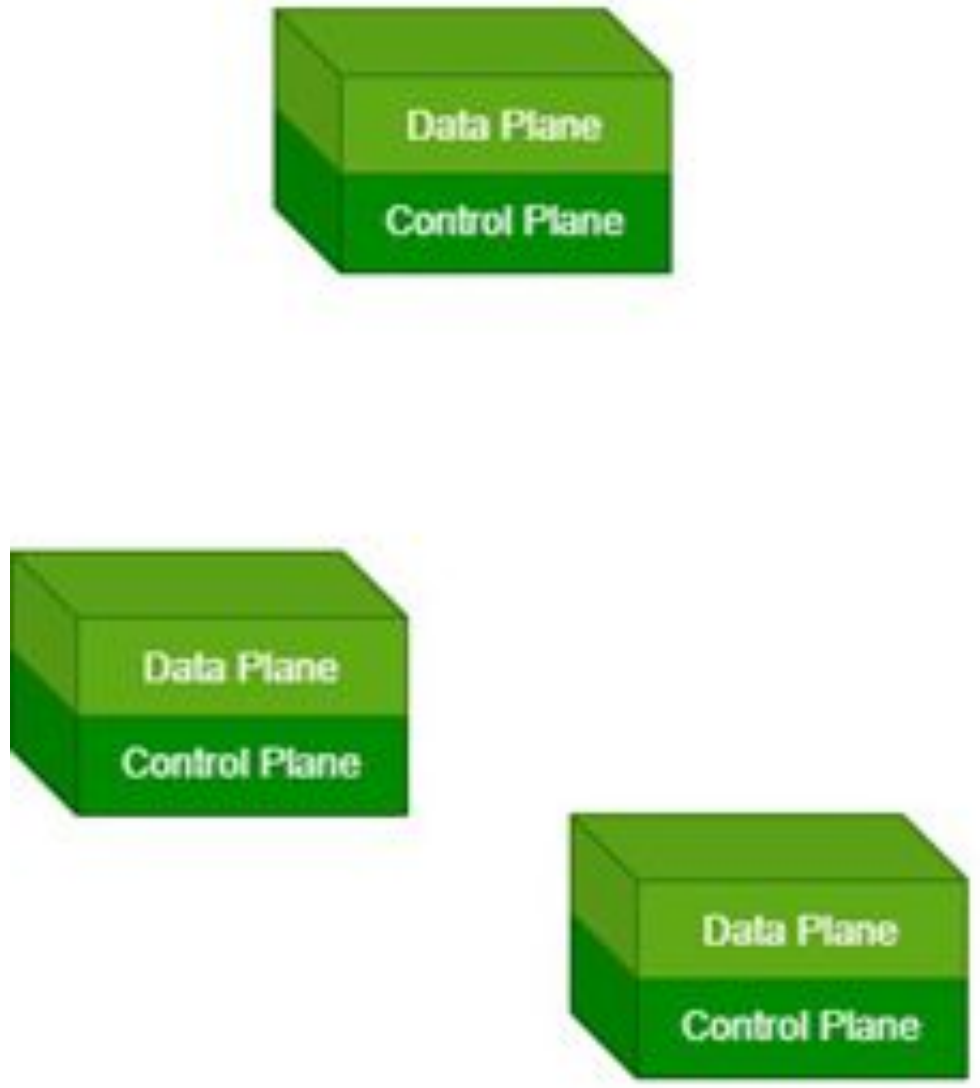


necos

Software-Defined Networking (SDN)



Traditional Network



Software-Defined Knowledge Plane

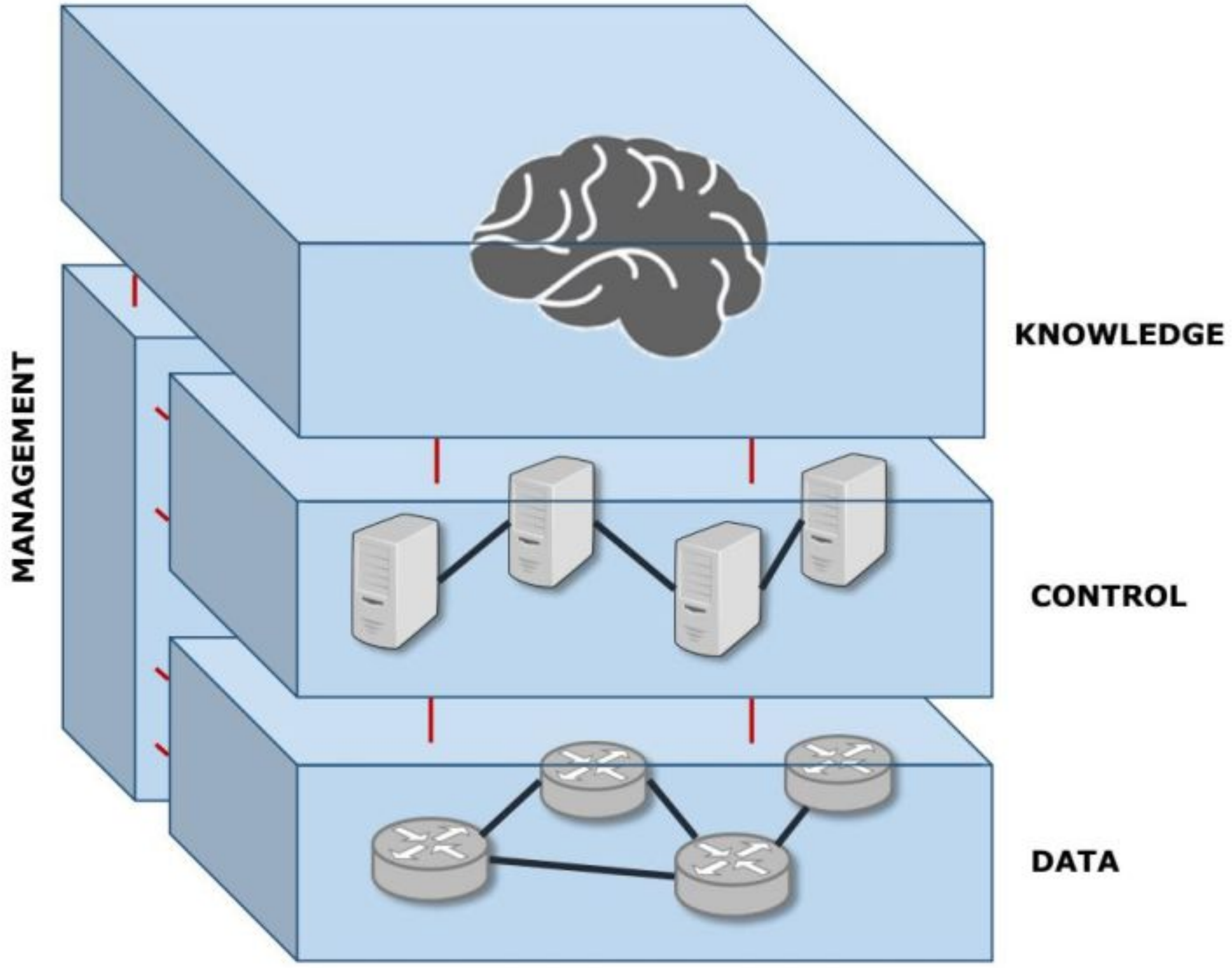


Fig. 1. KDN planes

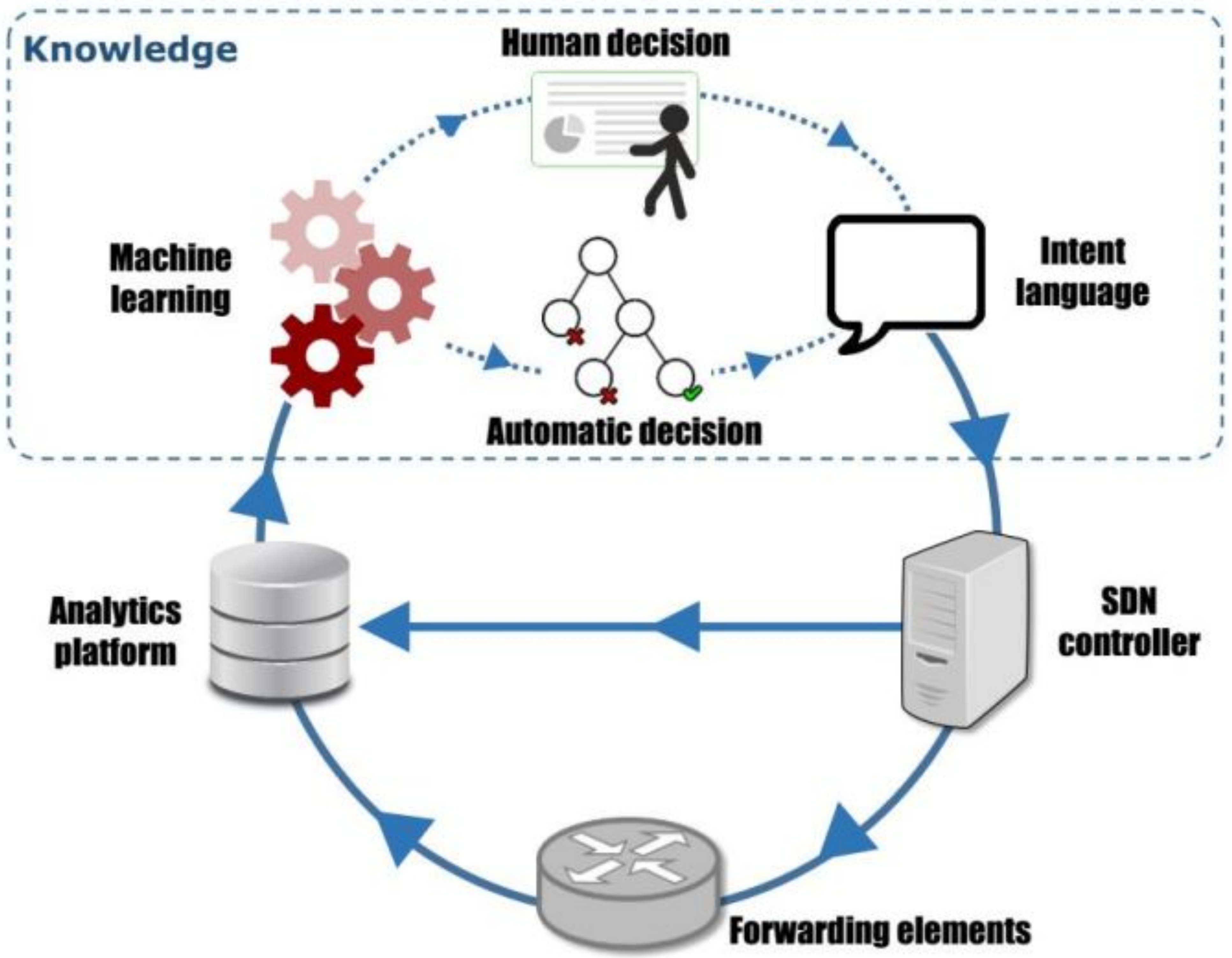
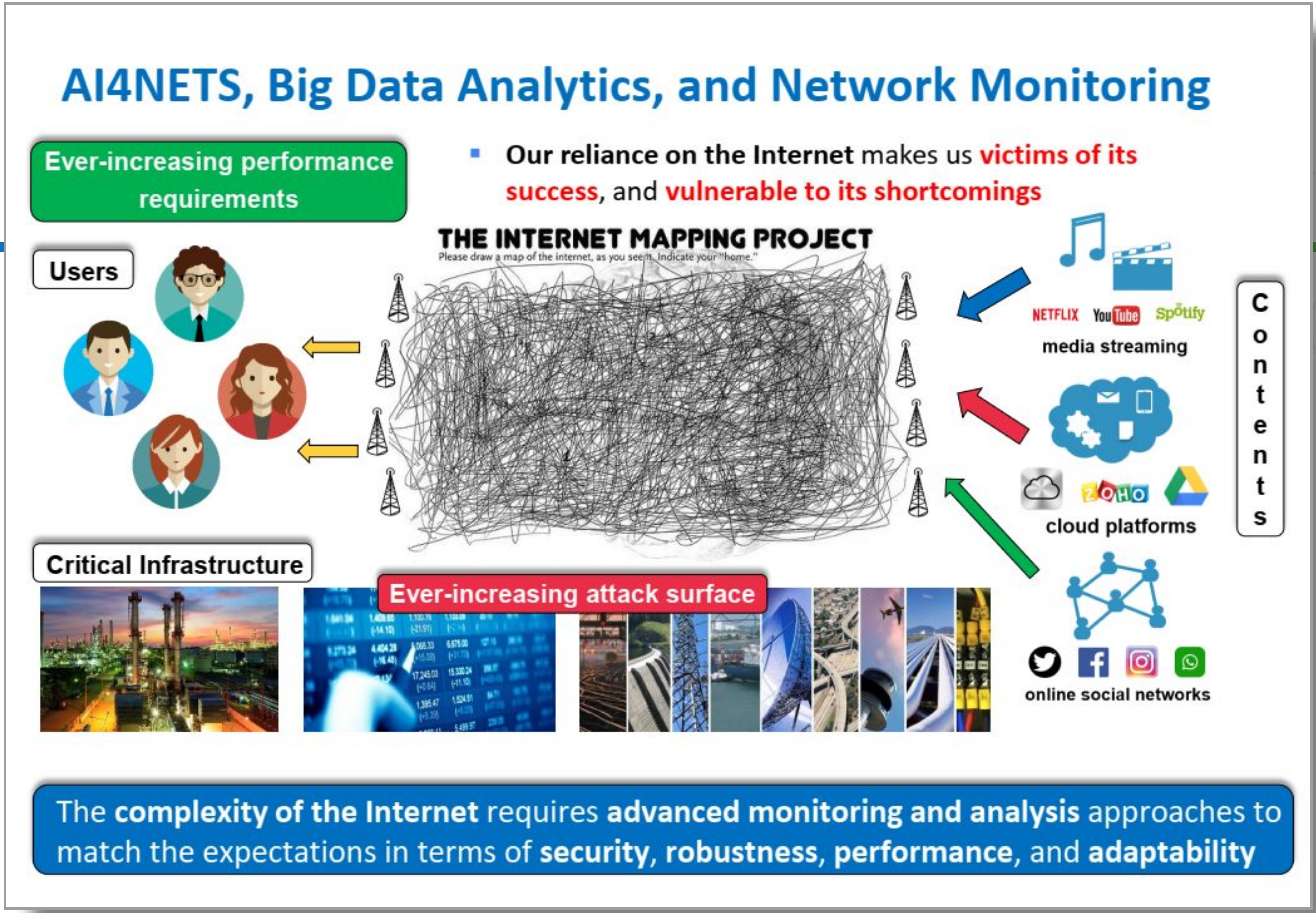


Fig. 2. KDN operational loop

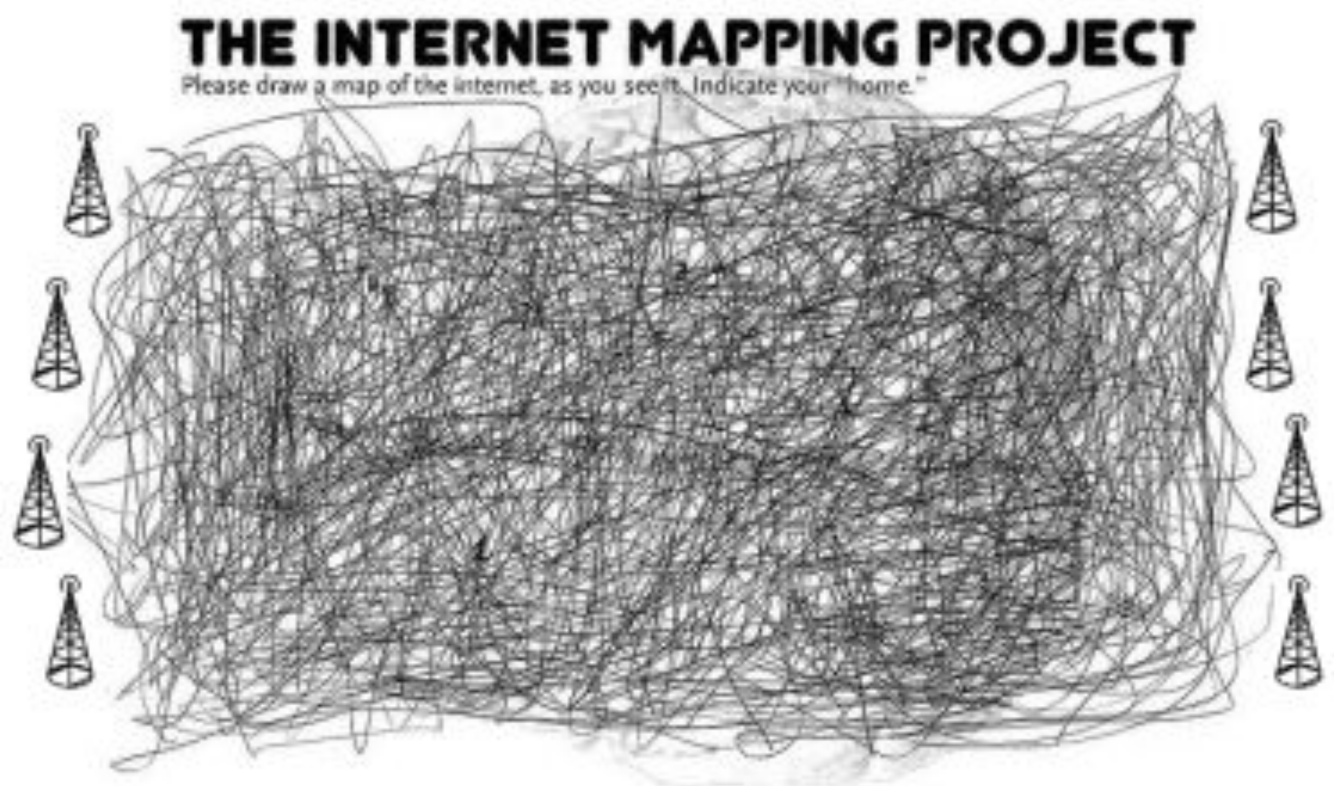
Source:
https://arxiv.org/pdf/1606.06222.pdf
http://knowledgedefinednetworking.org/



SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

AI4NETS, Big Data Analytics, and Network Monitoring



"The Internet is the first thing that humanity has built that humanity doesn't understand, the largest experiment in anarchy that we have ever had."

Eric Schmidt, former Google CEO

- **Hot Topic** in the agenda of **top Internet players**

Goals of AI4NETS

- A **radical change in the way we manage** communication **networks**, relying on **AI & Big-Data Analytics**
- The vision – turn the **Internet "transparent"** and **"liquid"**
- More **secure & robust**, better **performance**, **greener**, and **self-adaptive to end-user needs** in real-time

SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

Two decades of AI4NETS

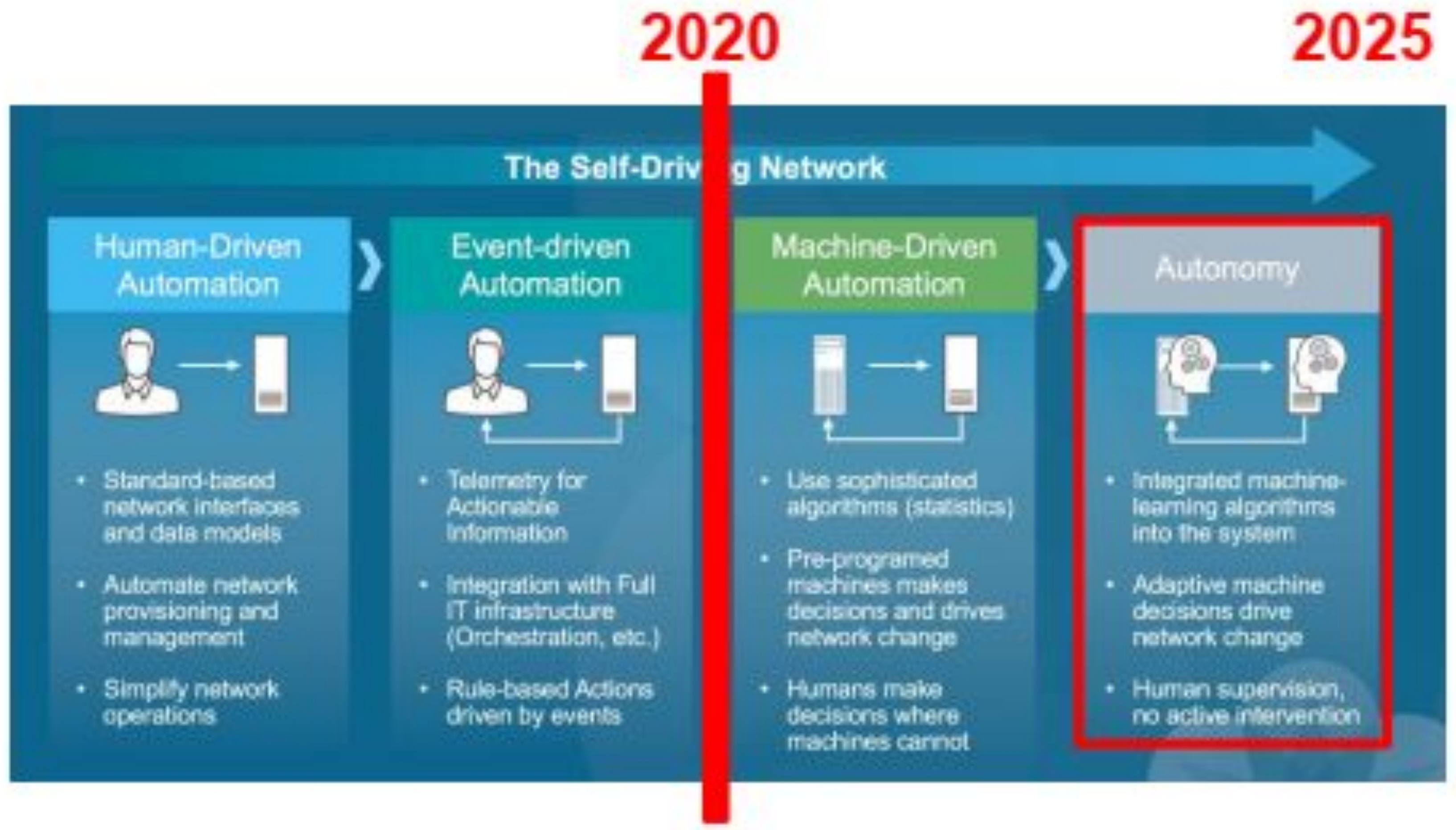
- **Cognitive Networking** (1998): networks with cognitive capabilities which could learn from past observations and behaviors, to better adapt to end-to-end requirements.

- Term **re-furbished along time**, referring to it as self-organizing networks, self-aware networks, **self-driving networks**, intelligent networks, etc.

- However, there is a **striking gap** between the **extensive academic research** and **the actual deployments of such AI-based systems in operational environments.**

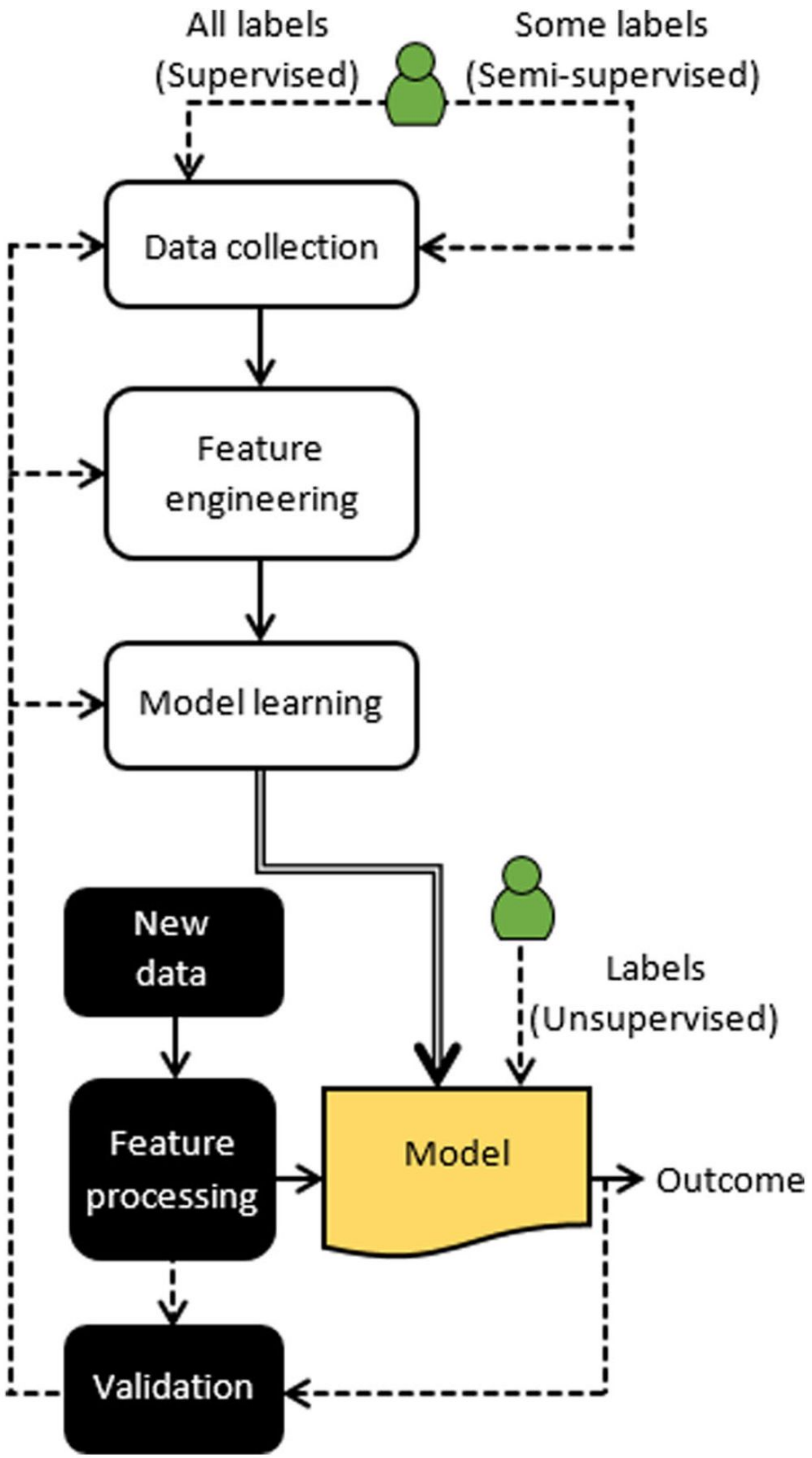
- **Why?** my take: there are still **many unsolved complex challenges** associated to the **analysis of Networking data through AI/ML.**

- **Hot Topic in the agenda of main Internet players:**
  - Network Operators
  - **Network Vendors (self-driving networks)**
  - Content Providers: the **Internet business of end-user engagement**



SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/



SMARTNESS

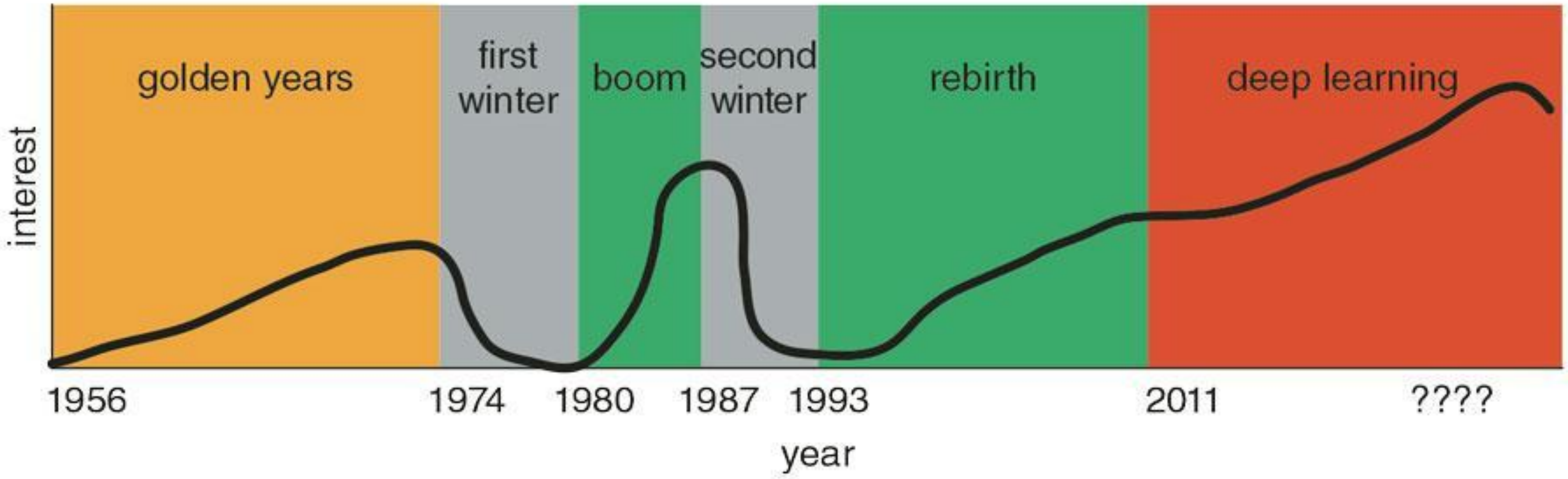
ML/AI Fundamentals & Application to Networking

ML 101

- **Fundamentos de IA/ML**

Historical Perspective

Artificial Intelligence, Machine Learning are not new areas (<u>and keep evolving</u>)

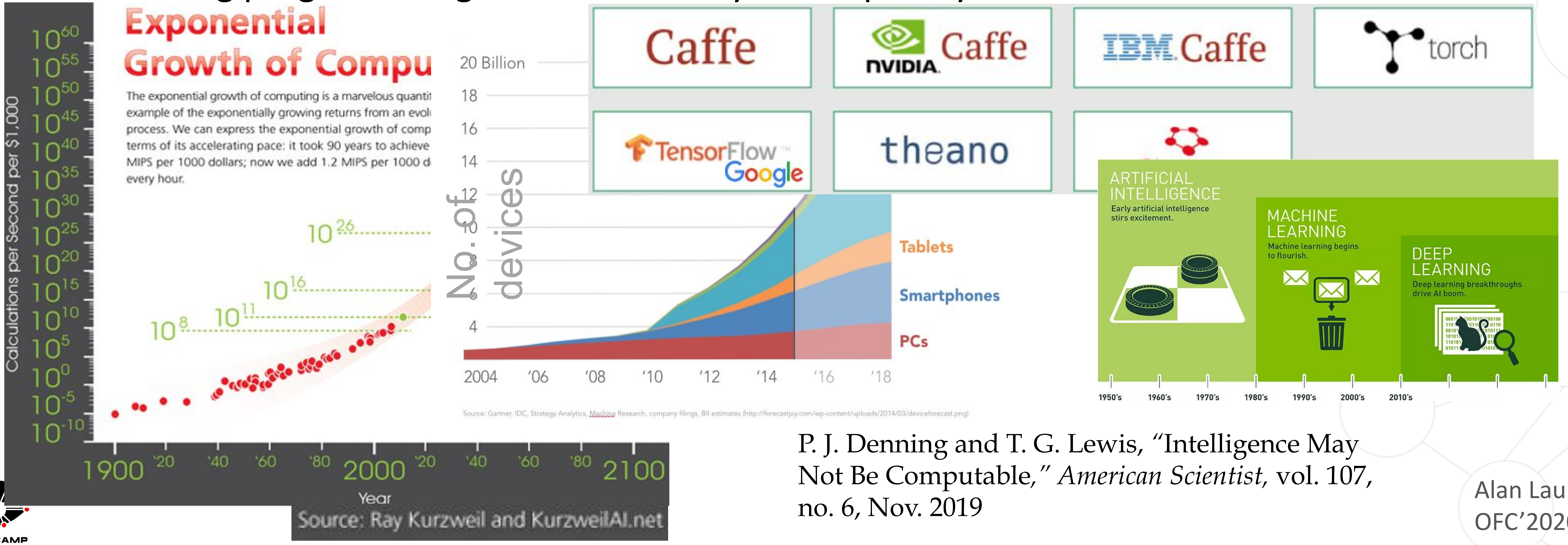


P. J. Denning and T. G. Lewis, "Intelligence May Not Be Computable," *American Scientist*, vol. 107, no. 6, Nov. 2019

What is Machine Learning?

How to explain the recent boom on ML?

- Increasing computational power
- Flood of available data – 20+ years of Internet worldwide
- Increasing support from industries
- Growing progress in algorithms & theory developed by researchers



P. J. Denning and T. G. Lewis, “Intelligence May Not Be Computable,” *American Scientist*, vol. 107, no. 6, Nov. 2019

Why AI/ML for networks?

- What are the “killer” apps for AI/ML? Potentially the ones “*where we are already implicitly employing machine learning, maybe badly*” (Schapira 2023)

Detection of <you name it> from data	Identifying existing **patterns** or **anomalies** in data, often in the present or recent past
Prediction of <you name it> from data	Forecasting **future outcomes** or **events** based on historical data

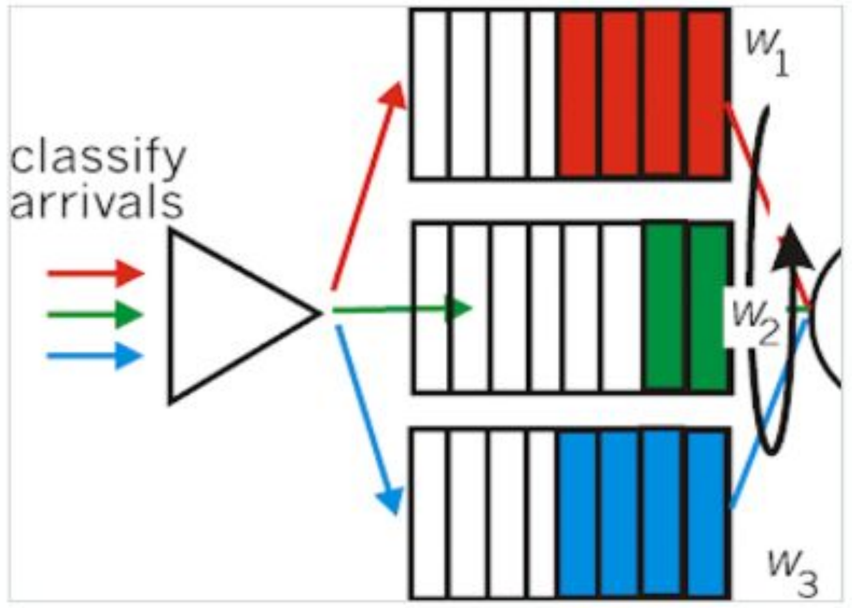
Source: Michael Schapira: AI for networking, and networking for AI. The Networking Channel, 2023. Available at https://www.youtube.com/watch?v=i6DvbfIUPSg

Why AI/ML for networks?

Carry out **intrusion detection** by looking at unexpected events

Source: https://towardsdatascience.com

Source: Kurose and Ross, 2020.

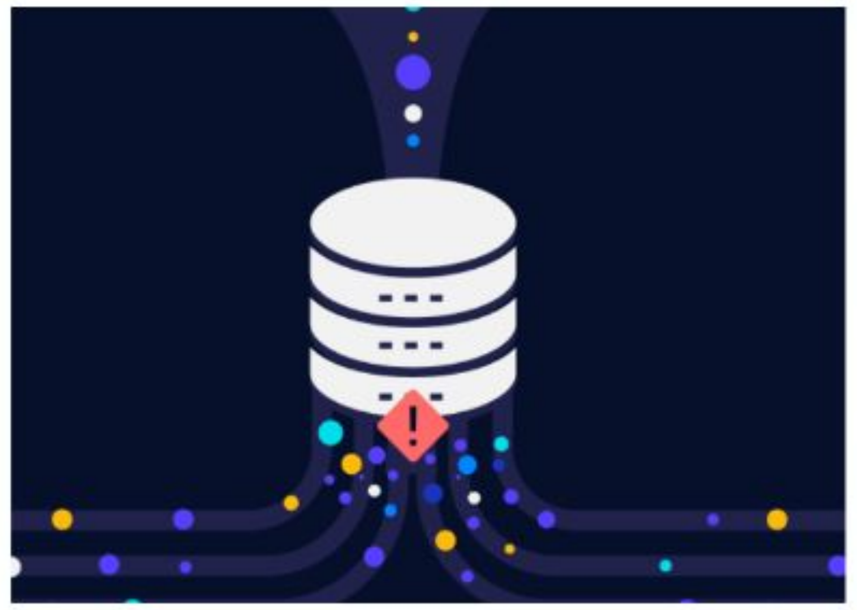


Perform **traffic classification** based on patterns existing on network packets

Execute **traffic engineering** based on predicted future traffic demands

Source: https://engineering.nyu.edu

Source: https://granulate.io



Run **congestion control** based on the prediction of the bottleneck bandwidth

Make **video streaming** decisions based on predicted download times of video chunks

Source: https://www.forbes.com

\+ **resource management** + **fault management** + **channel modeling** + ...

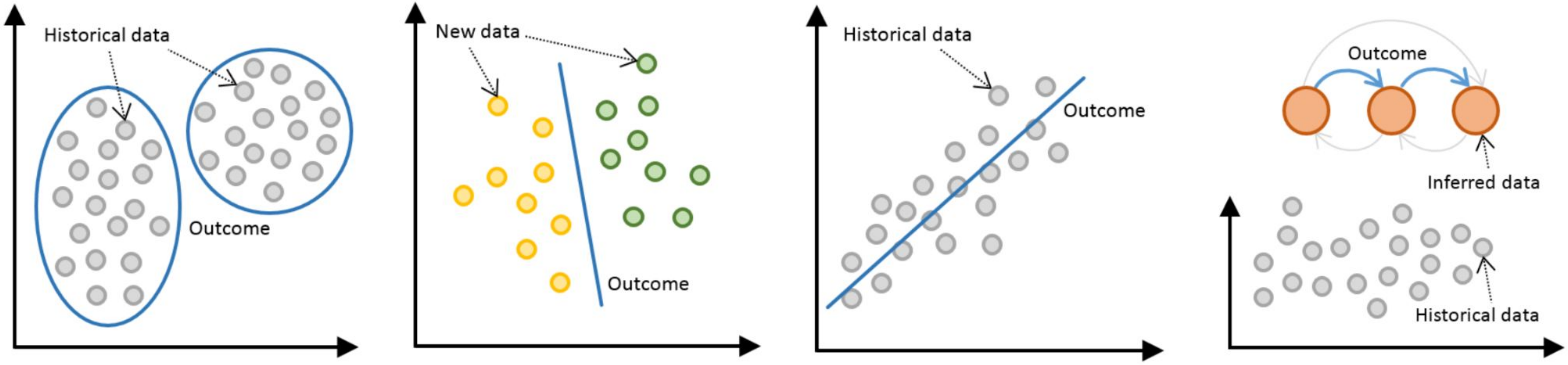
Source: Michael Schapira: AI for networking, and networking for AI. The Networking Channel, 2023. Available at https://www.youtube.com/watch?v=i6DvbfIUPSg

What is Machine Learning?

- We wish to give computers the **ability to learn**
  - Learning is the process of converting experience into expertise or knowledge
- Applied to:
  - Tasks that are too **complex** to program
  - Time-varying systems (**adaptivity**)

S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theoryto Algorithms. Wile y, 2014.

Boutaba et al., 2018)



(a) Clustering (b) Classification (c) Regression (d) Rule extraction

How AI/ML for networks?

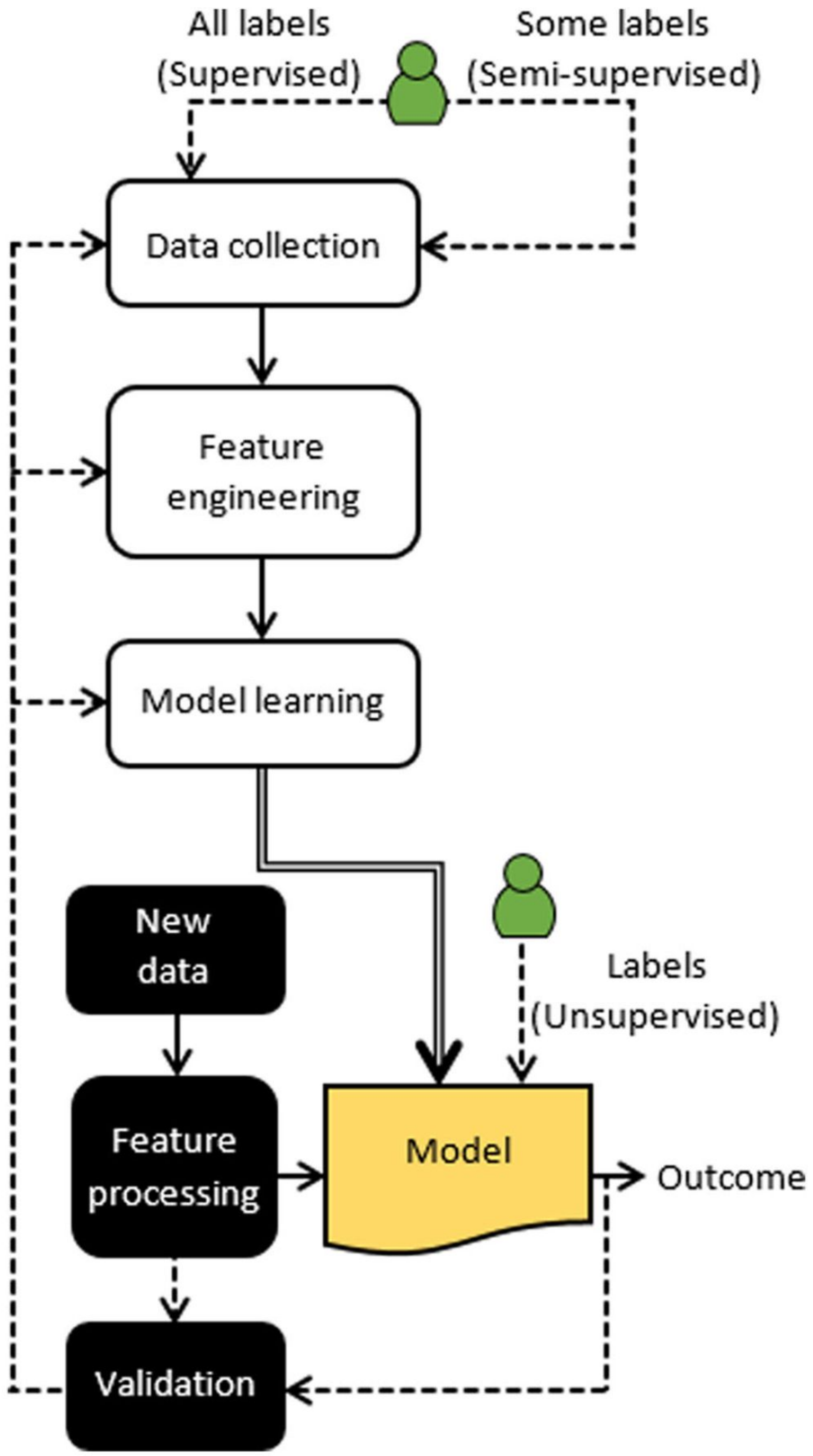
Main methods

Method	Characteristics
Supervised learning	• Two main phases: training and inference • Training data is labelled • SVM, Neural Network, Decision Tree, Ensemble Tree models, Naïve Bayes, K-Nearest Neighbor
Unsupervised learning	• Inputs of the learning algorithms are not labelled • Must learn patterns from the inputs • K-means, Isolation Forest, PCA, Autoencoder, SOM
Reinforcement learning	• Agent learns to achieve a goal by interacting with the environment (rewards and penalties) • Q-learning, SARSA

Source: Ricardo Parizotto, Bruno Loureiro Coelho, Diego Cardoso Nunes, Israat Haque, and Alberto Schaeffer-Filho: Offloading Machine Learning to Programmable Data Planes: A Systematic Survey. ACM Comput. Surv. 56, 1, Article 18 (2024).

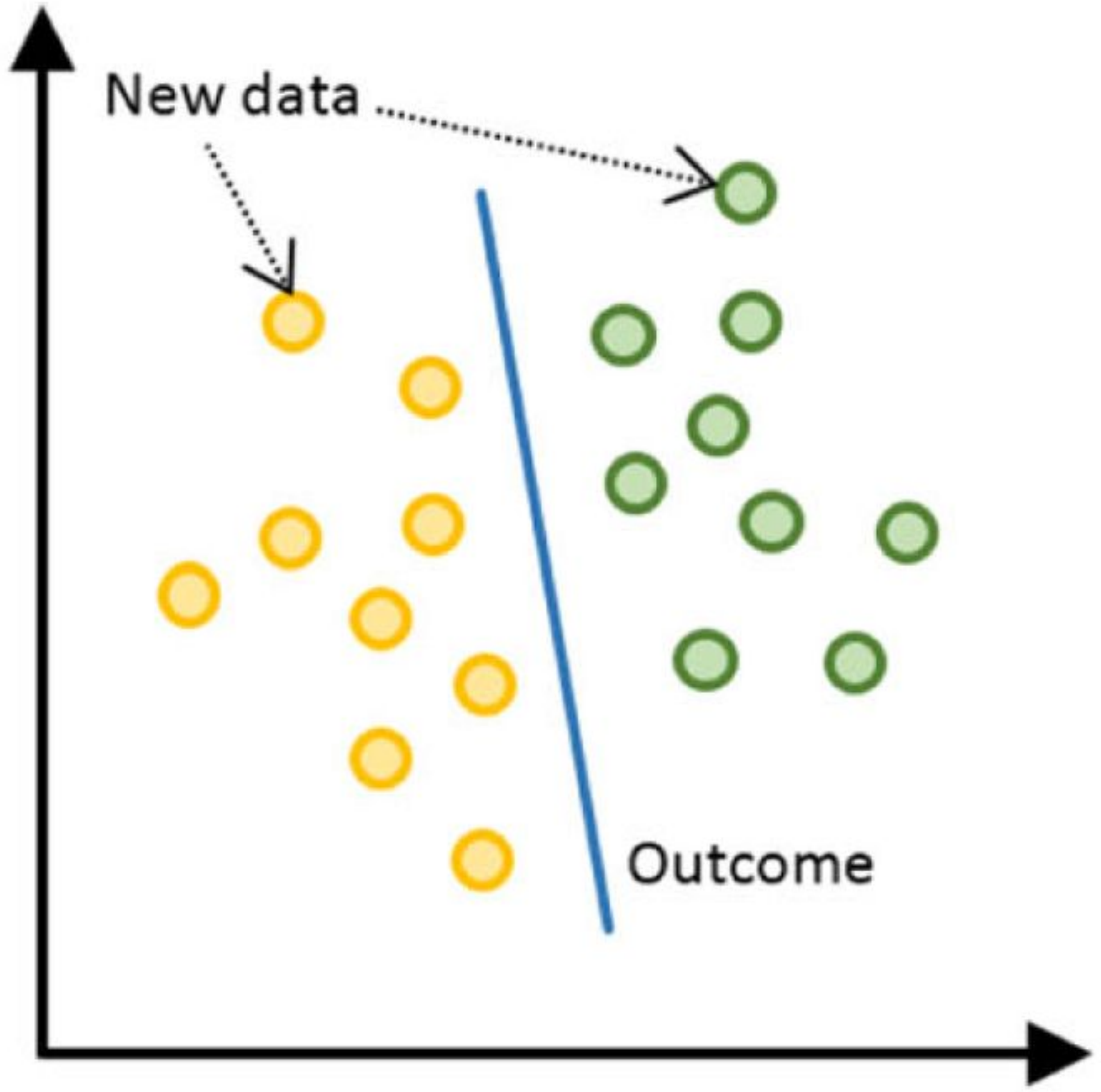
Types of ML - Intelligence (1/3)

Supervised Learning

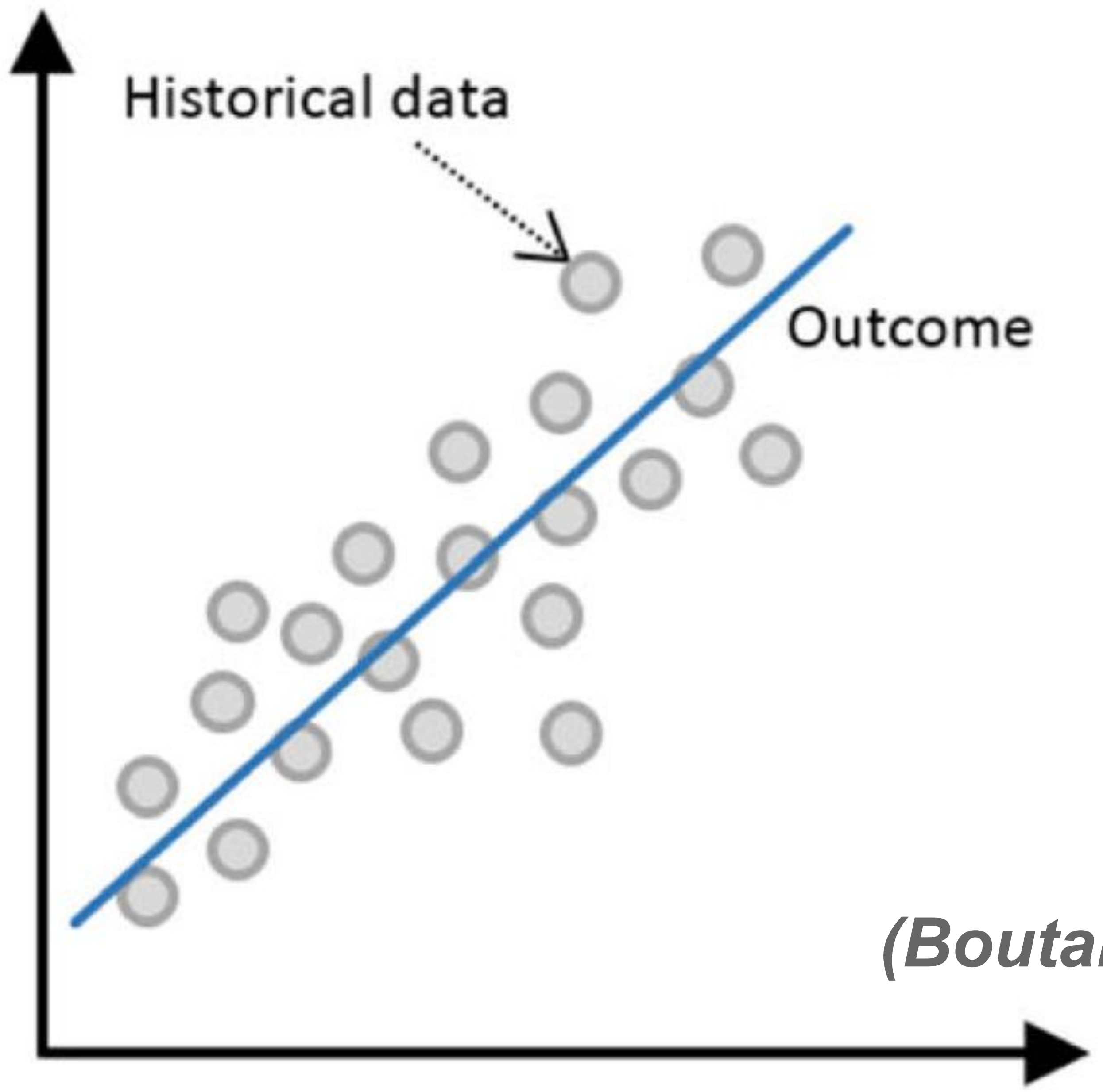


Learning approach	Training dataset	Problems aimed
Supervised	Labeled	Classification and regression
Semi-supervised	Incomplete labels	

Classification



Regression



(Boutaba et al., 2018)

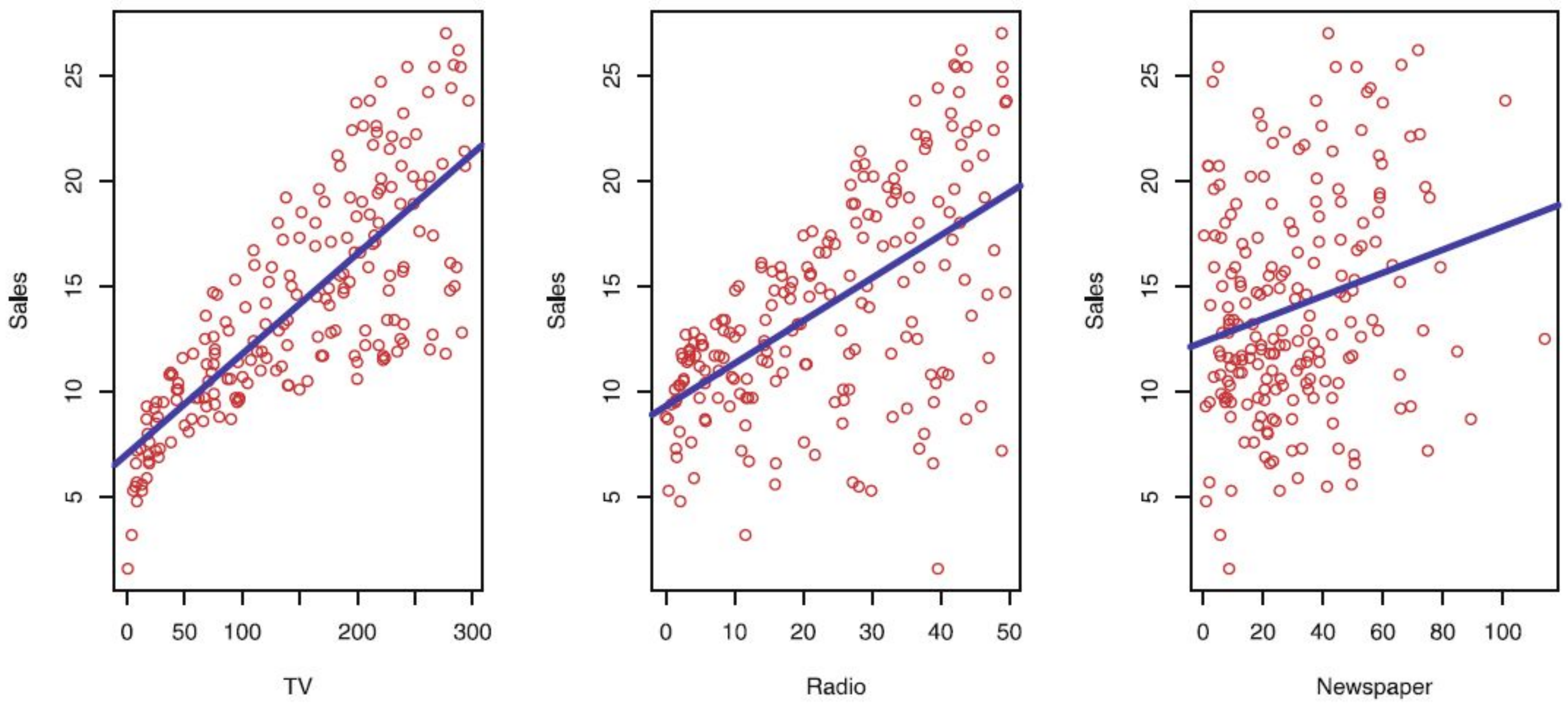
Types of ML - Intelligence (1/3)

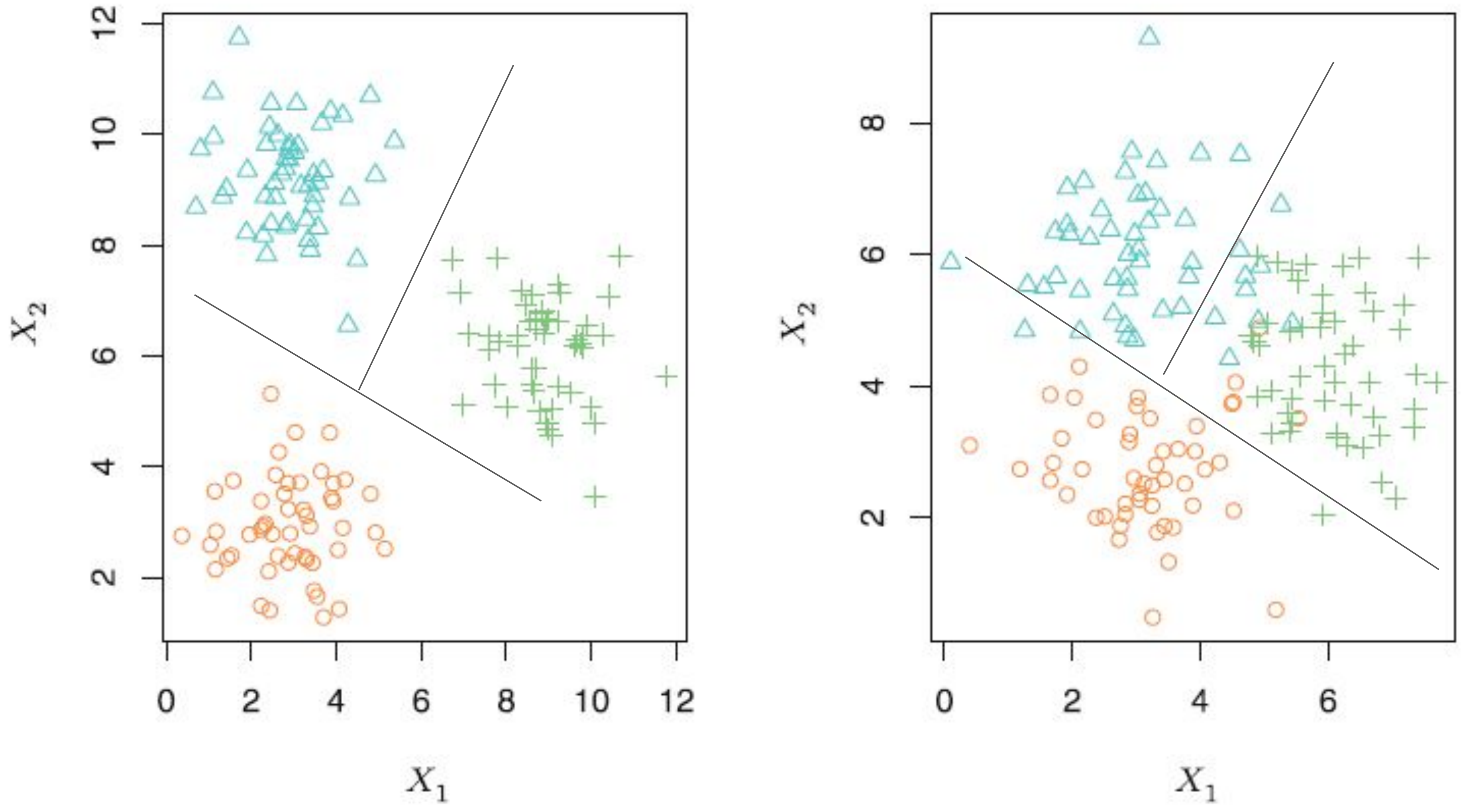
Supervised Learning

Detecting anomalies, intrusion detection, traffic classification.

For example, using labeled data

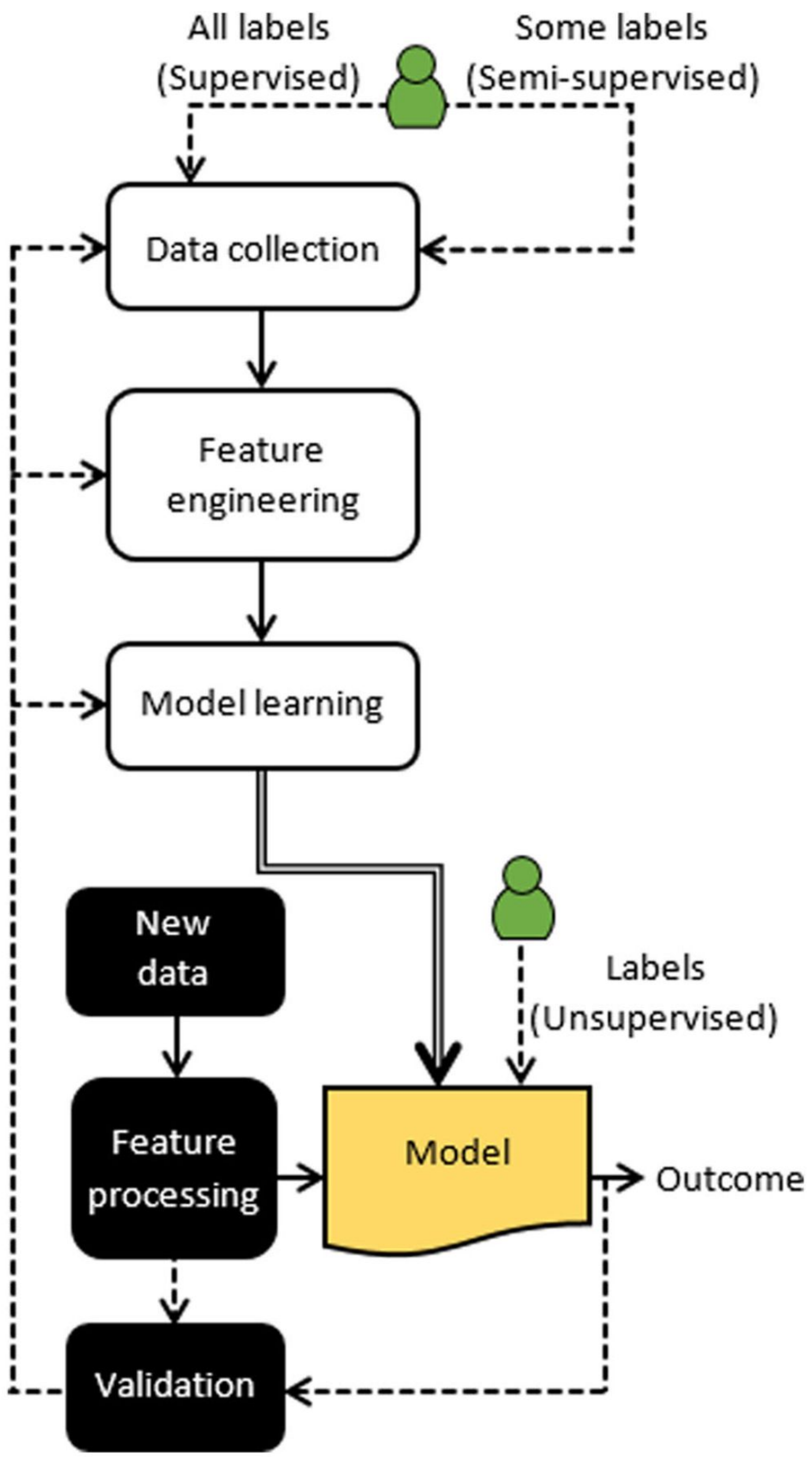
- to predict Mean Opinion Score (MOS) – Classification
- or forecast Quality of Experience (QoE) – Regression





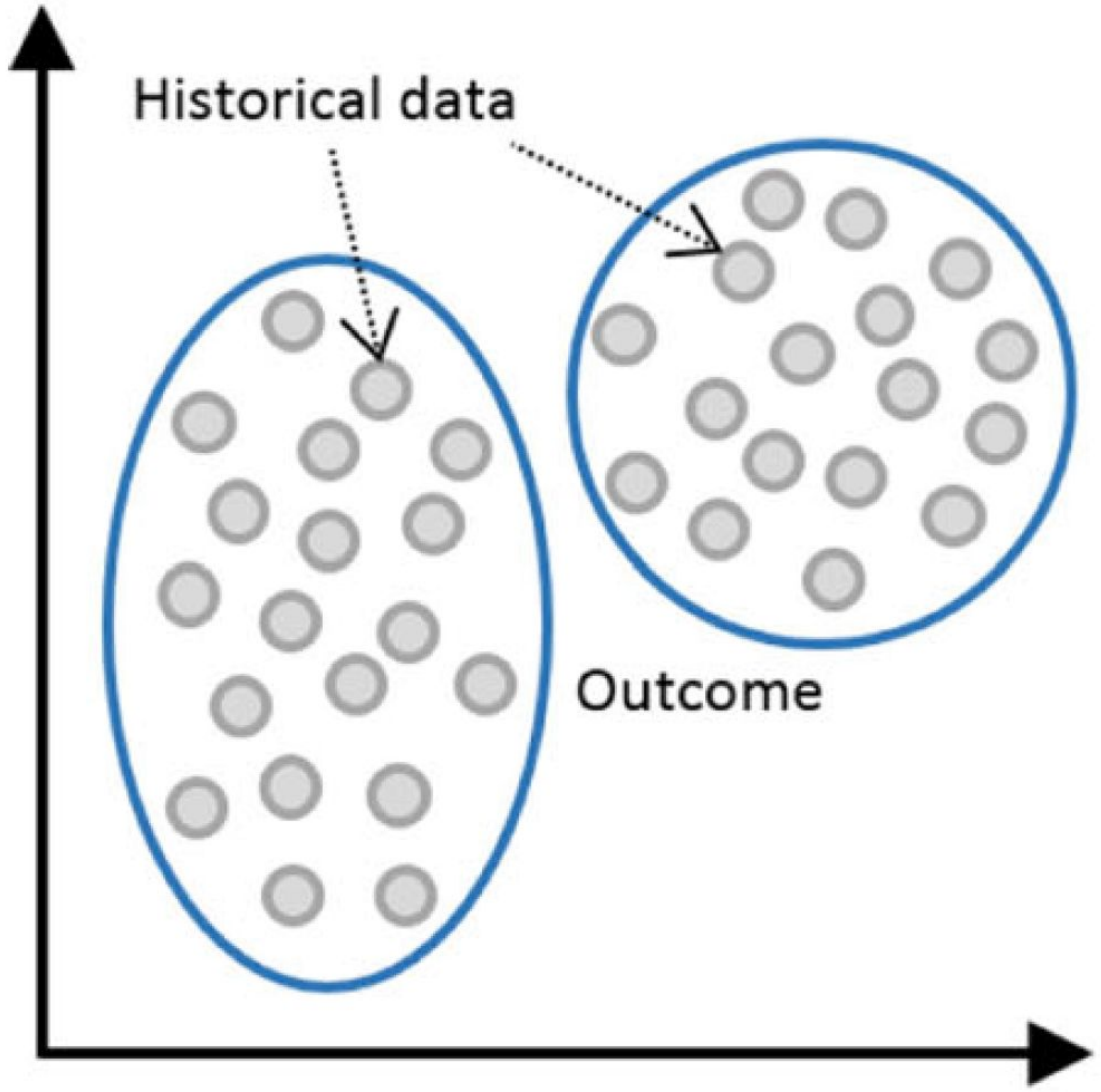
Types of ML - Intelligence (2/3)

Unsupervised Learning



Learning approach	Training dataset	Problems aimed
Unsupervised	Unlabeled	Clustering

Clustering



(Boutaba et al., 2018)

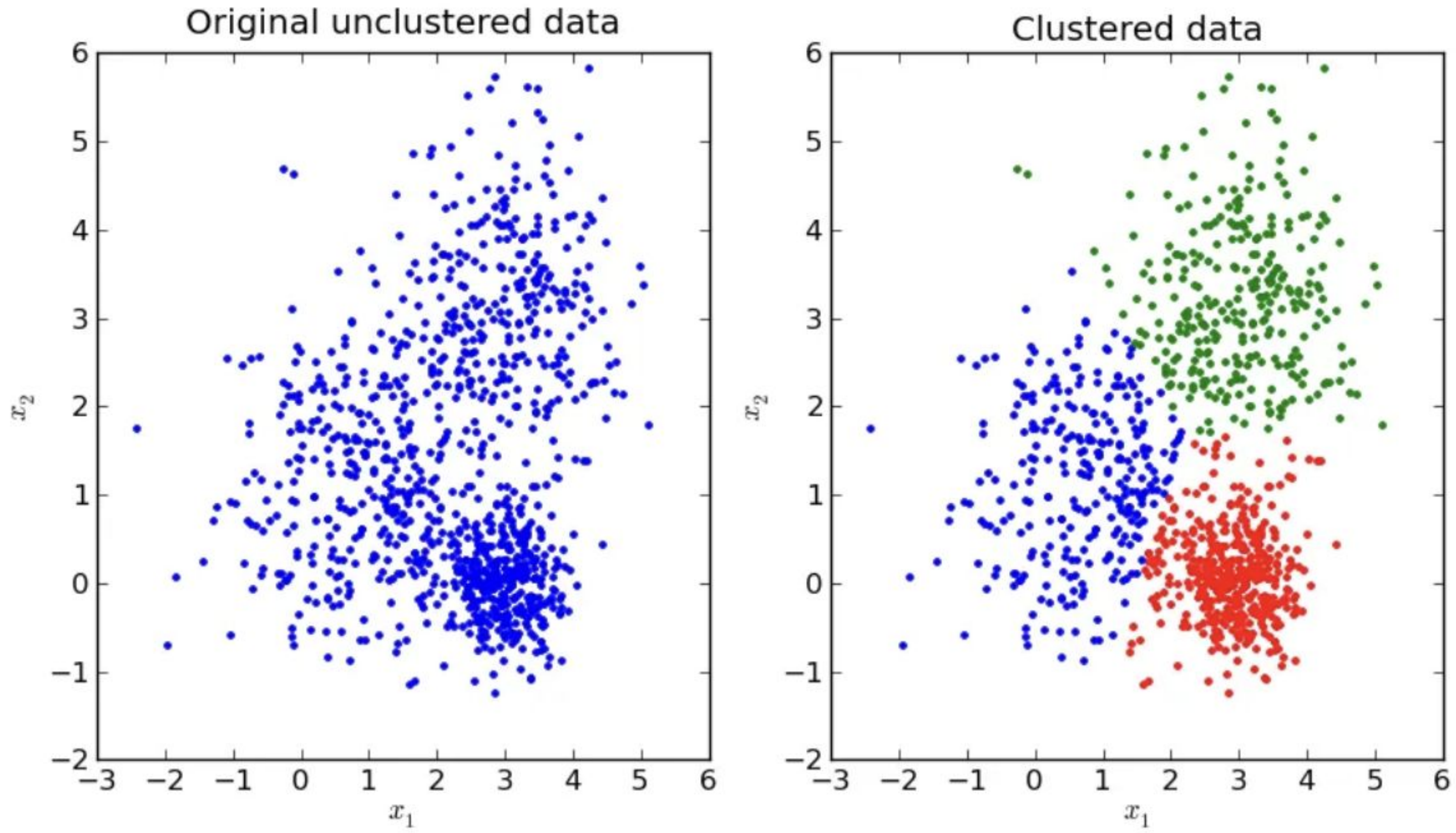
Types of ML - Intelligence (2/3)

Unsupervised Learning

Network clustering, identifying patterns in user behavior.

For example, Grouping similar network traffic

- for improved load balancing / routing / network function chains

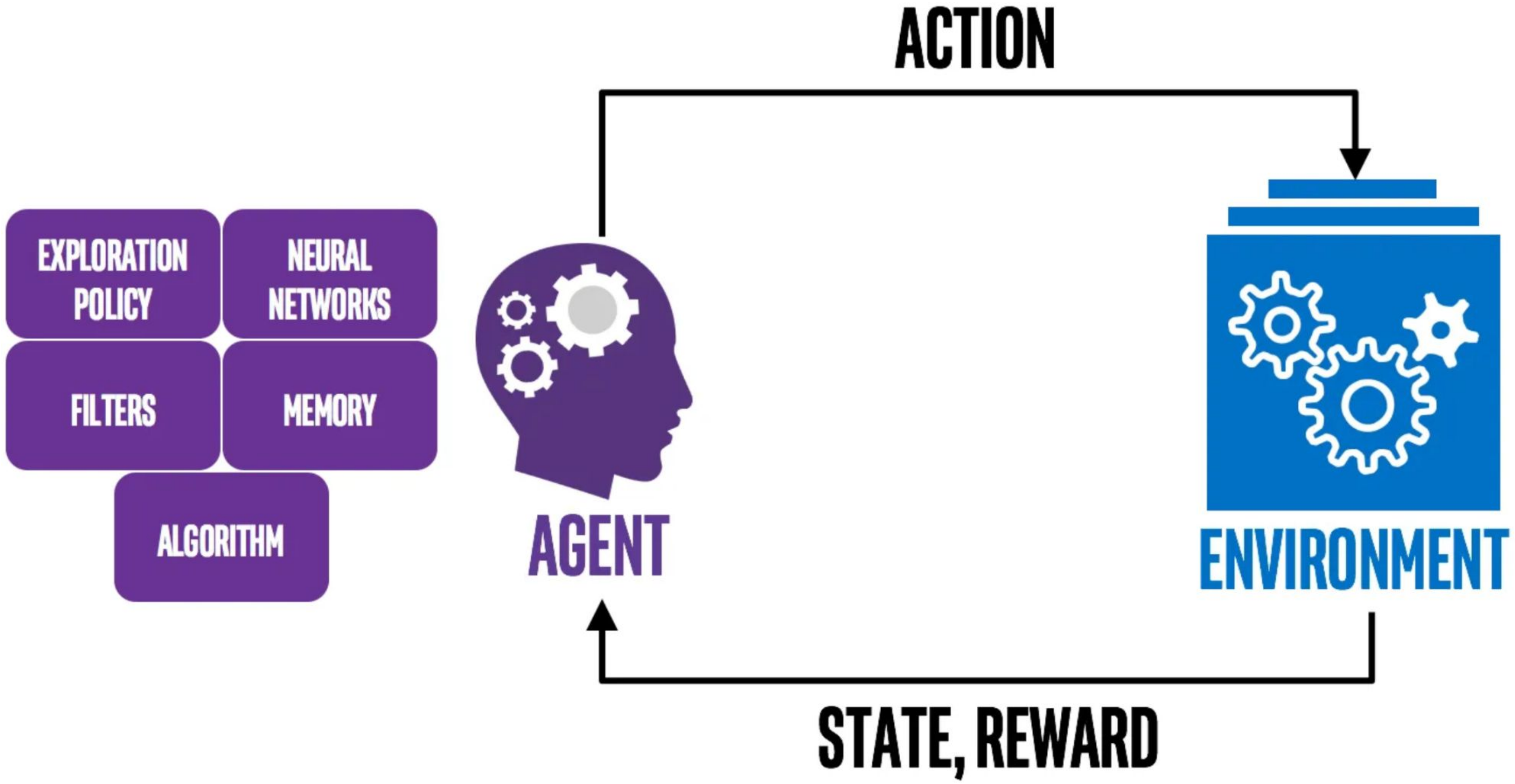


Types of ML - Intelligence (3/3)

Reinforcement Learning

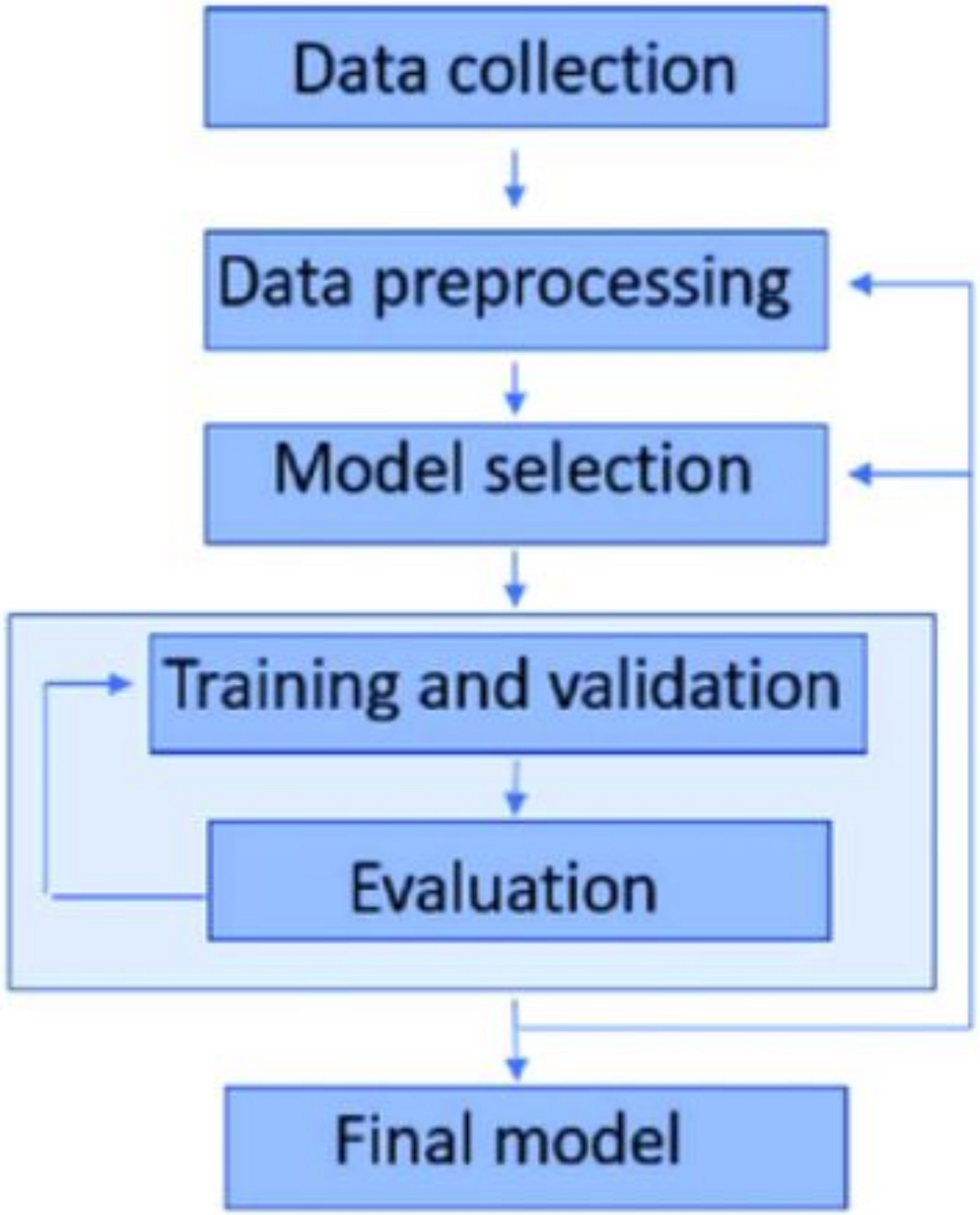
For example,

- Optimizing Flow Routing, Dynamic Routing, Adaptive Traffic Management, Resource Allocation, etc.



Machine Learning Pipeline

- Data Collection (**gathering data**)
- Data Preprocessing (**cleaning, transforming**)
- Data Splitting (**train/test split**)
- Model Selection (**algorithm choice**)
- Model Training (**fit model**)
- Model Evaluation (**assess performance**)
- Model Testing (**final test**)
- Model Deployment (**production use**)



ML: discipline vs tool to solve complex problems

ML in TMA **IS NOT** about trying different algorithms to obtain better results.

To build a solid house on your own, you need to know about architecture, as well as about the intrinsic characteristics of the construction toolbox. . .

Two commonly arising problems when coupling ML and Networking:

(I) You have to understand the problem:

- Even a ML expert fails to achieve a good networking solution if he neither knows the good descriptors nor understands the problem (e.g., try to classify flows using only port numbers).
- Keep the scope narrow, to better understand the overall process (i.e., from selecting features to evaluation and conclusions).
- The solution must be meaningful in practical terms (e.g., predicting QoE from descriptors that can't be controlled is pretty useless for QoE management).

SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

ML: discipline vs tool to solve complex problems

ML in TMA **IS NOT** about trying different algorithms to obtain better results.

To build a solid house on your own, you need to know about architecture, as well as about the intrinsic characteristics of the construction toolbox. . .

Two commonly arising problems when coupling ML and Networking:

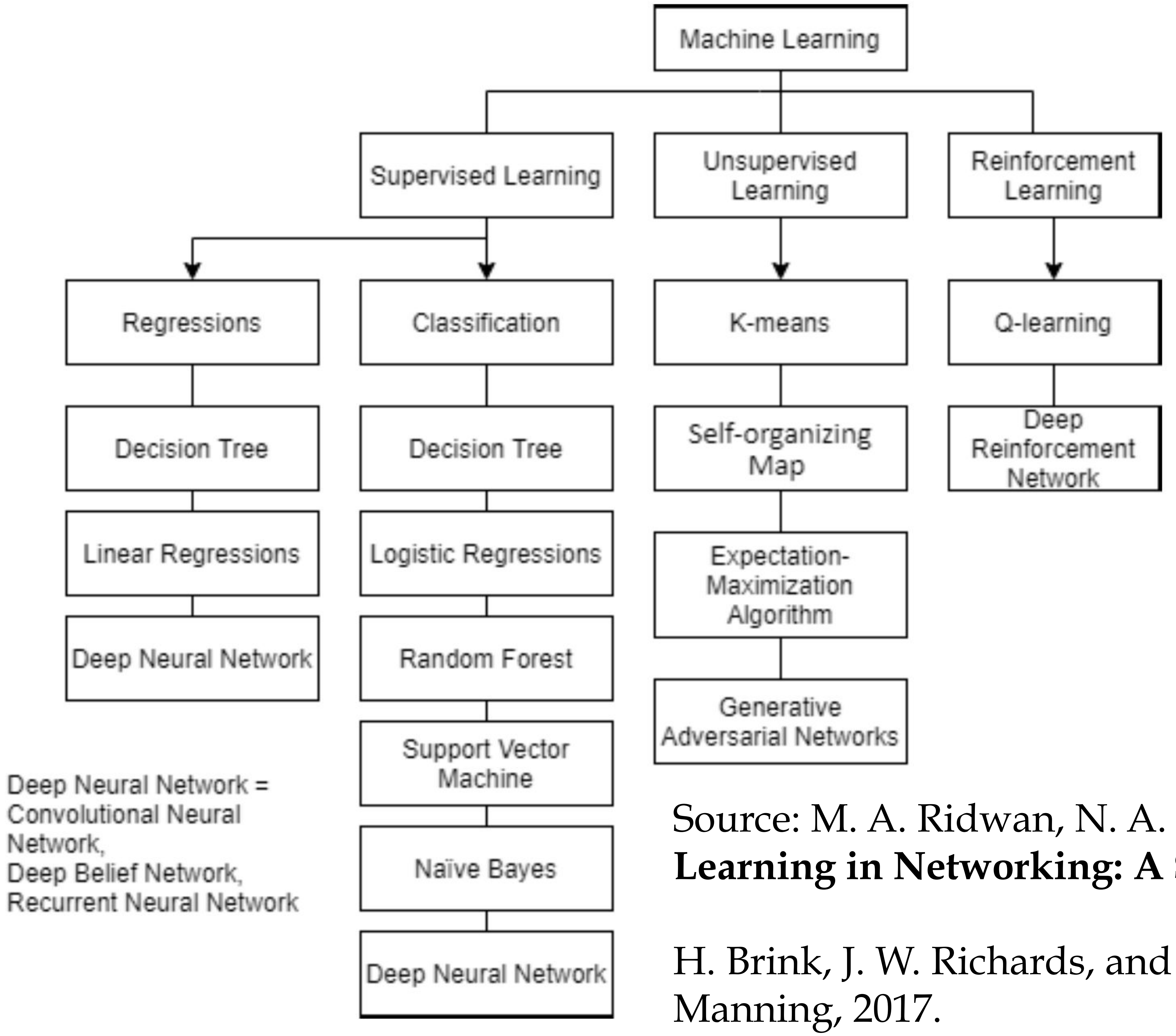
(II) You have to understand the tool:

- The broader overview you have about the particularities of each ML approach, the better chances to apply the correct one (e.g., avoid killing mosquitos with a hammer).
- The research community does not benefit any further from yet another untried ML approach (e.g., IDS based on KDD'99 dataset).
- A good grasp of calculus, linear algebra, and probability is essential for a clear understanding of ML and PR in TMA and Networking.

SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

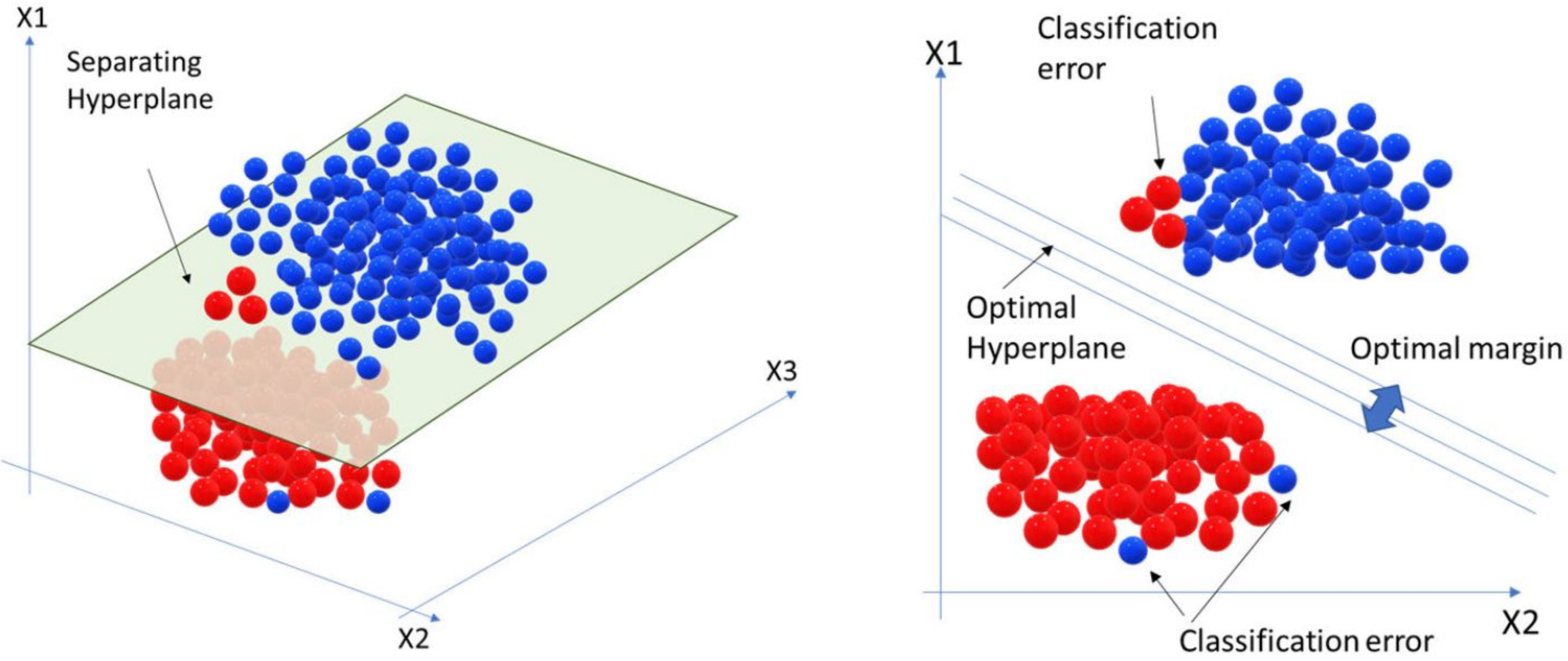
Machine Learning Model Categories



Source: M. A. Ridwan, N. A. M. Radzi, F. Abdullah and Y. E. Jalil, "**Applications of Machine Learning in Networking: A Survey of Current Issues and Future Challenges**," in IEEE Access 2021

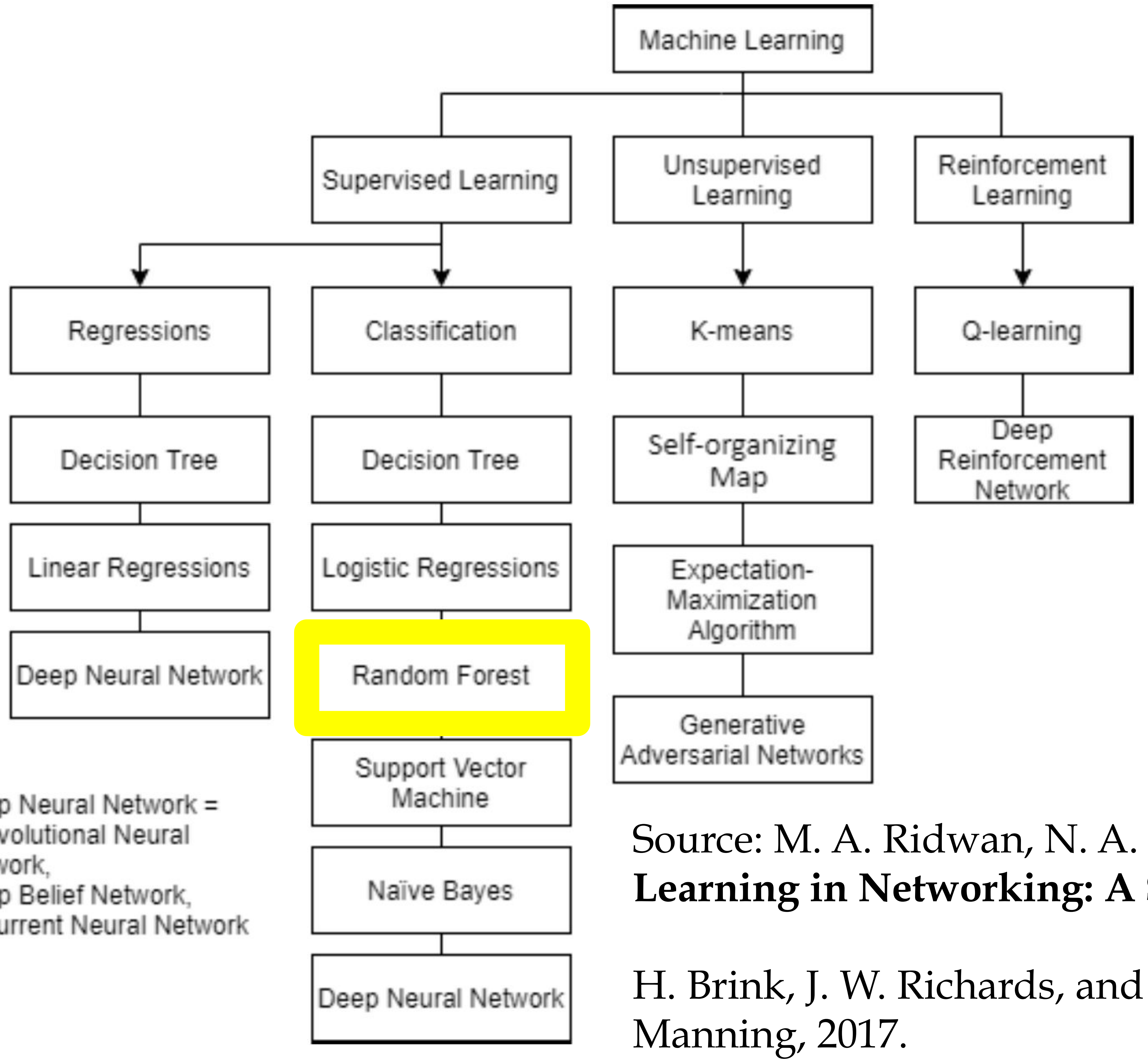
H. Brink, J. W. Richards, and M. Fetherolf, **Real-World Machine Learning**. Shelter Island, NY, USA: Manning, 2017.

Machine Learning Model Categories

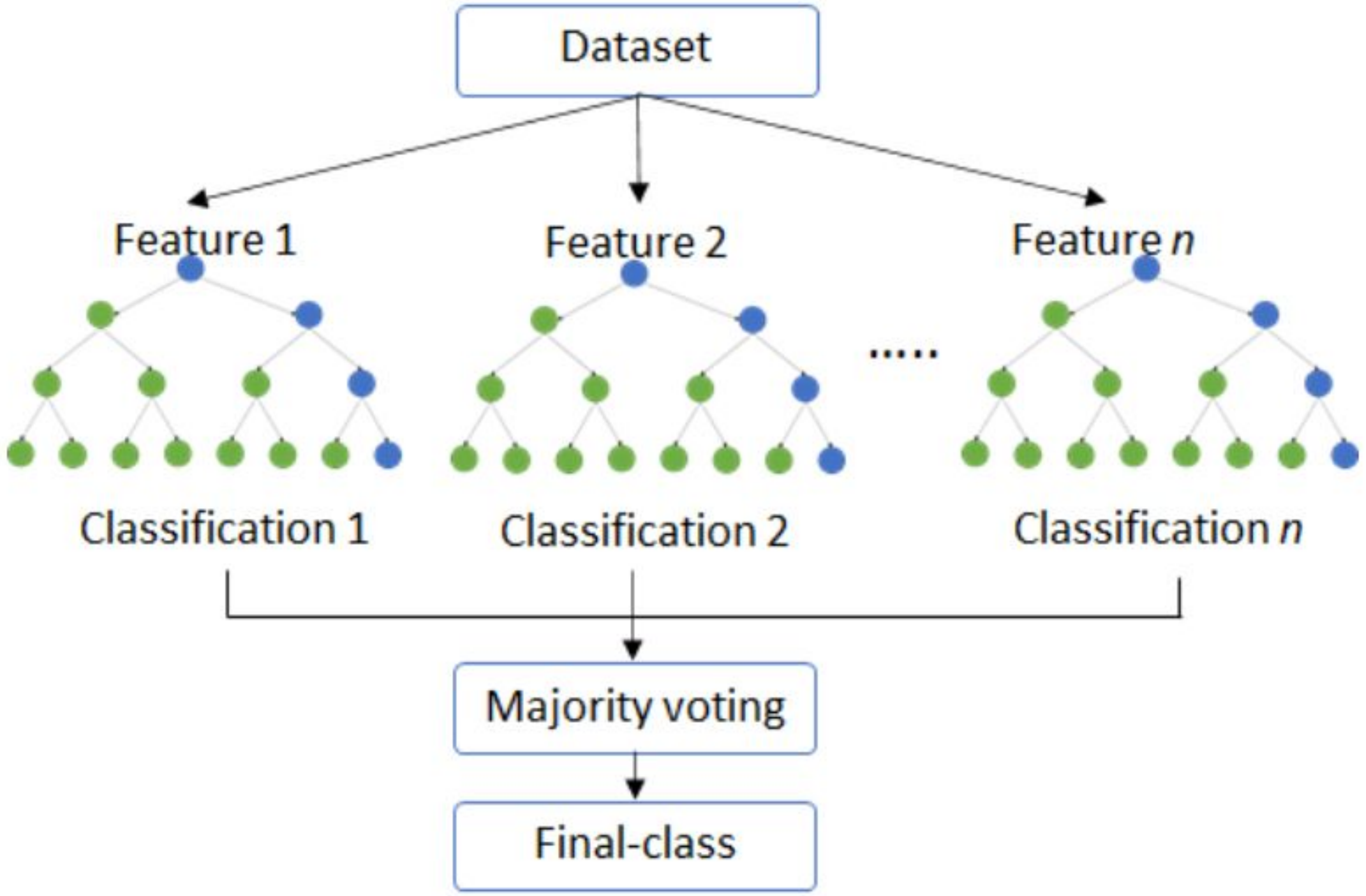


Source: M. A. Ridwan, N. A. M. Radzi, F. Abdullah and Y. E. Jalil, "Ap
Learning in Networking: A Survey of Current Issues and Future Challenges," in IEEE Access 2021
Illustration of SVM hyperplane in 3-dimension space and its
optimal hyperplane and margin in 2-dimension space [

Machine Learning Model Categories



Deep Neural Network = Convolutional Neural Network, Deep Belief Network, Recurrent Neural Network



. Illustration of Random Forest classifier

Source: M. A. Ridwan, N. A. M. Radzi, F. Abdullah and Y. E. Jalil, "**Applications of Machine Learning in Networking: A Survey of Current Issues and Future Challenges**," in IEEE Access 2021

H. Brink, J. W. Richards, and M. Fetherolf, **Real-World Machine Learning**. Shelter Island, NY, USA: Manning, 2017.

Artificial Intelligence – As Smart as a Donut!

- **Machine Learning** is still **very stupid** – the **big revolution** is on **big data processing** and **data availability**/accessibility

- **Current ML benefits** are fundamentally due to machines ability to **blindly**:
  - compute lots of math operations per second
  - handle large amounts of data
  - deal with data in high-dimensional spaces

- **A lot of data required** to “learn“ simple logical inter-relations

- **Shallow Learning**: less data but ***human expert knowledge required***, to properly guide the ***feature engineering*** process

- **Deep Learning**: ***automated feature engineering*** (representation learning) but ***needs much more data***

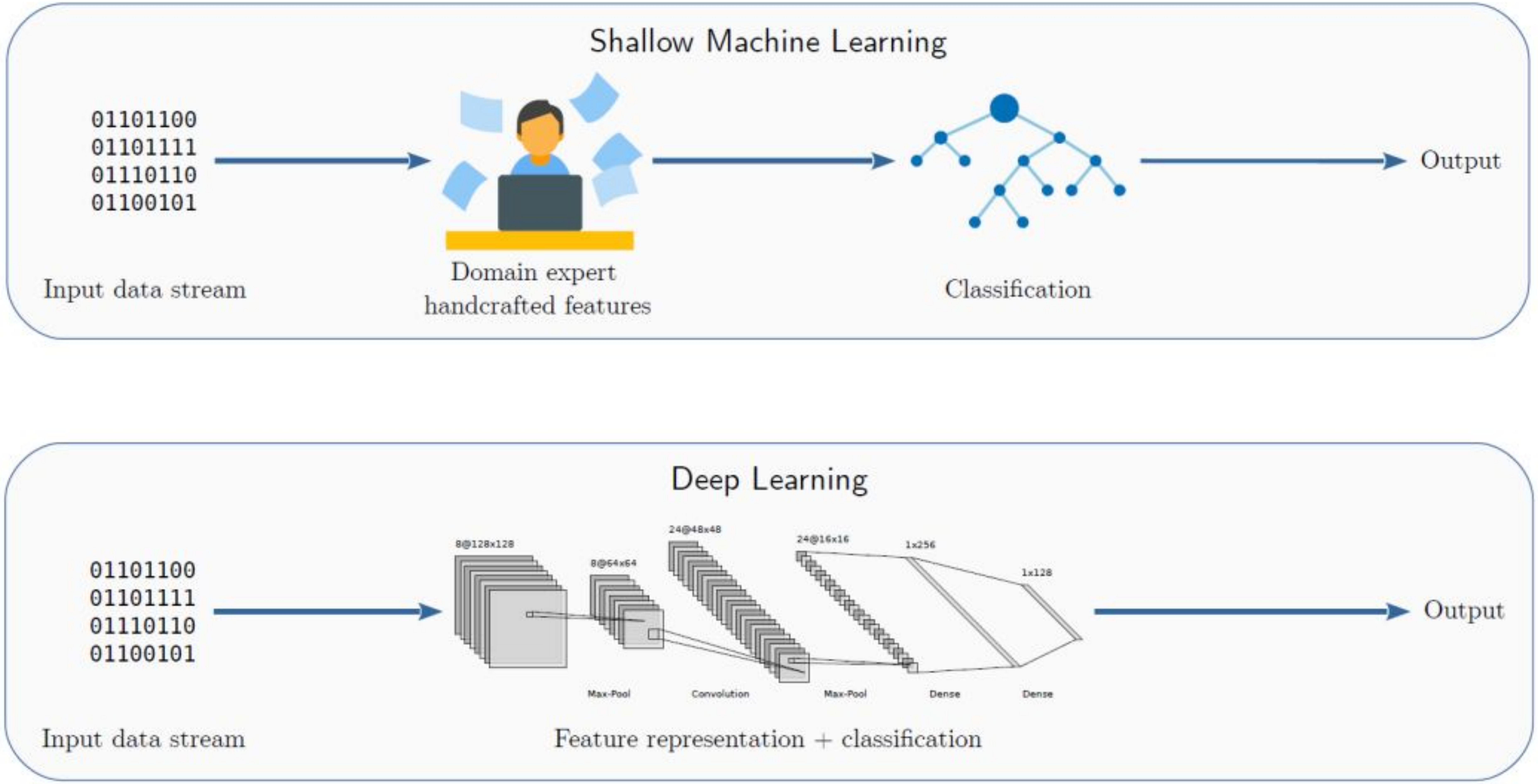
RawPower

we explore **deep learning** for ***blind*** malware detection in network traffic

SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

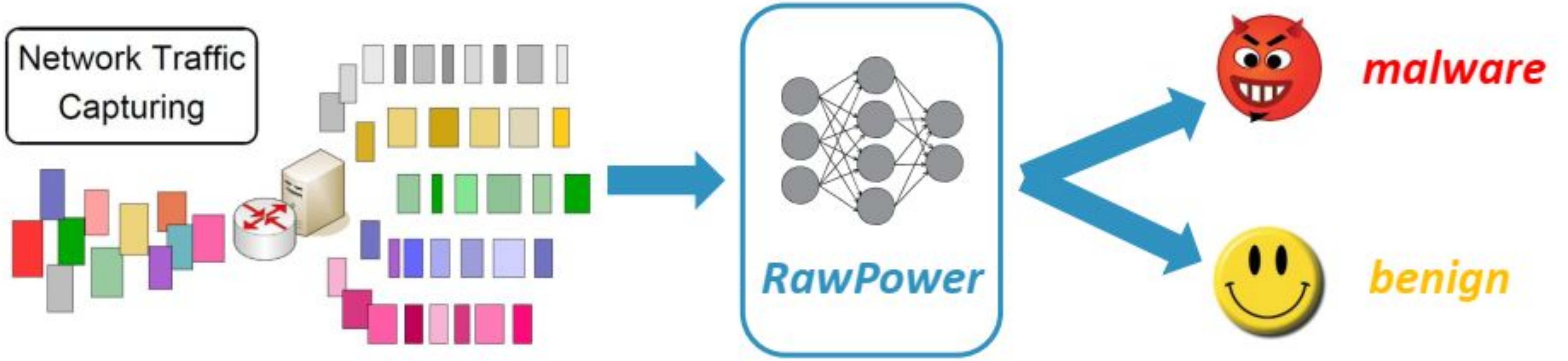
Shallow Learning vs Deep Learning



Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

Basic Concepts of RawPower

- The ***input to the Deep Learning*** model is **RAW** – ***only byte-streams***
- **No need to define** tailored, domain-knowledge-based **input features**



- Different architectures to **analyze both packet-based** and **flow-based byte aggregations**
- Models for **binary malware detection** – fully **supervised-based training**

SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI/ML in Networking?

- Challenges
- Hazards
- Trade-offs

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

AI4NETS – Application to Other Domains

- **AI4NETS** represents a complex **context for AI/ML,** opening the door to **other fields**
- The **challenge is huge:**
  - AI4NETS **involves** all of the **major learning and big data challenges** (**the 4 Vs**)
  - **Massive volumes** of **complex** and heterogeneous **data** (**V**olume and **V**ariety)
  - **Fast** and **highly dynamic** streams of **data** (**V**elocity)
  - **Lack of ground truth** for learning (**V**eracity)

IoT

Industry 4.0

Scalable Computing

Trading

Blockchains

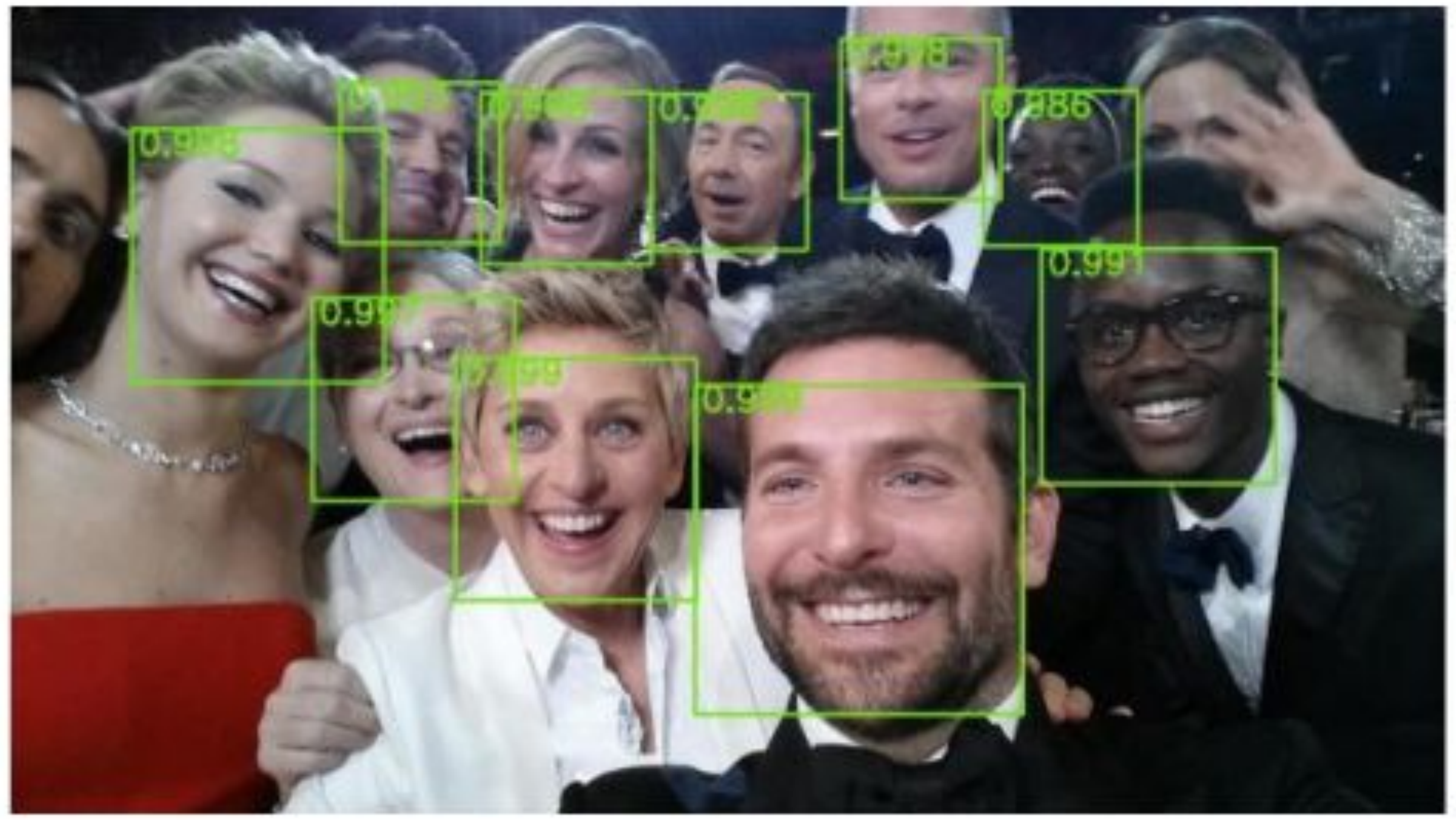
SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI Success in Networking?

- ***Data Complexity:*** the complexity (and heterogeneity) of the data related to Internet-like networks is one of the most significant bottlenecks to AI4NETS

- The Internet, and in general large-scale networks, are a **complex tangle** of networks, technologies, applications, services, devices and end-users



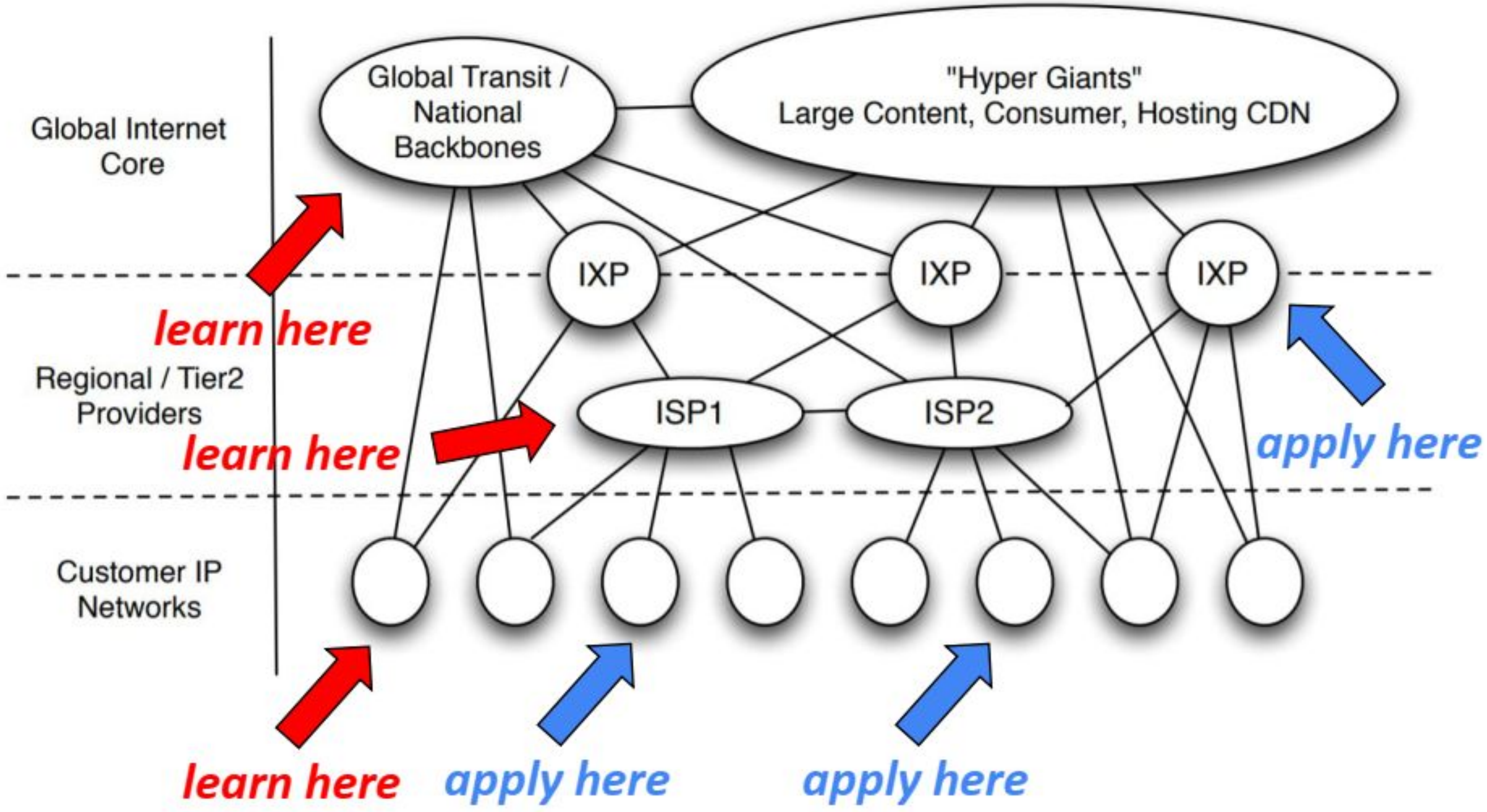
- **AI has so far shown very successful results** generally in **data from** more **predictable** and **easy to understand sources** (natural sources)

SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI Success in Networking?

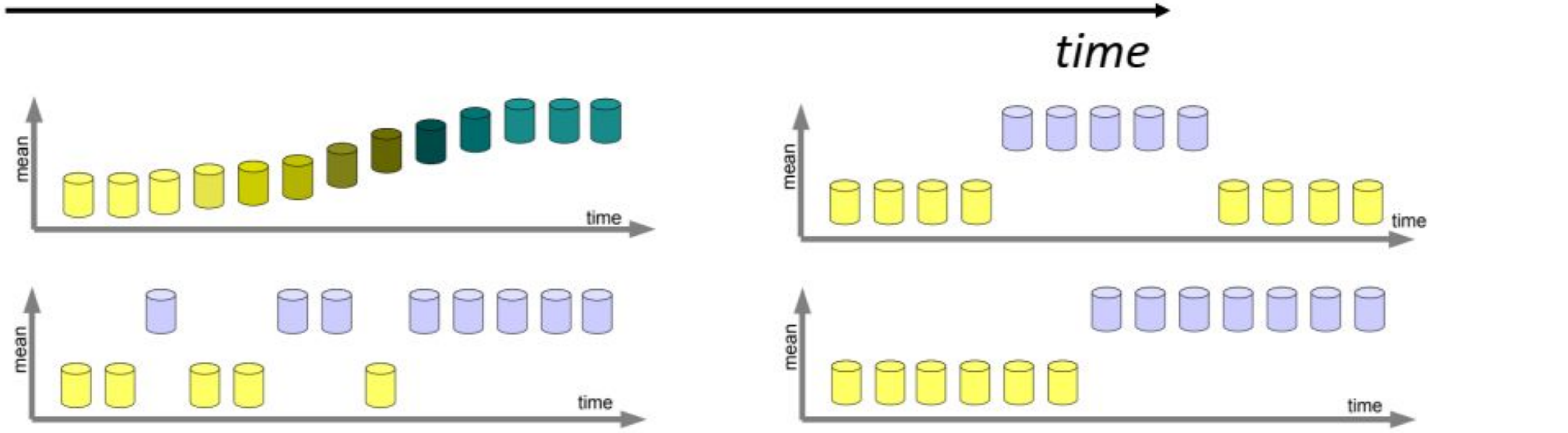
- ***Diversity of Network Data:*** besides complexity, network data often exhibits much more diversity than one would intuitively expect



Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI Success in Networking?

- ***Data Dynamics:*** networking data is non-stationary, generally comes in the form of data streams, and is full of constant concept drifts



Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI Success in Networking?

- ***Lack of Ground Truth:*** in the wild networking data is usually non-labeled
- ***Lack of Standardized and Representative Datasets***: datasets are generally biased, difficult to find appropriate public datasets to assess AI4NETS

- There is **no IMAGENET** or the like in Networking
- **Network data labeling**, and even **data interpretation**, is **too complex** for humans, even for **domain experts** (e.g., malware vs benign traffic instead of cat vs dog)
  - Easier for naturally generated data: images, text, audio

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI Success in Networking?

- ***Lack of Interpretability:*** this is a general problem of ML models (e.g., DL provides *beautiful* black-boxes)…but the issue is even more complex in AI4NETS

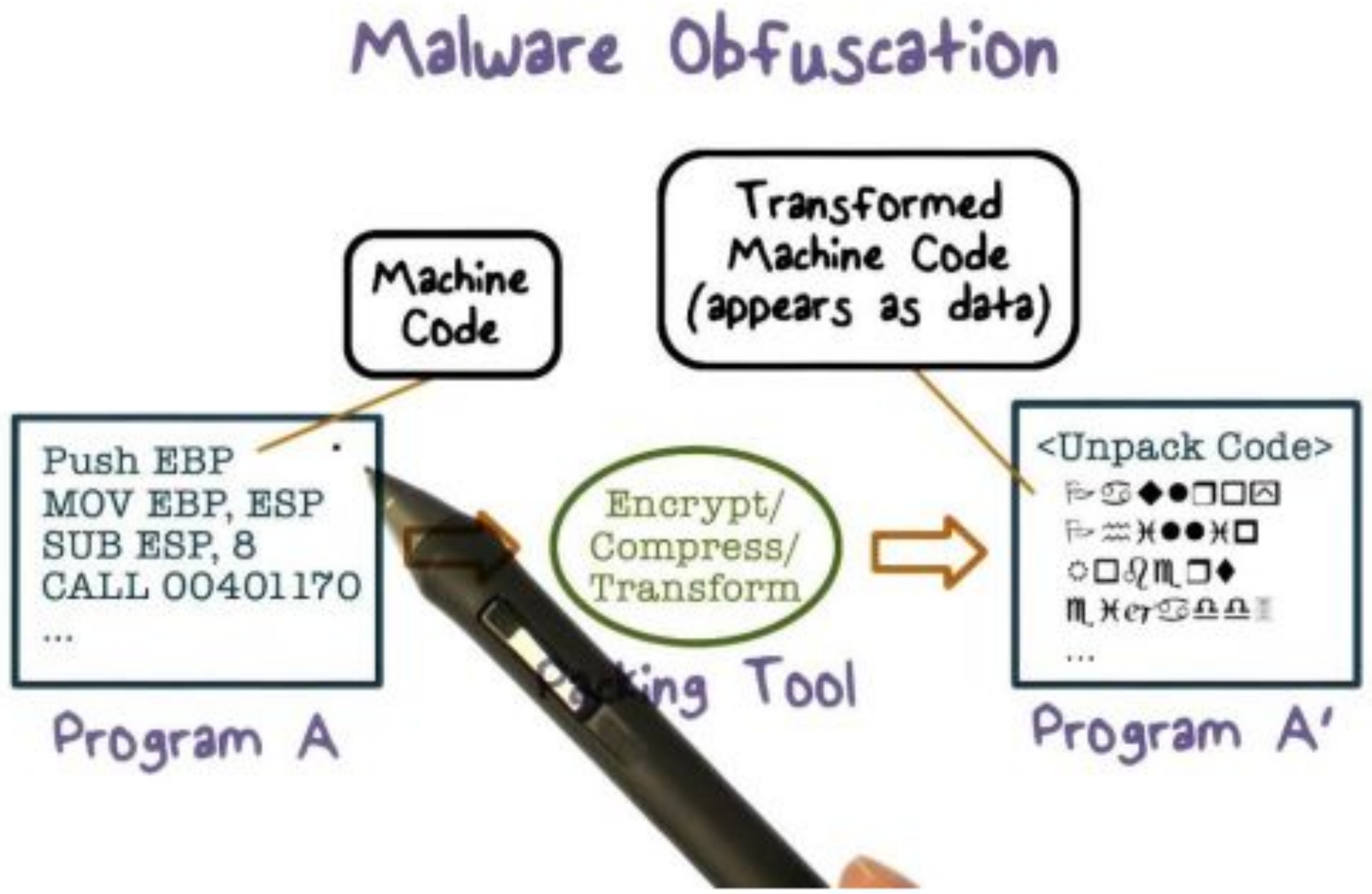
- To **improve trust**, the end-user (**humans**) has to **trust model predictions**, for example, by **understanding which inputs lead to a specific output**, but generally difficult to interpret networking features

- The **lack of interpretability and trust stops AI deployments**:
  - Network security – AI4NETSEC
  - Dynamic Traffic Engineering – AI4NETTE
  - Dynamic network instantiation (NFV) and (re)-configuration (SDN) – AI4SELFNET

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI Success in Networking?

- ***Learning occurs in an Adversarial Setting:*** services obfuscate and modify their functioning to bypass monitoring and avoid traffic engineering policies



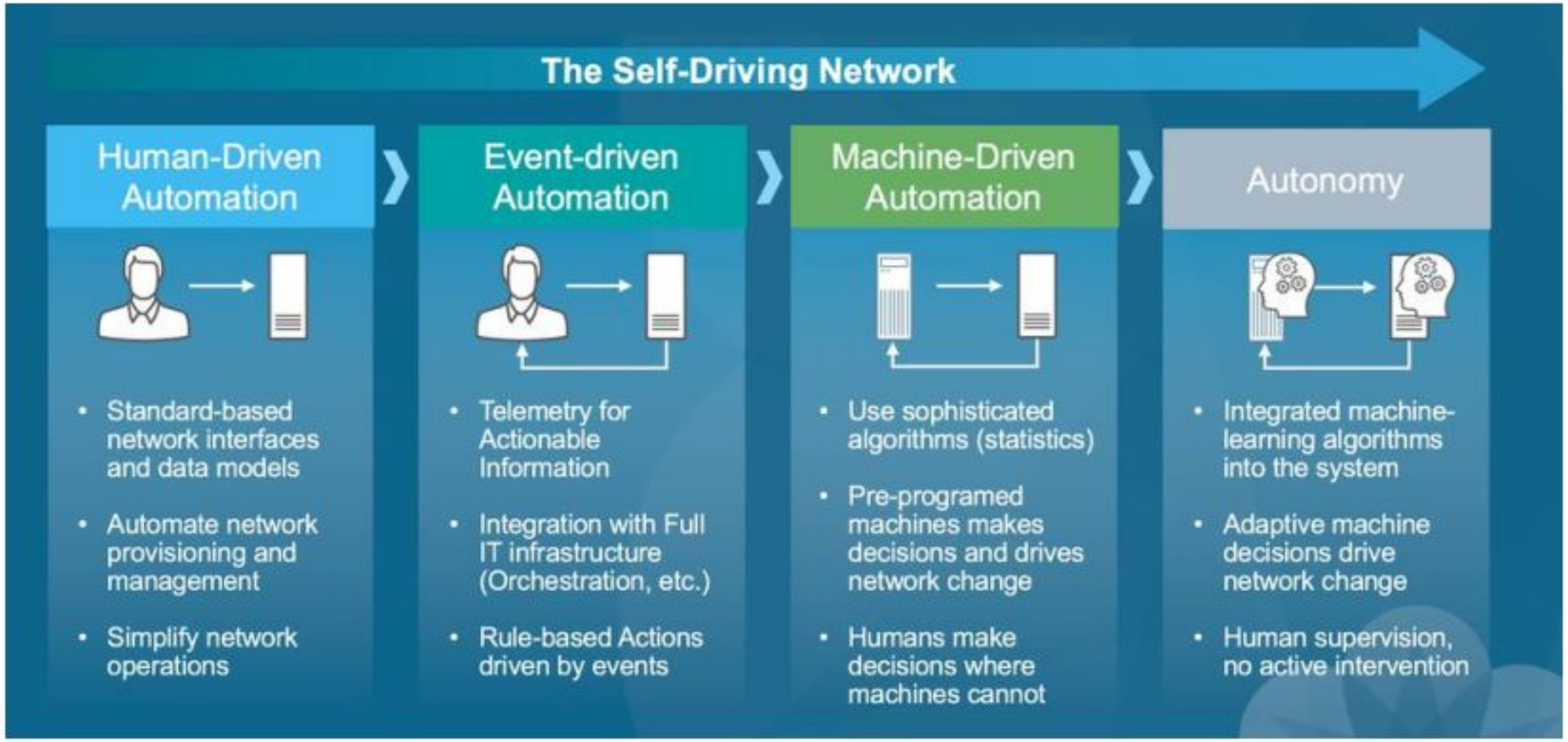
- It becomes even **more trickier to learn**, when the ***adversary*** constantly tries to fool the learner
- **Not only malign actors**, but standard *services*: Skype, QUIC, etc.

SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

What is Blocking AI Success in Networking?

- ***Lack of Learning Generalization:*** it becomes ***extremely difficult*** in the networking practice to learn ***models which can generalize to operational environments***

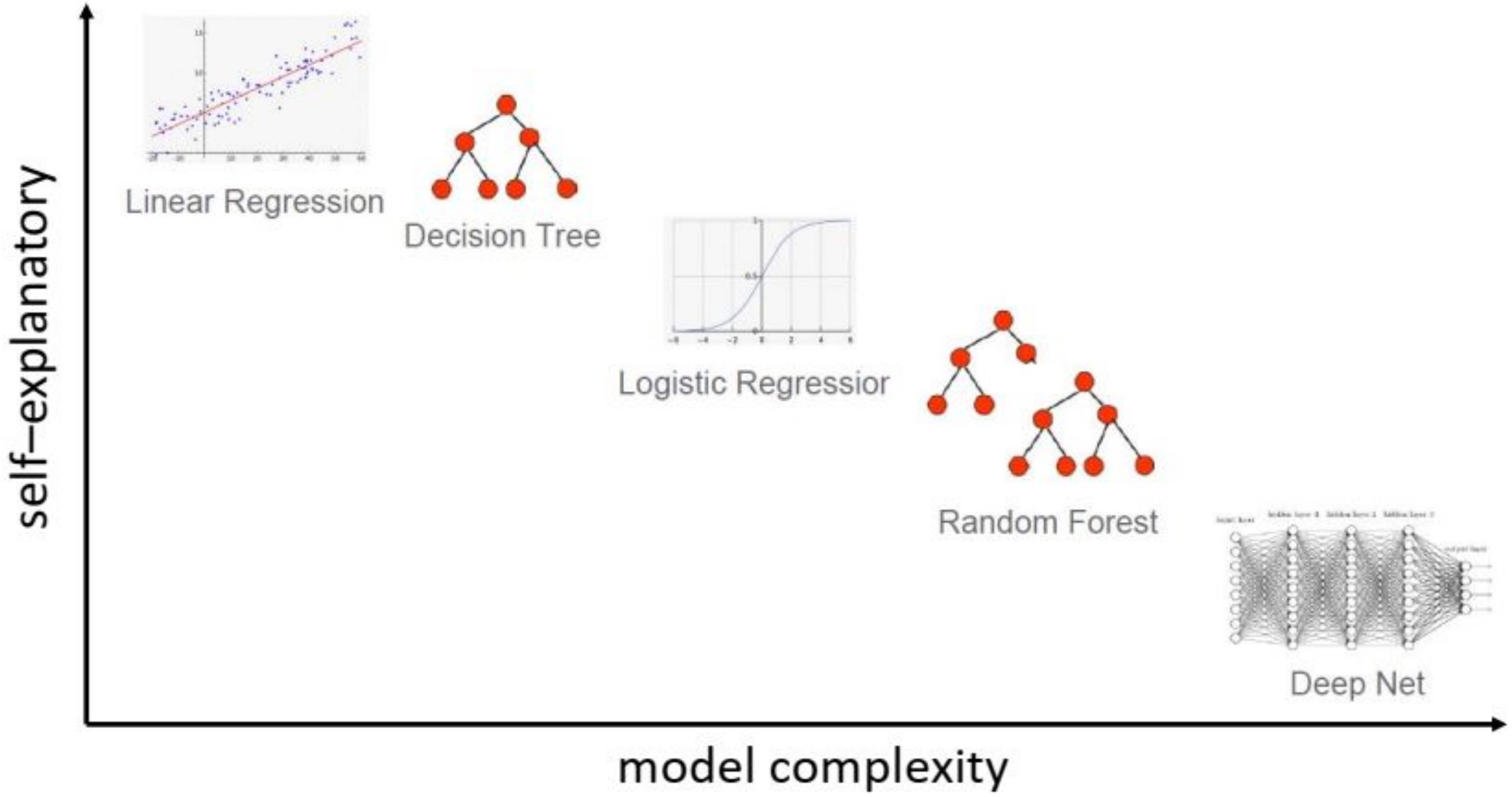


SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

Why XAI?

- **Ideally, ML models** should be ***self-explanatory:*** **improve end-user understanding** and **trust**, by offering simple **explanations** of the ***"whys" of certain decision***
- **Only few models are self-explanatory:**



SMARTNESS

Source: Slide Courtesy by Pedro Casas: https://bigdama.ait.ac.at/pcasas/

SMARTNESS

ML & Networking

- Surveys of Research Works

Table 2 Performance metrics for accuracy validation

Metric	Description
Mean Absolute Error (MAE)	Average of the absolute error between the actual and predicted values. Facilitates error interpretability.
Mean Squared Error (MSE)	Average of the squares of the error between the actual and predicted values. Heavily penalizes large errors.
Mean Absolute Prediction Error (MAPE)	Percentage of the error between the actual and predicted values. Not reliable for zero values or low-scale data.
Root MSE (RMSE)	Squared root of MSE. Represents the standard deviation of the error between the actual and predicted values.
Normalized RMSE (NRMSE)	Normalized RMSE. Facilitates comparing different models independently of their working scale.
Cross-entropy	Metric based on the logistic function that measures the error between the actual and predicted values.
Accuracy	Proportion of correct predictions among the total number of predictions. Not reliable for skewed class-wise data.
True Positive Rate (TPR) or recall	Proportion of actual positives that are correctly predicted. Represents the sensitivity or detection rate (DR) of a model.
False Positive Rate (FPR)	Proportion of actual negatives predicted as positives. Represents the significance level of a model.
True Negative Rate (TNR)	Proportion of actual negatives that are correctly predicted. Represents the specificity of a model.
False Negative Rate (FNR)	Proportion of actual positives predicted as negatives. Inversely proportional to the statistical power of a model.
Received Operating Characteristic (ROC)	Curve that plots TPR versus FPR at different parameter settings. Facilitates analyzing the cost-benefit of possibly optimal models.
Area Under the ROC Curve (AUC)	Probability of confidence in a model to accurately predict positive outcomes for actual positive instances.
Precision	Proportion of positive predictions that are correctly predicted.
F-measure	Harmonic mean of precision and recall. Facilitates analyzing the trade-off between these metrics.
Coefficient of Variation (CV)	Intra-cluster similarity to measure the accuracy of unsupervised classification models based on clusters.

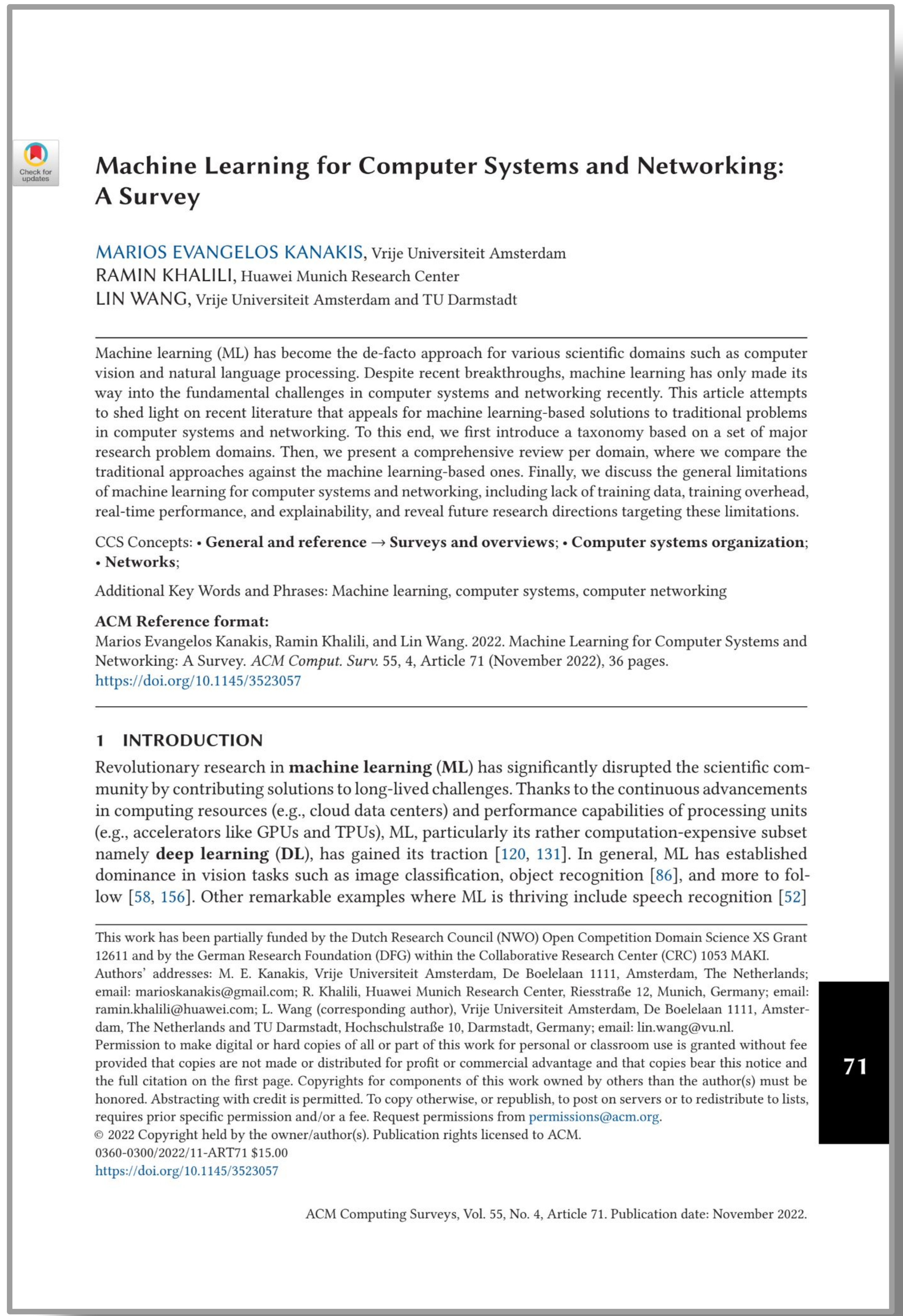
SMARTNESS

Surveys of ML for Networking

Tons of use case examples

ML for Computer Systems and Networking

Survey of selected problem and solution examples



Machine Learning for Computer Systems and Networking: A Survey

MARIOS EVANGELOS KANAKIS, Vrije Universiteit Amsterdam
RAMIN KHALILI, Huawei Munich Research Center
LIN WANG, Vrije Universiteit Amsterdam and TU Darmstadt

Machine learning (ML) has become the de-facto approach for various scientific domains such as computer vision and natural language processing. Despite recent breakthroughs, machine learning has only made its way into the fundamental challenges in computer systems and networking recently. This article attempts to shed light on recent literature that appeals for machine learning-based solutions to traditional problems in computer systems and networking. To this end, we first introduce a taxonomy based on a set of major research problem domains. Then, we present a comprehensive review per domain, where we compare the traditional approaches against the machine learning-based ones. Finally, we discuss the general limitations of machine learning for computer systems and networking, including lack of training data, training overhead, real-time performance, and explainability, and reveal future research directions targeting these limitations.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Computer systems organization**; • **Networks**;

Additional Key Words and Phrases: Machine learning, computer systems, computer networking

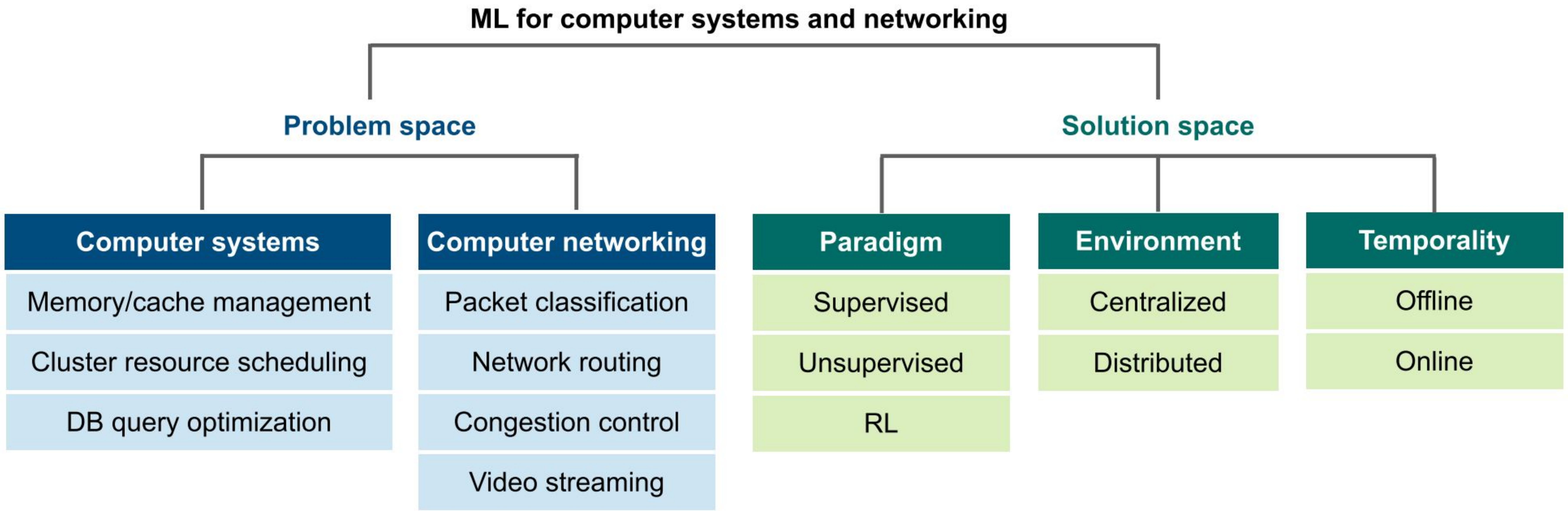
ACM Reference format:
Marios Evangelos Kanakis, Ramin Khalili, and Lin Wang. 2022. Machine Learning for Computer Systems and Networking: A Survey. *ACM Comput. Surv.* 55, 4, Article 71 (November 2022), 36 pages.
https://doi.org/10.1145/3523057

1 INTRODUCTION

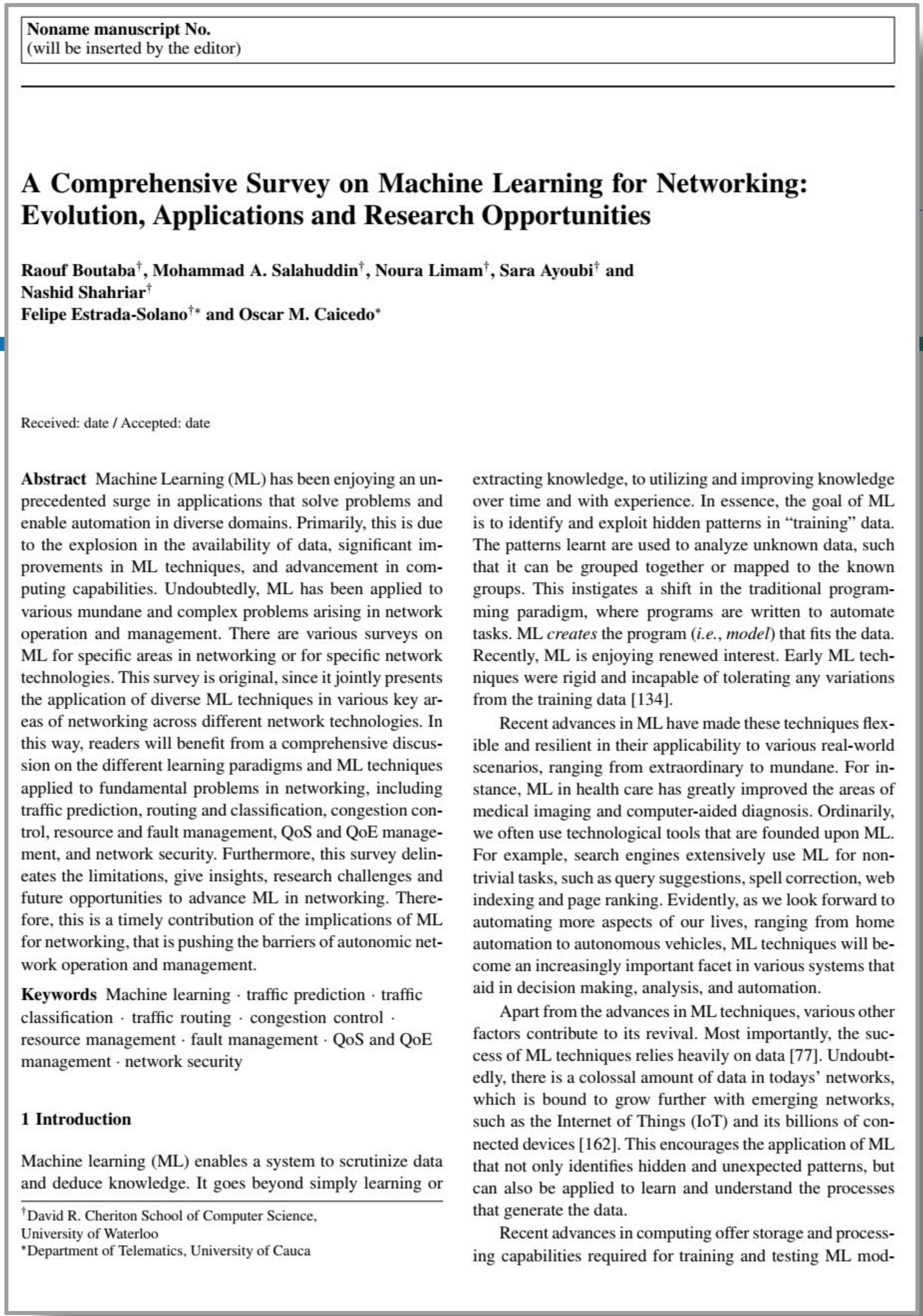
Revolutionary research in **machine learning** (**ML**) has significantly disrupted the scientific community by contributing solutions to long-lived challenges. Thanks to the continuous advancements in computing resources (e.g., cloud data centers) and performance capabilities of processing units (e.g., accelerators like GPUs and TPUs), ML, particularly its rather computation-expensive subset namely **deep learning** (**DL**), has gained its traction [120, 131]. In general, ML has established dominance in vision tasks such as image classification, object recognition [86], and more to follow [58, 156]. Other remarkable examples where ML is thriving include speech recognition [52]

This work has been partially funded by the Dutch Research Council (NWO) Open Competition Domain Science XS Grant 12611 and by the German Research Foundation (DFG) within the Collaborative Research Center (CRC) 1053 MAKI.
Authors' addresses: M. E. Kanakis, Vrije Universiteit Amsterdam, De Boelelaan 1111, Amsterdam, The Netherlands; email: marioskanakis@gmail.com; R. Khalili, Huawei Munich Research Center, Riesstraße 12, Munich, Germany; email: ramin.khalili@huawei.com; L. Wang (corresponding author), Vrije Universiteit Amsterdam, De Boelelaan 1111, Amsterdam, The Netherlands and TU Darmstadt, Hochschulstraße 10, Darmstadt, Germany; email: lin.wang@vu.nl.

0360-0300/2022/11-ART71 $15.00
https://doi.org/10.1145/3523057



Source: Marios Evangelos Kanakis, Ramin Khalili, and Lin Wang. **Machine Learning for Computer Systems and Networking: A Survey**. ACM Comput. Surv. 2023



Noname manuscript No.
(will be inserted by the editor)

A Comprehensive Survey on Machine Learning for Networking: Evolution, Applications and Research Opportunities

Raouf Boutaba[†], Mohammad A. Salahuddin[†], Noura Limam[†], Sara Ayoubi[†] and Nashid Shahriar[†]
Felipe Estrada-Solano[†*] and Oscar M. Caicedo[*]

Received: date / Accepted: date

Abstract Machine Learning (ML) has been enjoying an unprecedented surge in applications that solve problems and enable automation in diverse domains. Primarily, this is due to the explosion in the availability of data, significant improvements in ML techniques, and advancement in computing capabilities. Undoubtedly, ML has been applied to various mundane and complex problems arising in network operation and management. There are various surveys on ML for specific areas in networking or for specific network technologies. This survey is original, since it jointly presents the application of diverse ML techniques in various key areas of networking across different network technologies. In this way, readers will benefit from a comprehensive discussion on the different learning paradigms and ML techniques applied to fundamental problems in networking, including traffic prediction, routing and classification, congestion control, resource and fault management, QoS and QoE management, and network security. Furthermore, this survey delineates the limitations, give insights, research challenges and future opportunities to advance ML in networking. Therefore, this is a timely contribution of the implications of ML for networking, that is pushing the barriers of autonomic network operation and management.

Keywords Machine learning · traffic prediction · traffic classification · traffic routing · congestion control · resource management · fault management · QoS and QoE management · network security

1 Introduction

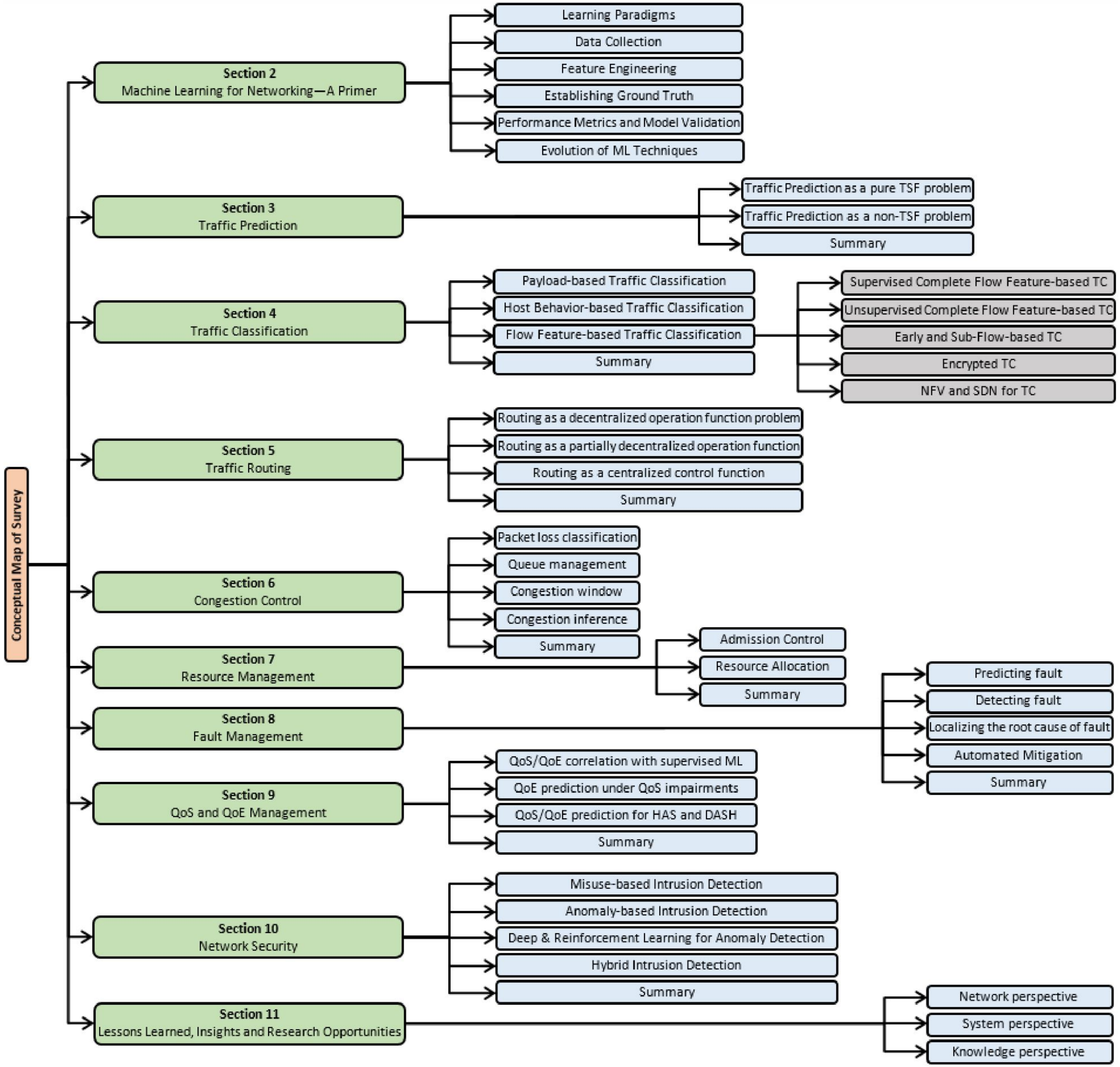
Machine learning (ML) enables a system to scrutinize data and deduce knowledge. It goes beyond simply learning or extracting knowledge, to utilizing and improving knowledge over time and with experience. In essence, the goal of ML is to identify and exploit hidden patterns in "training" data. The patterns learnt are used to analyze unknown data, such that it can be grouped together or mapped to the known groups. This instigates a shift in the traditional programming paradigm, where programs are written to automate tasks. ML *creates* the program (*i.e.*, *model*) that fits the data. Recently, ML is enjoying renewed interest. Early ML techniques were rigid and incapable of tolerating any variations from the training data [134].

Recent advances in ML have made these techniques flexible and resilient in their applicability to various real-world scenarios, ranging from extraordinary to mundane. For instance, ML in health care has greatly improved the areas of medical imaging and computer-aided diagnosis. Ordinarily, we often use technological tools that are founded upon ML. For example, search engines extensively use ML for non-trivial tasks, such as query suggestions, spell correction, web indexing and page ranking. Evidently, as we look forward to automating more aspects of our lives, ranging from home automation to autonomous vehicles, ML techniques will become an increasingly important facet in various systems that aid in decision making, analysis, and automation.

Apart from the advances in ML techniques, various other factors contribute to its revival. Most importantly, the success of ML techniques relies heavily on data [77]. Undoubtedly, there is a colossal amount of data in todays' networks, which is bound to grow further with emerging networks, such as the Internet of Things (IoT) and its billions of connected devices [162]. This encourages the application of ML that not only identifies hidden and unexpected patterns, but can also be applied to learn and understand the processes that generate the data.

Recent advances in computing offer storage and processing capabilities required for training and testing ML mod-

[†]David R. Cheriton School of Computer Science, University of Waterloo
[*]Department of Telematics, University of Cauca



Source: R. Boutaba et al. "**A comprehensive survey on machine learning for networking: evolution, applications and research opportunities**". JISA, 2018.

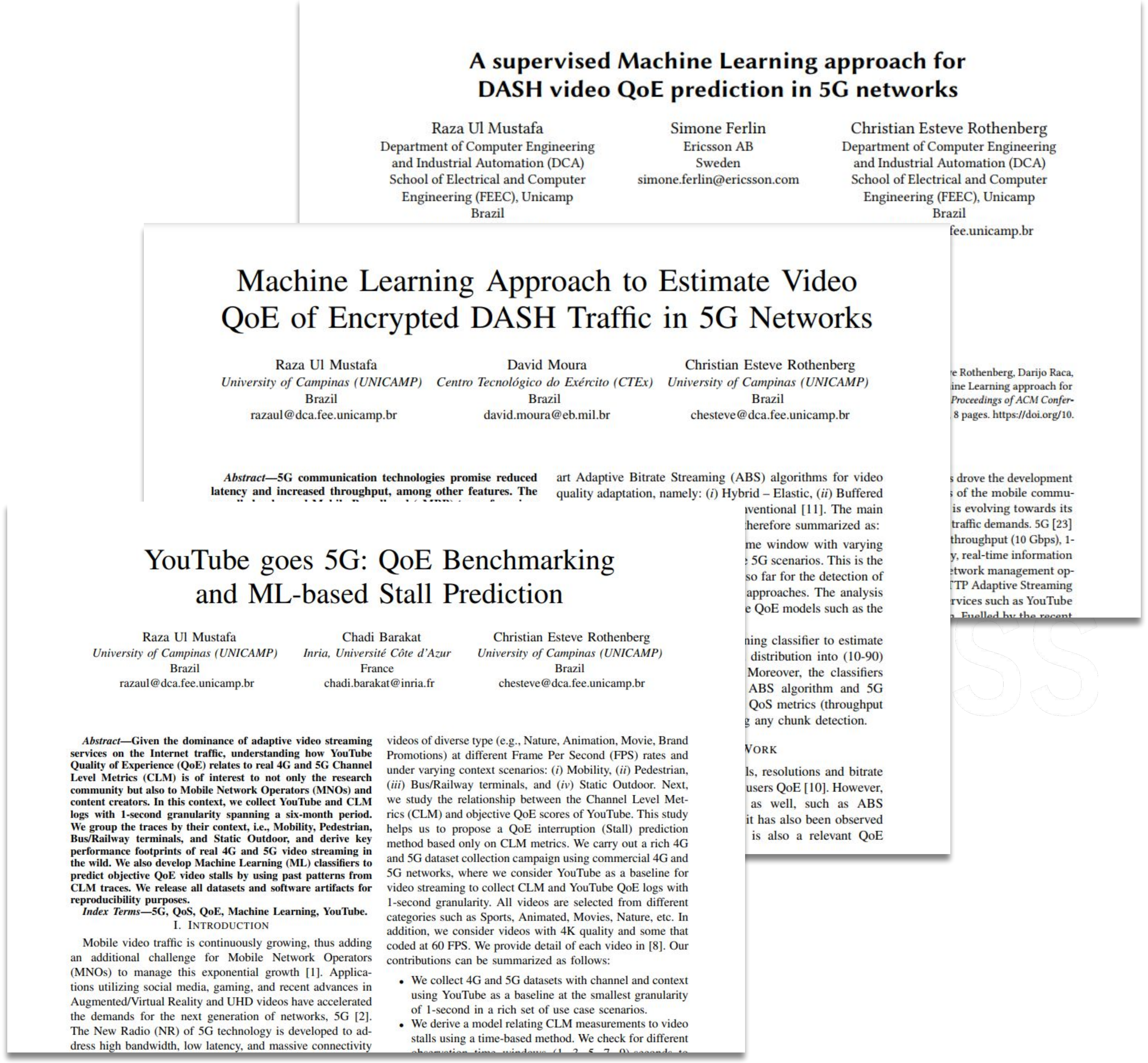
Survey of Machine Intelligence for Networking

Network Traffic Prediction via Time Series Forecasting (TSF)

Table 3 Summary of TSF and non-TSF-based Traffic Prediction

Ref.	ML Technique	Application *(approach)*	Dataset *(availability)*	Features	Output *(training)*	Evaluation: Settings*†	Evaluation: Results‡
NBP [141]	Supervised: · MLP-NN *(offline)*	End-to-end path bandwidth availability prediction *(TSF)*	NSF TeraGrid dataset *(N/A)*	Max, Min, Avg load observed in past 10s $\sim$ 30s	Available bandwidth on a end-to-end path in future epoch	Number of features= 3 MLP-NN: · *(N/A)*	MSE = 8%
Cortez *et al.* [104]	Supervised: · NNE trained with Rp *(offline)*	Link load and traffic volume prediction in ISP networks *(TSF)*	SNMP traffic data from 2 ISP nets, · traffic on a transatlantic link · aggregated traffic in the ISP backbone *(N/A)*	Traffic volume observed in past few minutes$\sim$several days	Expected traffic volume	Number of features= $6 \sim 9$ 5 NNs NNE: · all SLPs for dataset1 · 1 hidden layer MLPs with $6 \sim 8$ neurons for dataset2	1h lookahead: · MAPE = 1.43% $\sim$ 5.23% 1h $\sim$ 24h lookahead: · MAPE = 6.34% $\sim$ 23.48%)
Bermolen *et al.* [52]	Supervised: · SVR *(offline)*	Link load prediction in ISP networks *(TSF)*	Internet traffic collected at the POP of an ISP network *(N/A)*	Link load observed at τ time scale	Expected link load	Number of features= d samples with $d = 1..30$ Number of support vectors: · varies with d (*e.g.*, ~ 320 for $d = 10$)	RMSE < 2 for $\tau = 1ms$ and $d = 5$ · $\approx$ AR ·10% less than MA
Chabaa *et al.* [86]	Supervised: MLP-NN with different training algorithms (GD, CG, SS, LM, Rp) *(offline)*	Network traffic prediction *(TSF)*	1000 points dataset *(N/A)*	Past measurements	Expected traffic volume	Number of features *(N/A)* MLP-NN: · 1 hidden layer	LM: · RMSE= 0.0019 RPE = 0.0230% Rp: · RMSE= 0.0031 RPE= 0.0371%

Source: R. Boutaba et al. "**A comprehensive survey on machine learning for networking: evolution, applications and research opportunities**". JISA, 2018.



SMARTNESS

ML for Networking Use Cases

Selected Publications

- **Casos reales**
  - YouTube QoE en redes 4G/5G
  - Identificación de fallos suaves en redes ópticas usando Network Digital Twin

Why are traditional QoS metrics, like latency and bandwidth, insufficient alone for accurately predicting QoE in 5G networks

A supervised Machine Learning approach for DASH video QoE prediction in 5G networks

Raza Ul Mustafa
Department of Computer Engineering and Industrial Automation (DCA)
School of Electrical and Computer Engineering (FEEC), Unicamp
Brazil
r211580@dac.unicamp.br

Simone Ferlin
Ericsson AB
Sweden
simone.ferlin@ericsson.com

Christian Esteve Rothenberg
Department of Computer Engineering and Industrial Automation (DCA)
School of Electrical and Computer Engineering (FEEC), Unicamp
Brazil
chesteve@dca.fee.unicamp.br

Darijo Raca
Faculty of Electrical Engineering,
University of Sarajevo, Sarajevo, BiH
draca@etf.unsa.ba

Jason J. Quinlan
School of Computer Science & Information Technology,
University College Cork, Ireland
j.quinlan@cs.ucc.ie

ABSTRACT

Future fifth generation (5G) networks are envisioned to provide improved Quality-of-Experience (QoE) for applications by means of higher data rates, low and ultra-reliable latency and very high reliability. Proving increasing beneficial for mobile devices running multimedia applications. However, there exist two main co-related challenges in multimedia delivery in 5G. Namely, balancing operator provisioning and client expectations. To this end, we investigate how to build a QoE-aware network that guarantees at run-time that the end-to-end user experience meets the end users' expectations at the same that the network's Quality of Service (QoS) varies.

The contribution of this paper is twofold: First, we consider a Dynamic Adaptive Streaming over HTTP (DASH) video application in a realistic emulation environment derived from real 5G traces in static and mobility scenarios to assess the QoE performance of three state-of-art Adaptive Bitrate Streaming (ABS) algorithm categories: Hybrid - Elastic and Arbiter+; buffer-based - BBA and Logistic; and

ACM Reference Format:
Raza Ul Mustafa, Simone Ferlin, Christian Esteve Rothenberg, Darijo Raca, and Jason J. Quinlan. 2020. A supervised Machine Learning approach for DASH video QoE prediction in 5G networks. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The steady growth of Internet data services drove the development of third (3G) and fourth (4G) generations of the mobile communications standard. Now, the technology is evolving towards its fifth-generation (5G), motivated by similar traffic demands. 5G [23] is expected to support significantly higher throughput (10 Gbps), 1-millisecond end-to-end over-the-air latency, real-time information processing and transmission, and lower network management operation complexity. In video streaming, HTTP Adaptive Streaming (HAS) is the de-facto choice of popular services such as YouTube and Netflix for Internet video distribution. Fuelled by the recent

Machine Learning Approach to Estimate Video QoE of Encrypted DASH Traffic in 5G Networks

Raza Ul Mustafa
University of Campinas (UNICAMP)
Brazil
razaul@dca.fee.unicamp.br

David Moura
Centro Tecnológico do Exército (CTEx)
Brazil
david.moura@eb.mil.br

Christian Esteve Rothenberg
University of Campinas (UNICAMP)
Brazil
chesteve@dca.fee.unicamp.br

***Abstract*—5G communication technologies promise reduced latency and increased throughput, among other features. The so-called enhanced Mobile Broadband (eMBB) type of services will contribute to further adoption of video streaming services. In this work, we use a realistic emulation environment based on 5G traces to investigate how Dynamic Adaptive Streaming over HTTP (DASH) video content using three state-of-art Adaptive Bitrate Streaming (ABS) algorithms is impacted in static and mobility scenarios. Given the wide adoption of end-to-end encryption, we use machine learning (ML) models to estimate multiple key video Quality of Experience (QoE) indicators (KQIs) taking network-level Quality of Service (QoS) metrics as input features. The proposed feature extraction method does not require chunk-detection, significantly reducing the complexity of the monitoring approach and providing new means for QoE evaluation of HAS protocols. We show that our ML classifiers achieve a QoE prediction accuracy above 91%.**

***Index Terms*—5G, QoE, TLS, machine learning, QoS, HTTPS, DASH, HAS**

I. Introduction

Video content providers such as YouTube, Netflix, Amazon Prime, and Hulu use HTTP adaptive streaming (HAS) with HTTPS to deliver end-to-end encrypted video streaming services [2]. Cisco predicts that the HAS traffic will be at the top of traffic load with 82% from all Internet traffic by 2022 [1]. This increase in HAS traffic has opened

art Adaptive Bitrate Streaming (ABS) algorithms for video quality adaptation, namely: (*i*) Hybrid – Elastic, (*ii*) Buffered – BBA, and (*iii*) Rate-based – Conventional [11]. The main contributions of this work can be therefore summarized as:

- QoE assessment in 500 ms time window with varying bandwidth in static and mobile 5G scenarios. This is the smallest granularity proposed so far for the detection of anomaly and troubleshooting approaches. The analysis is undertaken through objective QoE models such as the P.1203 QoE standard [8, 6].
- A proposal of a machine learning classifier to estimate QoE based on packets length distribution into (10-90) percentile in 0.5 s intervals. Moreover, the classifiers are unaware of the specific ABS algorithm and 5G scenarios, using only network QoS metrics (throughput and packets) and not requiring any chunk detection.

II. Related Work

Previous works support that stalls, resolutions and bitrate are the main reasons that affect end users QoE [10]. However, other factors cannot be ignored as well, such as ABS adaptation mechanisms. Similarly, it has also been observed that continuous quality switching is also a relevant QoE factor [4].

Key Points

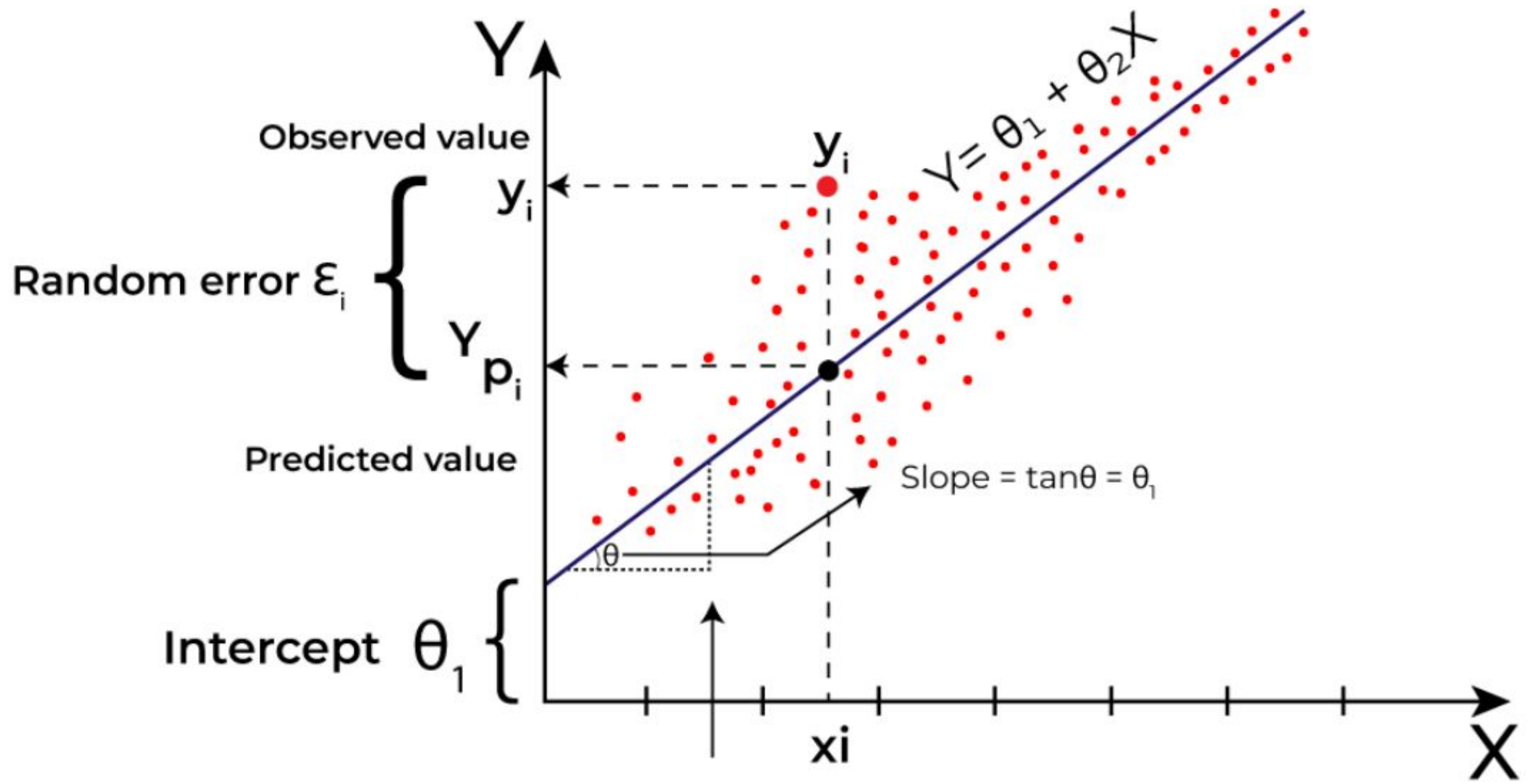
- **Interpacket Gap – IPG** can be used as a key metrics for objective QoE assessment
  - From IPG we can derive a few more metrics for QoE, such as EMA, DEMA, CUSUM
- IPGs along with other traditional QoS metrics are highly correlated to objective QoE KPIs
- Machine Learning can be used to predict QoE
  - correlation with flow-level and network QoS metrics

Regression - Unencrypted Traffic (1/2)

Which QoS features can be effectively used from the network level of unencrypted DASH traffic to estimate the QoE of adaptive video streaming?

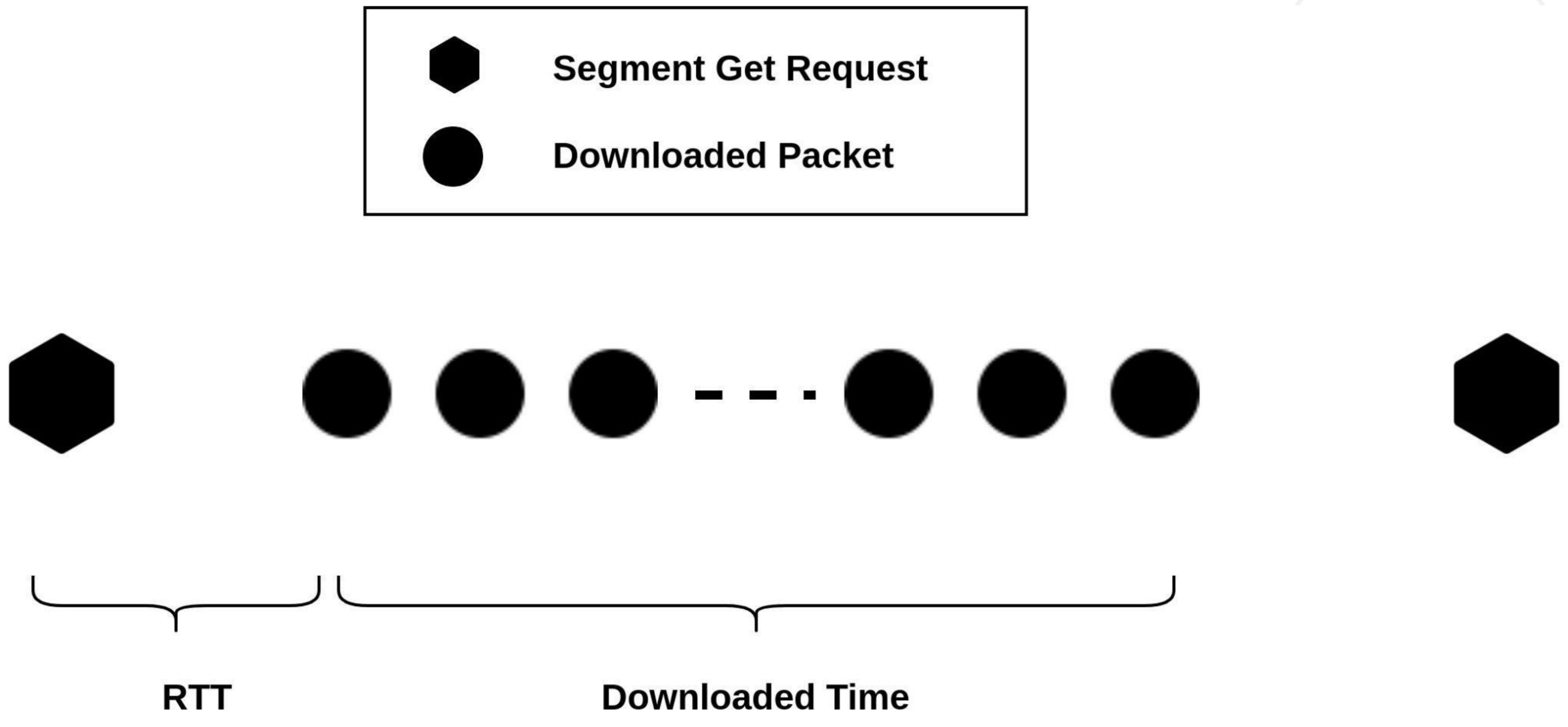
Per-segment QoS features

- RTT
- Packets
- Throughput relation to QoE



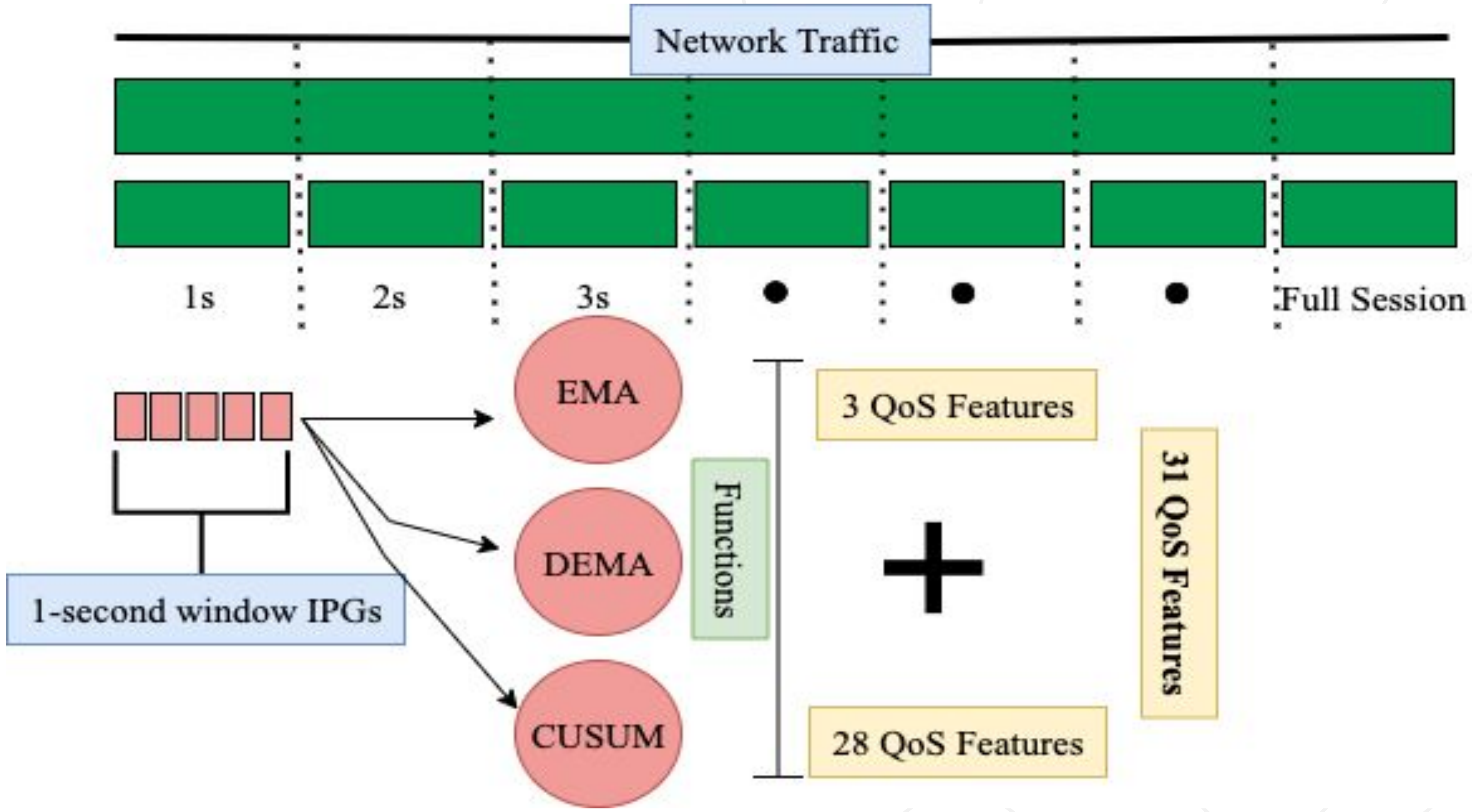
Regression - QoS Features (2/2)

1. Per-segment based QoS features.
   a. RTT
   b. Throughput
   c. Packets
2. Adaptation algorithms correlation with different network use cases.
   a. Buffered – BBA
   b. Hybrid – Elastic
   c. Throughput – Conventional



Throughput = Downloaded Packet Size / (RTT + Downloaded Time)

Classification - Encrypted Traffic (1/3)



Classification - QoS Features (2/3)

Real Time (Window)	**Comments**
Packets count (total) [w/ gt 100B]	*Ignoring ack packets of size 100B*
Packet size distribution [w, (10-90)p]	*10-90 percentile packet size distribution in a window*
Throughput [w, distribution (10-90)p]	*10-90 percentile throughput distribution in a window*
Packet Time [IPGs, Inter Arrival Time]	*Inter Packet Gap (IPGs) of a window*
IPGs features [EMA, DEMA, CUSUM]	*See the continuity of packets*
IPGs [Avg, Std, w/ gt100B]	*Average, Standard deviation of window*

Classification - ML Algorithms (3/3)

Supervised Machine Learning based model for objective assessment, such as,

- Decision Tree
- Random Forest
- K-nearest neighbors
- Artificial Neural Network (ANN)

Mapping between QoS and QoE KPIs for certain actions, i.e., resources optimization, SLA, SDN decisions.

Objective QoE

- Stalls
- Bitrate
- Shifts

QoE = Given the input features, predict QoE values

Models accuracy = Highly correlated features

Models – Regressions (continuous) and classifications (categorical)

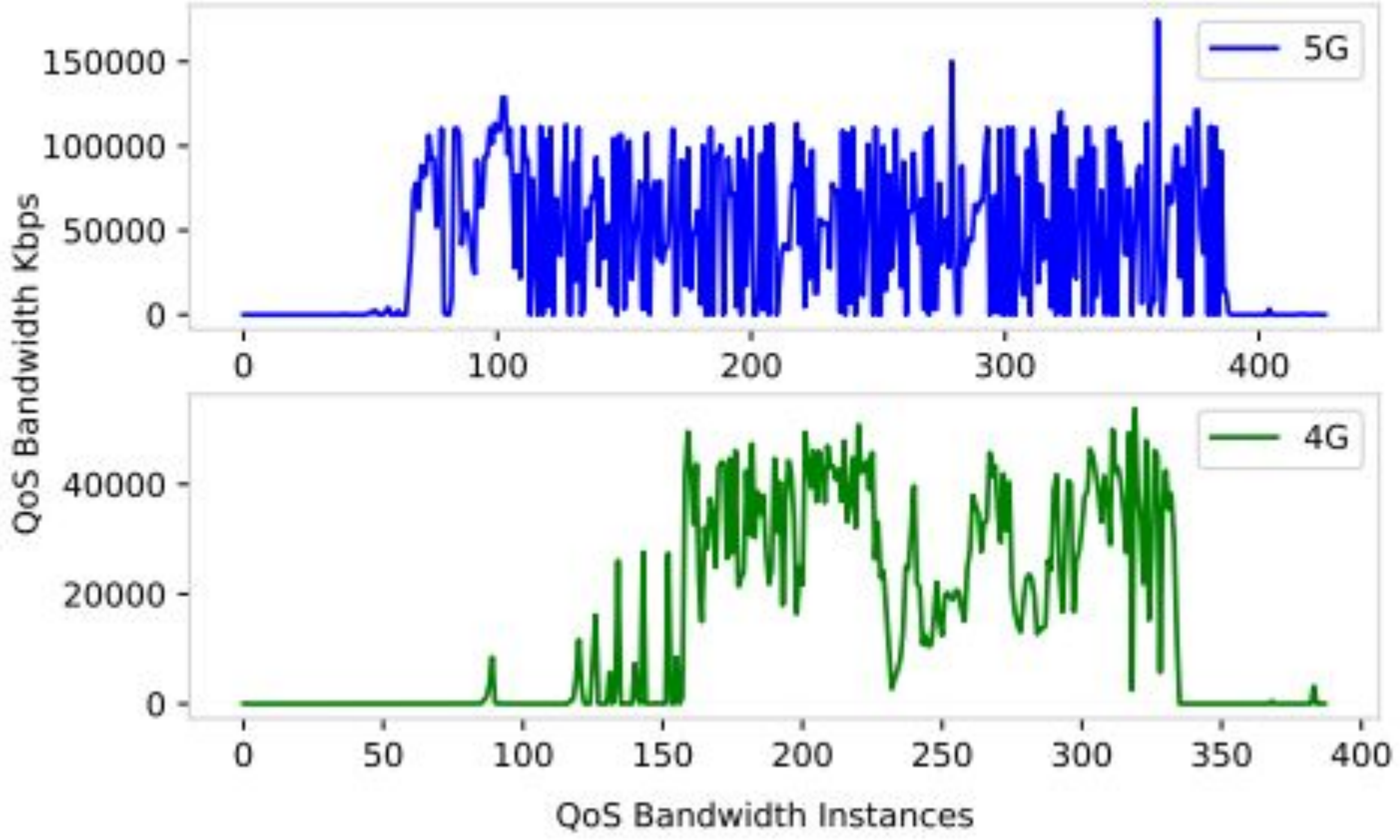
Correlation - QoS to QoE

Use Case

QoS features correlation with the objective QoE stall and quality shifts.

Shifts and stalls are the main QoE indicators *

- 5G – Blue (Real Traces)
- 4G – Green (Real Traces)



* Fan Zhang, Long Xu, and Qian Zhang. 2013. Maximum-likelihood visual quality based on additive log-logistic model. In 2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 470–475.

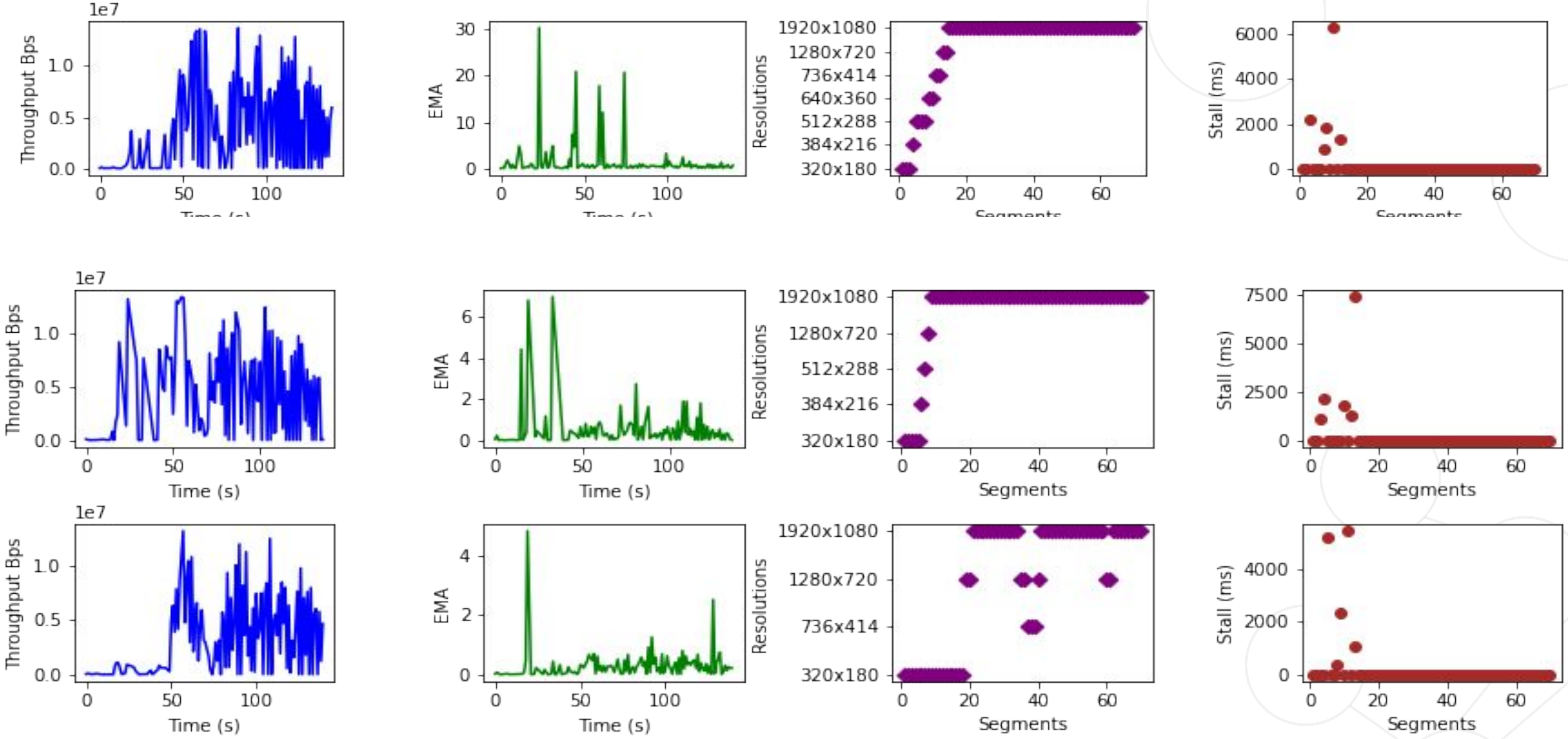
Adaptive Bitrate Streaming Algorithms

Correlation - 4G

BBA- Buffered

Conv - Throughput

Elastic: Hybrid



Adaptive Bitrate Streaming Algorithms

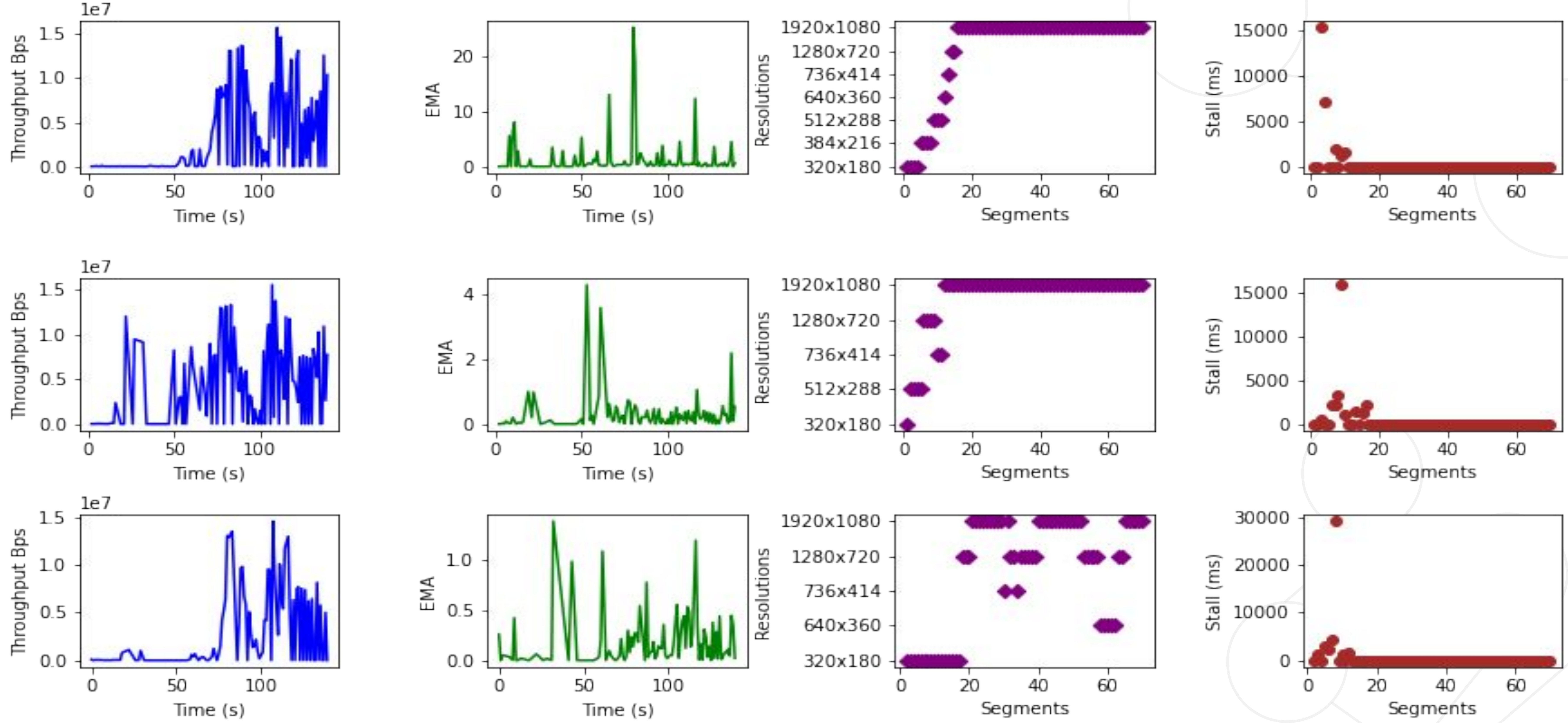
Correlation - 5G

SMARTNESS

BBA- Buffered

Conv - Throughput

Elastic: Hybrid



How does ML help correlate complex QoS data, such as RSRP and RSRQ, with user-perceived video quality on YouTube?

SMARTNESS

YouTube goes 5G: QoE Benchmarking and ML-based Stall Prediction

Raza Ul Mustafa
University of Campinas (UNICAMP)
Brazil
razaul@dca.fee.unicamp.br

Chadi Barakat
Inria, Université Côte d'Azur
France
chadi.barakat@inria.fr

Christian Esteve Rothenberg
University of Campinas (UNICAMP)
Brazil
chesteve@dca.fee.unicamp.br

***Abstract*—Given the dominance of adaptive video streaming services on the Internet traffic, understanding how YouTube Quality of Experience (QoE) relates to real 4G and 5G Channel Level Metrics (CLM) is of interest to not only the research community but also to Mobile Network Operators (MNOs) and content creators. In this context, we collect YouTube and CLM logs with 1-second granularity spanning a six-month period. We group the traces by their context, i.e., Mobility, Pedestrian, Bus/Railway terminals, and Static Outdoor, and derive key performance footprints of real 4G and 5G video streaming in the wild. We also develop Machine Learning (ML) classifiers to predict objective QoE video stalls by using past patterns from CLM traces. We release all datasets and software artifacts for reproducibility purposes.**

***Index Terms*—5G, QoS, QoE, Machine Learning, YouTube.**

I. Introduction

Mobile video traffic is continuously growing, thus adding an additional challenge for Mobile Network Operators (MNOs) to manage this exponential growth [1]. Applications utilizing social media, gaming, and recent advances in Augmented/Virtual Reality and UHD videos have accelerated the demands for the next generation of networks, 5G [2]. The New Radio (NR) of 5G technology is developed to address high bandwidth, low latency, and massive connectivity

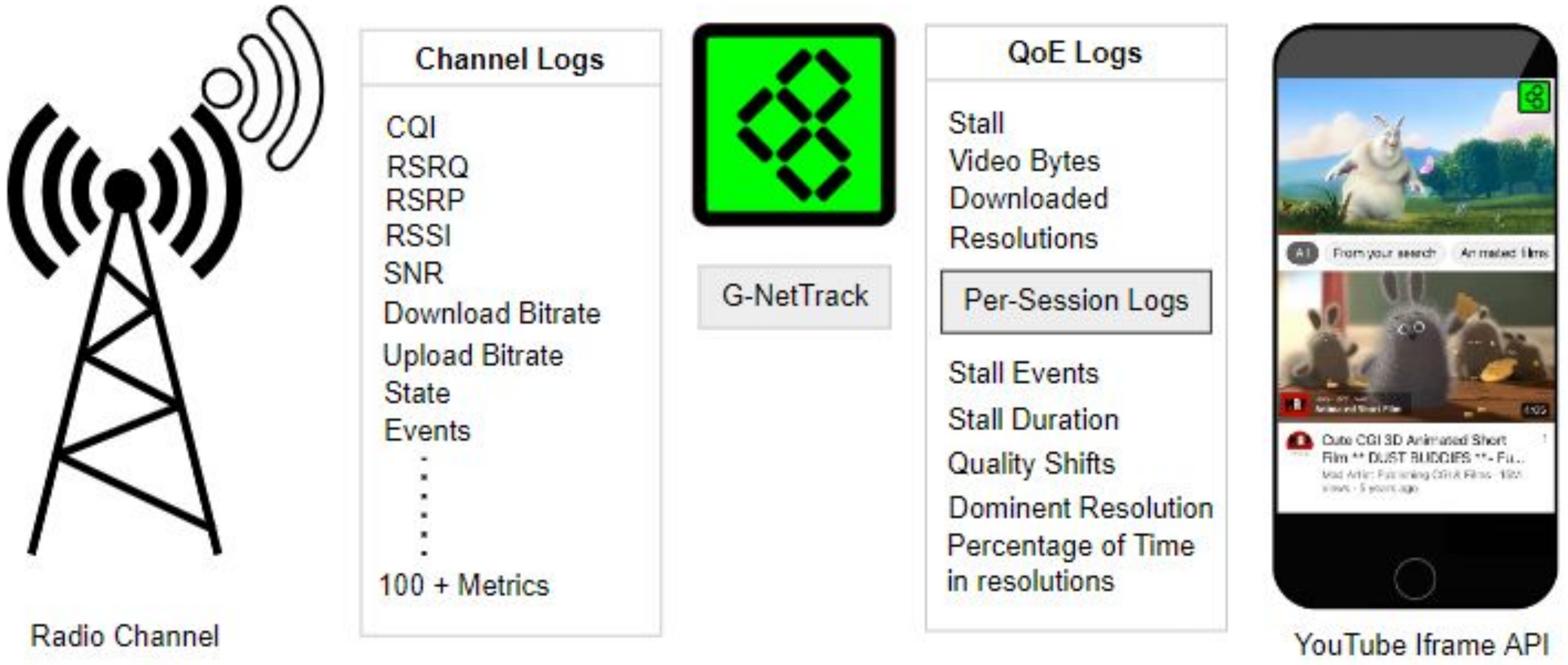
videos of diverse type (e.g., Nature, Animation, Movie, Brand Promotions) at different Frame Per Second (FPS) rates and under varying context scenarios: (*i*) Mobility, (*ii*) Pedestrian, (*iii*) Bus/Railway terminals, and (*iv*) Static Outdoor. Next, we study the relationship between the Channel Level Metrics (CLM) and objective QoE scores of YouTube. This study helps us to propose a QoE interruption (Stall) prediction method based only on CLM metrics. We carry out a rich 4G and 5G dataset collection campaign using commercial 4G and 5G networks, where we consider YouTube as a baseline for video streaming to collect CLM and YouTube QoE logs with 1-second granularity. All videos are selected from different categories such as Sports, Animated, Movies, Nature, etc. In addition, we consider videos with 4K quality and some that coded at 60 FPS. We provide detail of each video in [8]. Our contributions can be summarized as follows:

- We collect 4G and 5G datasets with channel and context using YouTube as a baseline at the smallest granularity of 1-second in a rich set of use case scenarios.
- We derive a model relating CLM measurements to video stalls using a time-based method. We check for different

Key Points

Running YouTube in multiple 4G/5G scenarios

- 5G outperforms 4G in YouTube streaming, as expected, but performance over currently deployed 5G nets is still not ideal
- 5G behaves greedily under mobility
- 5G experiences more stalling events compared to 4G in mobility
  - In 5G mobility, we observe a **16.67 %** increase in stalling events compared to those in 4G.



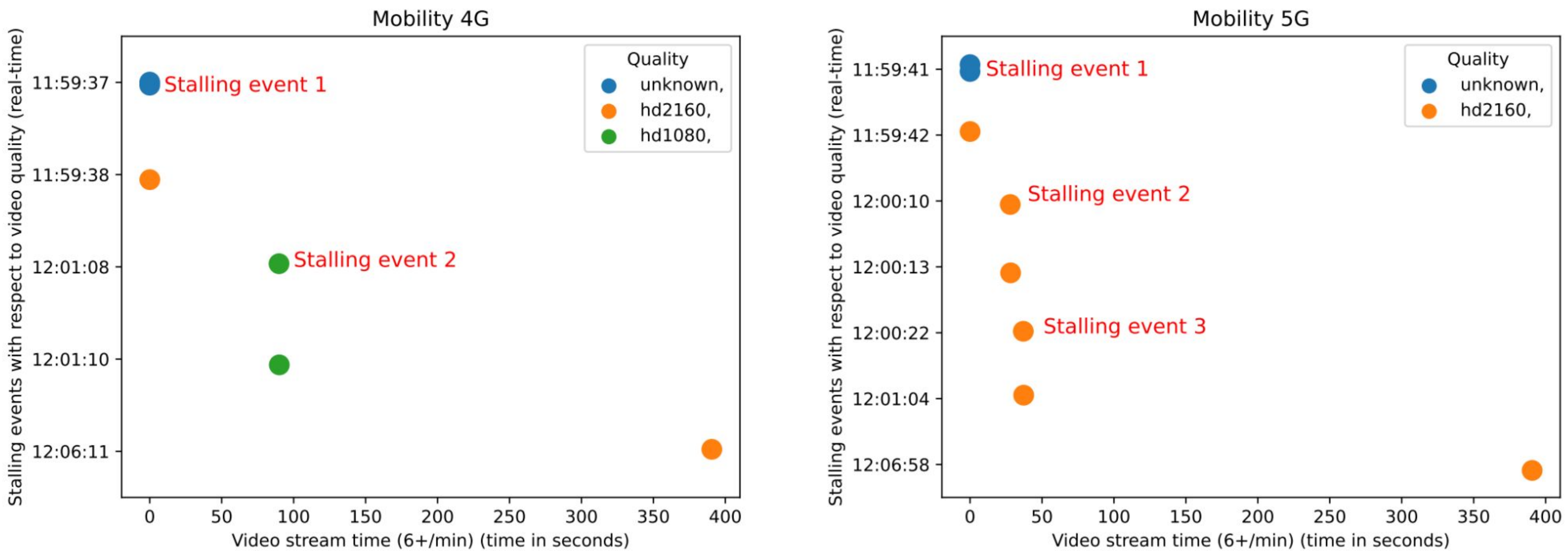
CLM - Player logs - Player events

TABLE I: CLM and their corresponding YouTube player logs and events for use case – Mobility, Technology – 5G

Channel metrics						Player logs				Player events		
Time	RSRP	RSRQ	SNR	CQI	DL_bitrate	Time	Quality	VBD	LP	Time	Quality	Event
17.43.30	-97	-3	20	13	5	17.43.30	hd2160	0.128010425	12.8	17:43:30	hd2160	buffering
17.43.31	-97	-3	20	13	61350	17.43.31	hd2160	0	0	17:43:32	hd2160	playing
17.43.32	-97	-3	20	13	86264	17.43.32	hd2160	0.08171254	8.2	-	-	-
17.43.33	-87	-3	21	13	94897	17.43.33	hd2160	0.102403732	10.2	-	-	-
17.43.34	-87	-3	21	15	4	17.43.34	hd2160	0.102403732	10.2	-	-	-

- <u>Machine Learning Classifiers Features</u>
  - Previous times (window), i.e., (1, 3, 5, 7, 9)-seconds to see if there is any correlation with the radio channel CLM, i.e., CQI, RSRQ, RSRP, SNR
  - i) Majority of a window, ii) Standard deviation, iii) 25, 50, and 75 percentile of a window.

Quality shifts – Resolutions



(a) 4G – Stalls (b) 5G – Stalls

TABLE II: 4G vs. 5G percentage of player resolutions, case – mobility and pedestrian.

Case	480p	hd720	hd1080	hd1440	hd2160
			4G		
Mobility	7.7	-	3.2	25.9	63.2
Pedestrian	0.4	3.7	33.2	24.8	37.9
			5G		
Mobility				0.2	99.8
Pedestrian		0.6	3.7	0.4	95.3

- 5G behaves greedily under mobility
  - Even with stalling events during mobility, the player remains in higher resolutions instead of choosing a segment with a lower resolution and bitrate to avoid stalls
- 5G outperformed 4G by 36.6 percent in hd2160 resolution. But! More stalling events in Mobility in 5G.

SMARTNESS

ML for Networking Use Cases

- **Casos reales**
  - Identificación de fallos suaves en redes ópticas usando Network Digital Twin

D. A. A. Mello, K. S. Mayer, A. F. Escallón-Portilla, D. S. Arantes, R. P. Pinto and C. E. Rothenberg, "**When Digital Twins Meet Optical Networks Operations**," *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, CA, USA, 2023, pp. 1-3, doi: 10.1364/OFC.2023.W4A.3.

Kayol S. Mayer, Rossano P. Pinto, Jonathan A. Soares, Dalton S. Arantes, Christian E. Rothenberg, Vinicius Cavalcante, Leonardo L. Santos, Filipe D. Moraes, and Darli A. A. Mello, "**Demonstration of ML-Assisted Soft-Failure Localization Based on Network Digital Twins,**" J. Lightwave Technol. 40, 4514-4520 (2022)

BR102023017526-0. "Método para estimar a qualidade de transmissão em redes ópticas e mídia de armazenamento legível por computador." Kayol Mayer, Darli Mello, Marcos Almeida, Humberto Melo, Rossano Pinto, Christian Rothenberg e Dalton Soares.

What Problems in Optical Communications and Networks need ML?

H. Bock, Infinera, OSA PC, 2020

"NOC-less operation"

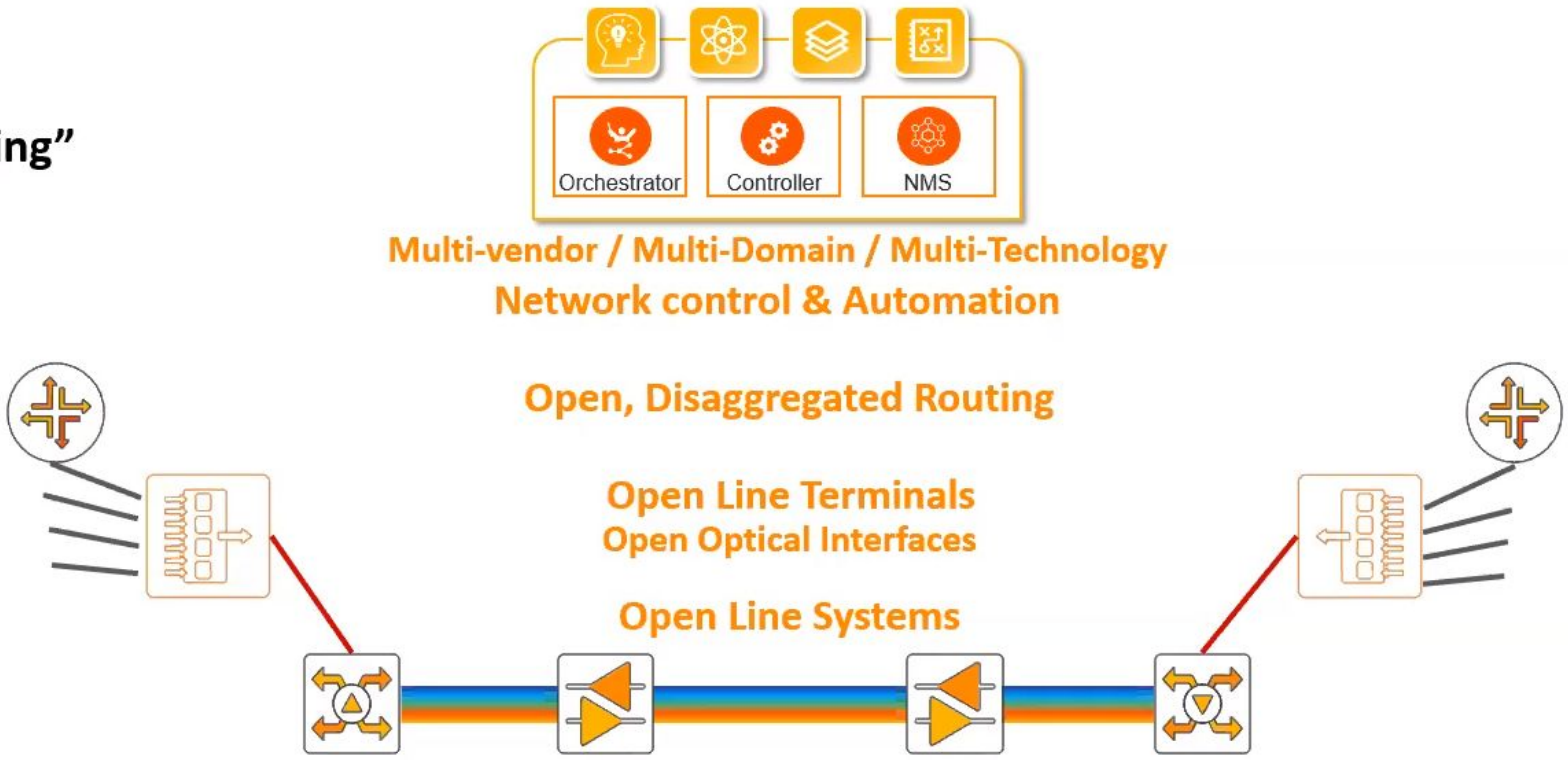
"Network Analytics"
- "Pre-emptive Maintenance"
- "Fault-identification / Troubleshooting"
- "Health check"

"Network Recovery"

"Traffic Prediction"
- "Predictive load balancing"
- "Optimized power utilization"

"Network Defragmentation"

"Reach x capacity optimization"



What Problems in Optical Communications and Networks need ML?

H. Bock, Infinera, OSA PC, 2020

"NOC-less operation"

"Network Analytics"

"Pre-emptive Maintenance"

"Fault-identification / Troubleshooting"

"Health check"

Failure Management

OPM

"Network Recovery"

"Traffic Prediction"

"Predictive load balancing"

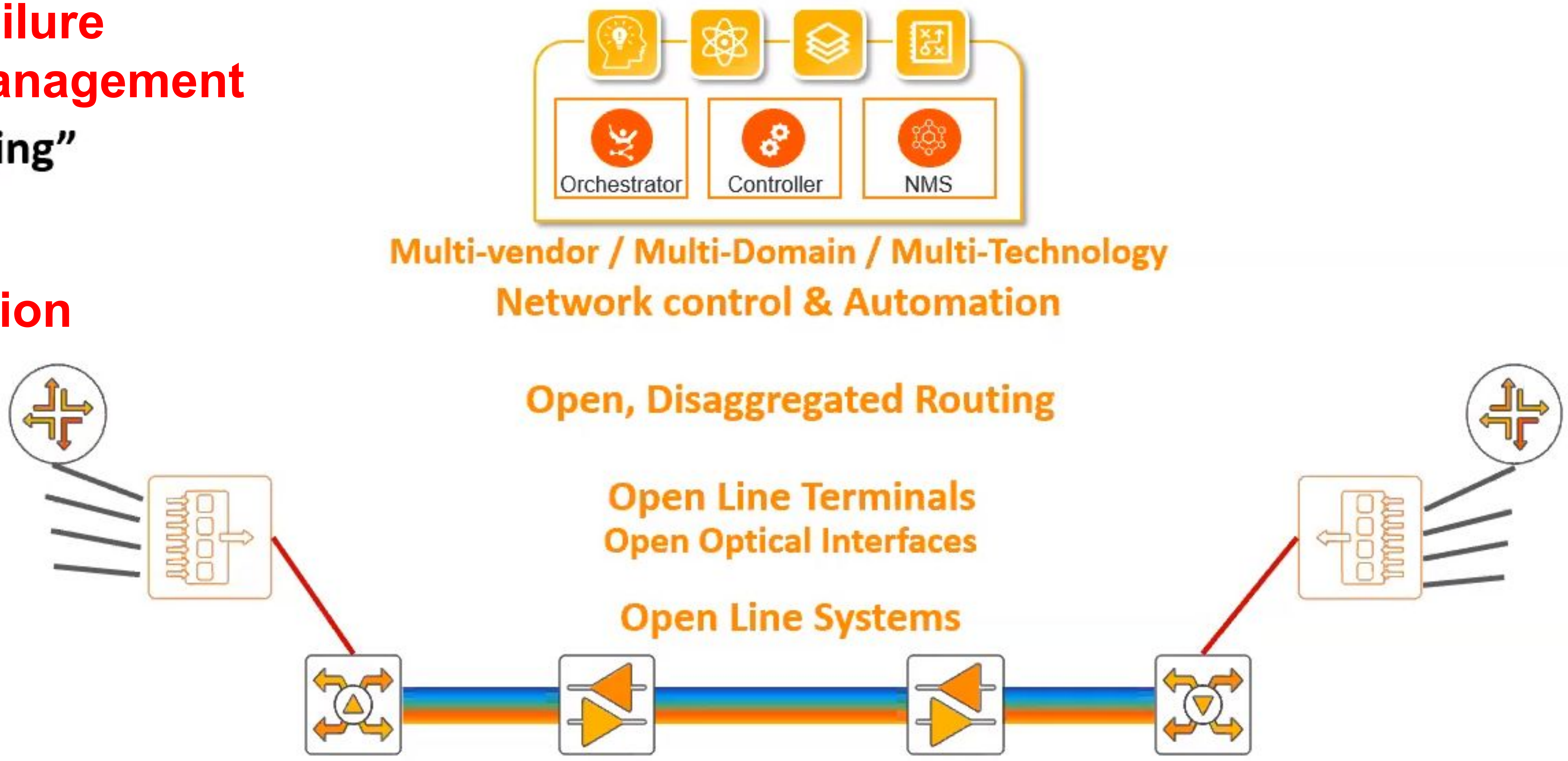
"Optimized power utilization"

Traffic Prediction

"Network Defragmentation"

"Reach x capacity optimization"

QoT



What problems in optical communications and networks need ML?

AI is a high-quality prediction technology, getting better every day

Prediction is the process of filling missing information		
Past	**Present**	**Future**
Predicting YouTube video has cats (pattern detection with CNNs)	Predicting a driving action to be taken (reinforcement learning with DNNs)	Predicting future stocks prices (forecasting with RNNs)

AI can find "missing information" in optical networks		
Planning & Design	**Operations**	**Maintenance**
• Network traffic demands • Expected QoT for new waves **QoT** • Physical parameters for unlit paths • Fiber power response • Etc.	• Future QoT of existing waves **OPM** • Root cause of subpar QoT • Undetected anomalies • Future SNR margin • Etc.	• Future equipment failures • Future fiber cuts • When fibers will exhaust • Dirty/loose connectors • Etc. **Failure Management**

P. Djukic, Ciena, OSA PC, 2020

UNICAMP

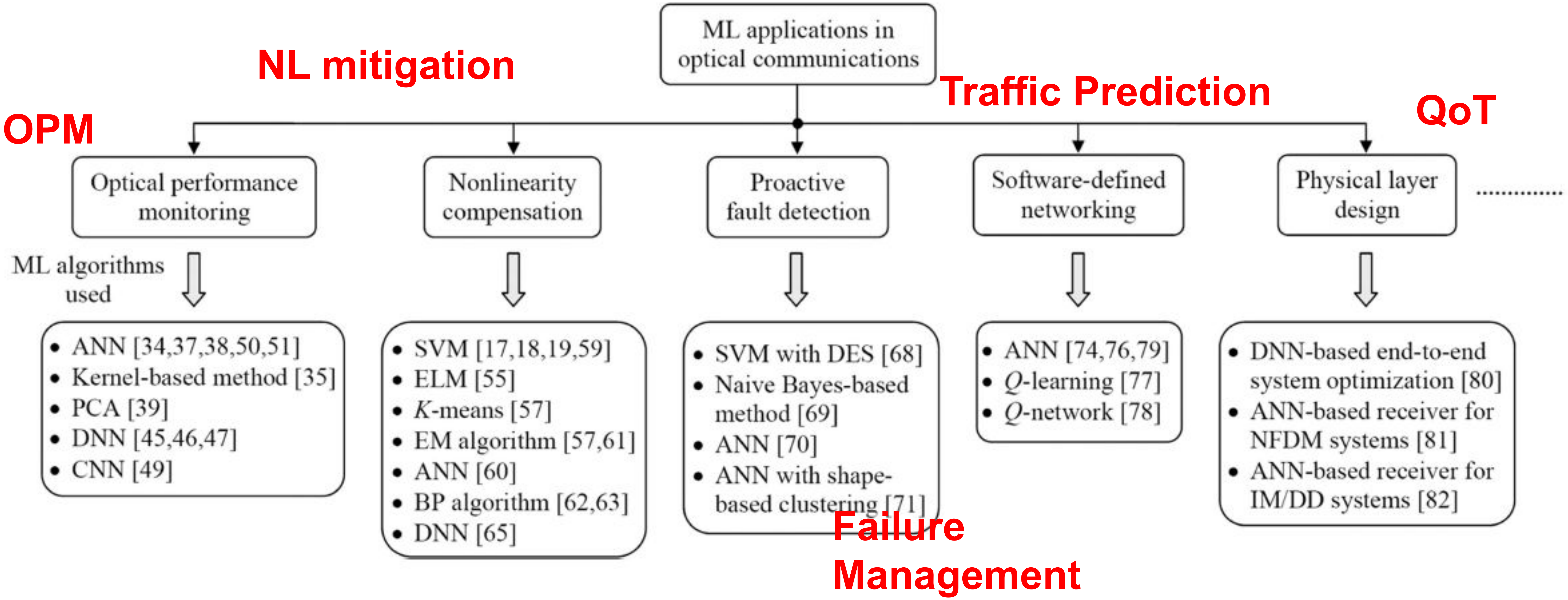
What problems in optical communications and networks need ML?

JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 37, NO. 2, JANUARY 15, 2019 493

An Optical Communication's Perspective on Machine Learning and Its Applications

Faisal Nadeem Khan, Qirui Fan, Chao Lu, and Alan Pak Tao Lau

(Invited Paper)



What problems in optical communications and networks need ML?

An Overview on Application of Machine Learning Techniques in Optical Networks

Francesco Musumeci, *Member, IEEE,* Cristina Rottondi, *Member, IEEE,* Avishek Nag, *Member, IEEE,* Irene Macaluso, Darko Zibar, *Member, IEEE,* Marco Ruffini, *Senior Member, IEEE,* and Massimo Tornatore, *Senior Member, IEEE*

F. Musumeci *et al*., in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383-1408, 2018.

UNICAMP

What problems in optical communications and networks need ML?

- Optical Transmission

TABLE I: Different use cases at physical layer and their characteristics.

QoT

OPM

Use Case	ML category	ML methodology	Input data	Output data	Training data	Ref.
QoT estimation	supervised	kriging, L_2-norm minimization	OSNR (historical data)	OSNR	synthetic	[34]
			OSNR/Q-factor	BER	synthetic	[35], [36]
			OSNR/PMD/CD/SPM	blocking prob.	synthetic	[37]
		CBR	error vector magnitude, OSNR	Q-factor	real	[38]
			lightpath route, length, number of co-propagating lightpaths	Q-factor	synthetic	[39], [40]
		RF	lightpath route, length, MF, traffic volume	BER	synthetic	[41]
		regression	SNR (historical data)	SNR	synthetic	[42]
		NN	lightpath route and length, number of traversed EDFAs, degree of destination, used channel wavelength	Q-factor	synthetic	[43], [44]
		k-nearest neighbor, RF, SVM	total link length, span length, channel launch power, MF and data rate	BER	synthetic	[45]
		NN	channel loadings and launch power settings	Q-factor	real	[46]
		NN	source-destination nodes, link occupation, MF, path length, data rate	BER	real	[47]
OPM	supervised	NN	eye diagram and amplitude histogram param.	OSNR/PMD/CD	real	[48]
		NN, SVM	asynchronous amplitude histogram	MF	real	[49]
		NN	asyncrhonous constellation diagram and amplitude histogram param.	OSNR/PMD/CD	synthetic	[50]–[53]
		Kernel-based ridge regression	eye diagram and phase portraits param.	PMD/CD	real	[54]
		NN	Horizontal and Vertical polarized I/Q samples from ADC	OSNR, MF, symbol rate	real	[55]
		Gaussian Processes	monitoring data (OSNR vs λ)	Q-factor	real	[56]
Optical amplifiers control	supervised	CBR	power mask param. (NF, GF)	OSNR	real	[57], [58]
		NNs	EDFA input/output power	EDFA operating point	real	[59], [60]
		Ridge regression, Kernelized Bayesian regr.	WDM channel usage	post-EDFA power discrepancy	real	[61]
	unsupervised	evolutional alg.	EDFA input/output power	EDFA operating point	real	[62]

F. Musumeci *et al.*, in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383-1408, 2018.

UNICAMP

What problems in optical communications and networks need ML?

- Optical Transmission

MF recognition	unsupervised	6 clustering alg.	Stokes space param.	MF	synthetic	[63]
		k-means	received symbols	MF	real	[64]
	supervised	NN	asynchronous amplitude histogram	MF	synthetic	[65]
		NN, SVM	asynchronous amplitude histogram	MF	real	[66], [67], [49]
		variational Bayesian techn. for GMM	Stokes space param.	MF	real	[68]
Non-linearity mitigation	supervised	Bayesian filtering, NNs, EM	received symbols	OSNR, Symbol error rate	real	[31], [32], [69]
		ELM	received symbols	self-phase modulation	synthetic	[70]
		k-nearest neighbors	received symbols	BER	real	[71]
		Newton-based SVM	received symbols	Q-factor	real	[72]
		binary SVM	received symbols	symbol decision boundaries	synthetic	[73]
		NN	received subcarrier symbols	Q-factor	synthetic	[74]
		GMM	post-equalized symbols	decoded symbols with impairment estimated and/or mitigated	real	[75]
		Clustering	received constellation with nonlinearities	nonlinearity mitigated constellation points	real	[76]
		NN	sampled received signal sequences	equalized signal with reduced ISI	real	[77]–[82]
	unsupervised	k-means	received constellation	density-based spatial constellation clusters and their optimal centroids	real	[83]

NL mitigation

F. Musumeci *et al.*, in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383-1408, 2018.

UNICAMP

What problems in optical communications and networks need ML?

- Optical Networking

Traffic Prediction

Failure Management

Use Case	ML category	ML methodology	Input data	Output data	Training data	Ref.
Traffic prediction and virtual topology (re)design	supervised	ARIMA	historical real-time traffic matrices	predicted traffic matrix	synthetic	[84], [85]
		NN	historical end-to-end maximum bit-rate traffic	predicted end-to-end traffic	synthetic	[86], [87]
		Reinforcement learning	previous solutions of a multi-objective GA for VTD	updated VT	synthetic	[88], [89]
		Recurrent NN	historical aggregated traffic at different BBU pools	predicted BBU pool traffic	real	[90]
		NN	historical traffic in intra-DC network	predicted intra-DC traffic	real	[91]
	unsupervised	NMF, clustering	CDR, PoI matrix	similarity patterns in base station traffic	real	[92]
Failure management	supervised	Bayesian Inference	BER, received power	list of failures for all lightpaths	real	[93]
		Bayesian Inference, EM	FTTH network dataset with missing data	complete dataset	real	[94], [95]
		Kriging	previously established lightpaths with already available failure localization and monitoring data	estimate of failure localization at link level for all lightpaths	real	[96]
		(1) LUCIDA: Regression and classification (2) BANDO: Anomaly Detection	(1) LUCIDA: historic BER and received power, notifications from BANDO (2) BANDO: maximum BER, threshold BER at set-up, monitored BER	(1) LUCIDA: failure classification (2) BANDO: anomalies in BER	real	[97]
		Regression, decision tree, SVM	BER, frequency-power pairs	localized set of failures	real	[98]
		SVM, RF, NN	BER	set of failures	real	[99]
		regression and NN	optical power levels, amplifier gain, shelf temperature, current draw, internal optical [illegible]	detected faults	real	[100]

F. Musumeci *et al.*, in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383-1408, 2018.

UNICAMP

Digital Twins in Optical Networks: Insights and Applications

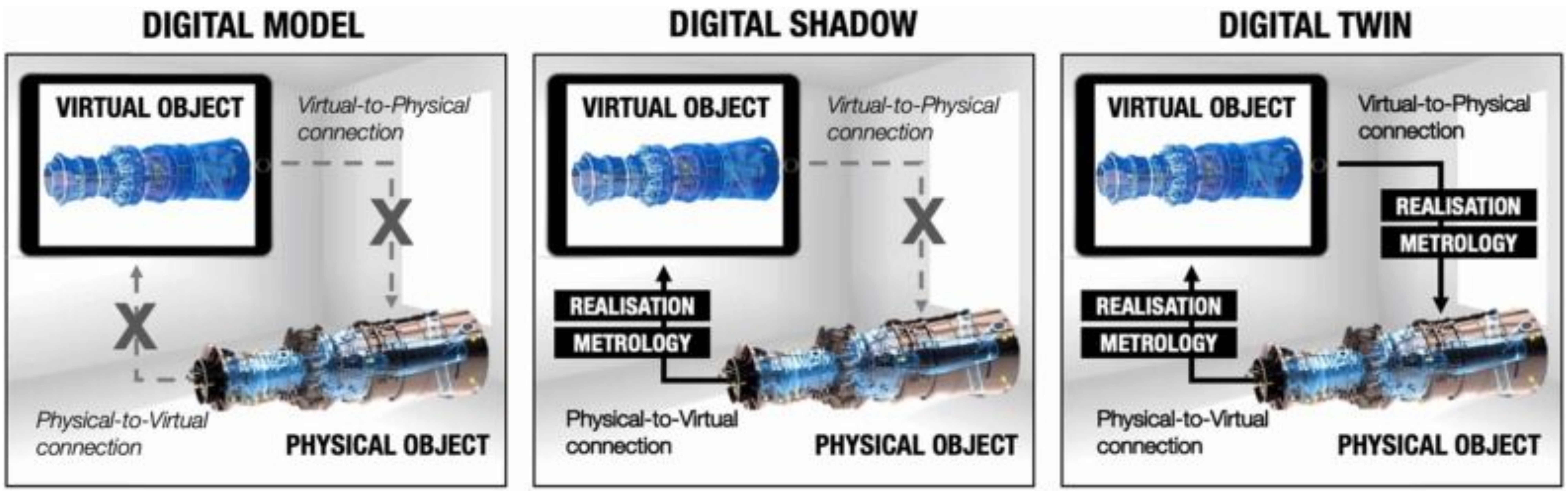
Darli A. A. Mello
Kayol Mayer
Rossano Pinto
Christian Rothenberg
Dalton Arantes

School of Electrical and Computer Engineering (FEEC)
University of Campinas (UNICAMP)
Brazil

UNICAMP

What is a Network Digital Twin?

- But first, what is a digital twin? It is much more than a digital model or a simulator …



M. Bertoni, A. Bertoni, (2022), Designing solutions with the product-service systems digital twin: What is now and what is next?, Computers in Industry, Volume 138, https://doi.org/10.1016/j.compind.2022.103629.

W. Kritzinger, M. Karner, G. Traar, J. Henjes, W. Sihn, Digital Twin in manufacturing: A categorical literature review and classification,IFAC-PapersOnLine, Volume 51, Issue 11, 2018, Pages 1016-1022.

What is a Network Digital Twin?

- Network Digital Twin (ou Digital Twin Network)

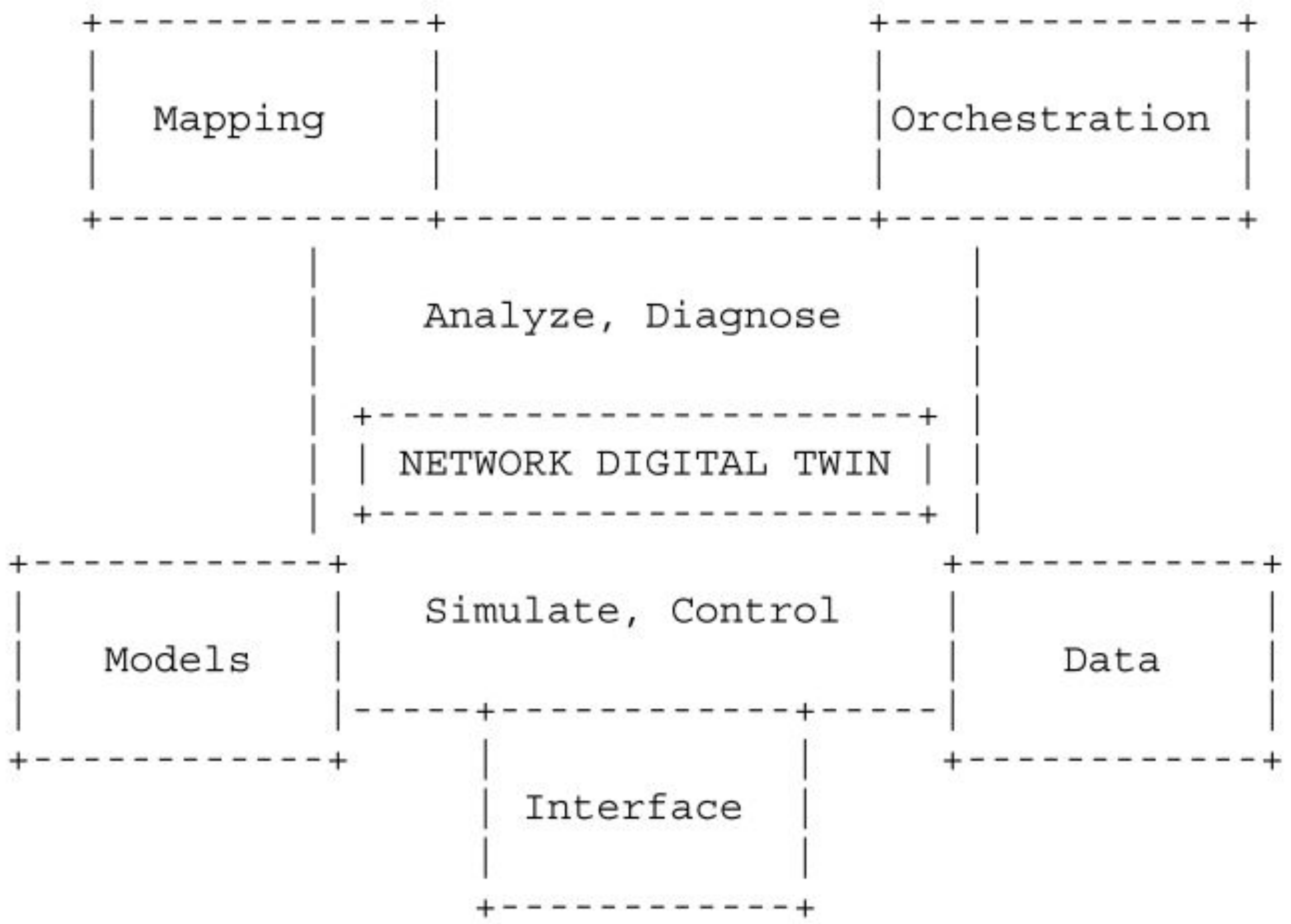
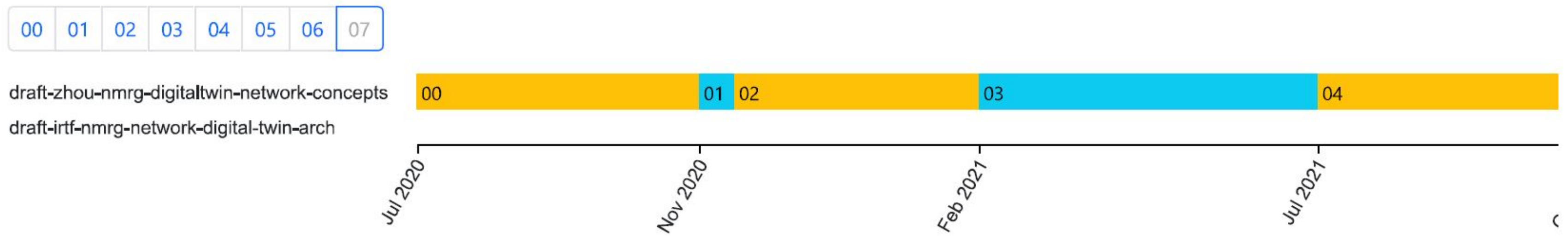


Figure 1: Key Elements of Digital Twin Network

Digital Twin Network: Concepts and Reference Architecture
draft-zhou-nmrg-digitaltwin-network-concepts-07

Status | IESG evaluation record | IESG writeups | Email expansions | History

Versions:



Document Type Replaced Internet-Draft (individual)

Authors Cheng Zhou, Hongwei Yang, Xiaodong Duan, Diego Lopez, Antonio Pastor, Qin Wu, Mohamed Boucadair, Christian Jacquenet

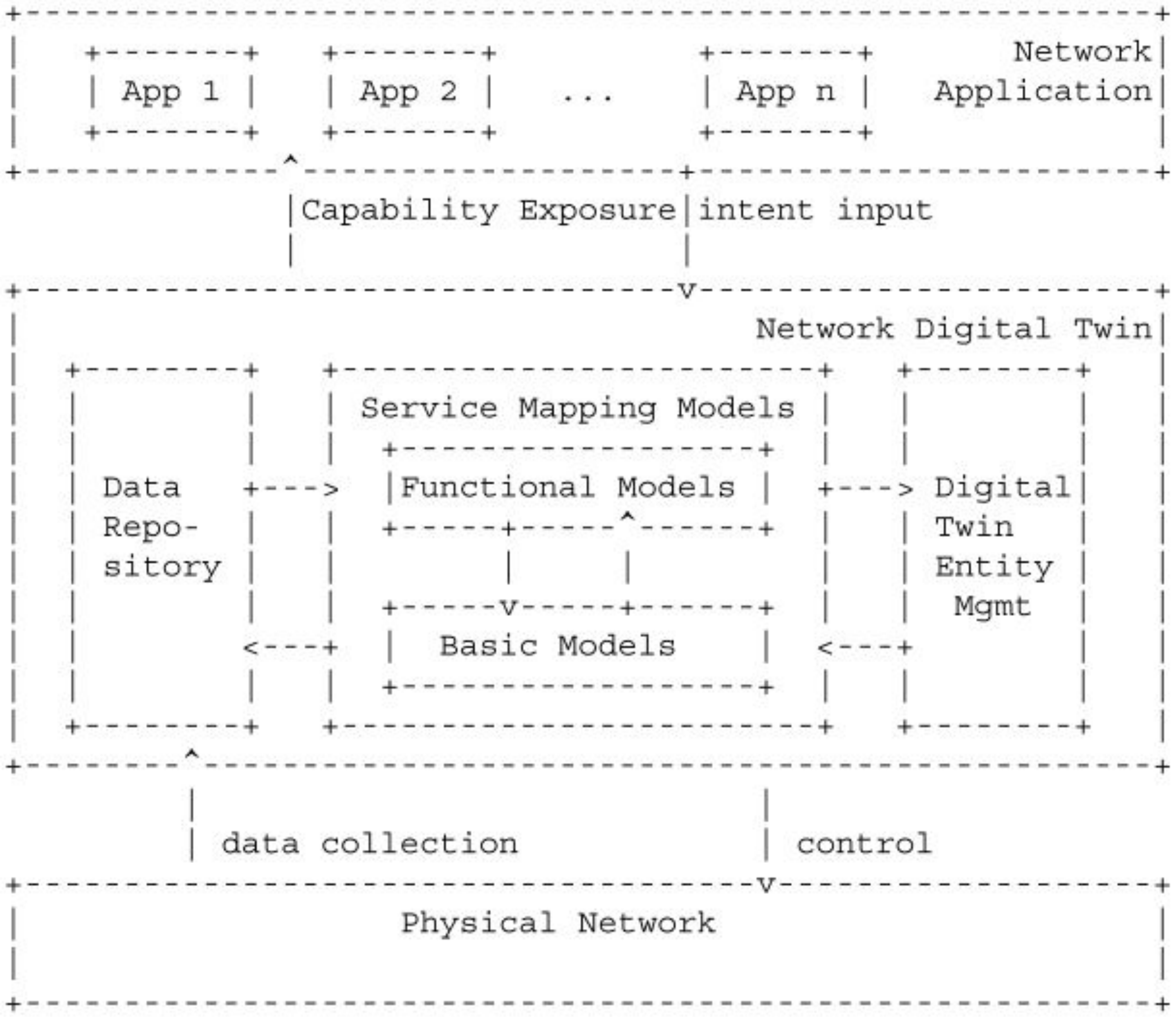


Figure 2: Reference Architecture of Digital Twin Network

Source: IRTF draft-zhou-nmrg-digitaltwin-network-concepts-00

What is a Network Digital Twin?


IEEE/IFIP Network Operations and Management Symposium
25-29 April 2022 // Budapest, Hungary
Network and Service Management in the Era of Cloudification, Softwarization and Artificial Intelligence

IEEE ComSoc ifip IEEE

HOME | ABOUT | COMMITTEES | AUTHORS | PROGRAM | REGISTRATION | HOTEL / TRAVEL | PATRONS / EXHIBITORS | Search

WS4 - 1ST INTERNATIONAL WORKSHOP ON TECHNOLOGIES FOR NETWORK TWINS (TNT 2022)

All times listed are in Central European Summer Time (CEST)

TNT 2022 Program

Time - CEST	Duration	Session	Chair
09:00	45 min.	**Opening and Keynote** Keynote speaker: Adam Drobot, OpenTechWorks Title: Network Requirements for Digital Twins and Digital Twin Requirements for Networks	Laurent Clavaglia
09:45	45 min.	**Invited talks**	Diego Lopez
	15 min.	Building a digital twin network leveraging model orchestration, Hongwei Yang – China Mobile *(remote)*	
	15 min.	An Industrial Network Digital Twin for enhanced cyber-security, Andrea Melis – Università di Bologna *(remote)*	
	15 min.	MTV: A Network Emulator for Digital Twins, Will Fantom – Lancaster University *(on site)*	
10:30	30 min.	*Coffee break*	
11:00	75 min.	**Technical session 1**	Roberto Minerva
	25 min.	Smart DC: An AI and Digital Twin-based Energy-Saving Solution for Data Centers, Ziting Zhang *(remote)*	
	25 min.	B5GEMINI: a Digital Twin Network for 5G and Beyond, Alberto Mozo Velasco *(remote)*	
	25 min.	Digital Twin for the Optical Network: Key Technologies and Enabled Automation Applications, Chris Janz *(on site)*	
12:30	60 min.	*Lunch break*	
13:30	75 min.	**Technical session 2**	Laurent Clavaglia
	25 min.	Stopping the Data Flood: Post-Shannon Traffic Reduction in Digital-Twins Applications, Caspar von Lengerke *(on site)*	
	25 min.	Accelerating Deep Reinforcement Learning for Digital Twin Network Optimization with Evolutionary Strategies, Carlos Güemes Palau *(on site)*	
	25 min.	A Chaos Engineering Approach for Improving the Resiliency of IT Services Configurations, Filippo Poltronieri *(on site)*	
15:00	30 min.	*Coffee break*	
		Panel on Hot Topics in Network Digital Twins	

https://noms2022.ieee-noms.org/ws4-1st-international-workshop-technologies-network-twins-tnt-2022

1ST INTERNATIONAL WORKSHOP ON TECHNOLOGIES FOR NETWORK TWINS (TNT 2022)

UNICAMP

What is a Network Digital Twin?

- Digital Twins in Optical Networks

OPTICAL COMMUNICATIONS AND NETWORKS

The Role of Digital Twin in Optical Communication: Fault Management, Hardware Configuration, and Transmission Simulation

Danshi Wang, Zhiguo Zhang, Min Zhang, Meixia Fu, Jin Li, Shanyong Cai, Chunyu Zhang, and Xue Chen

IEEE Comm. Magazine 2021

Architecture to Deploy and Operate a Digital Twin Optical Network

R. Vilalta[1], R. Casellas[1], Ll. Gifre[1] R. Muñoz[1], R. Martínez[1],
A. Pastor[2] , D. López[2], J.P. Fernández-Palacios[2]

[1] Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels (Barcelona), Spain
[2] Telefónica I+D, Madrid, Spain
e-mail: ricard.vilalta@cttc.es

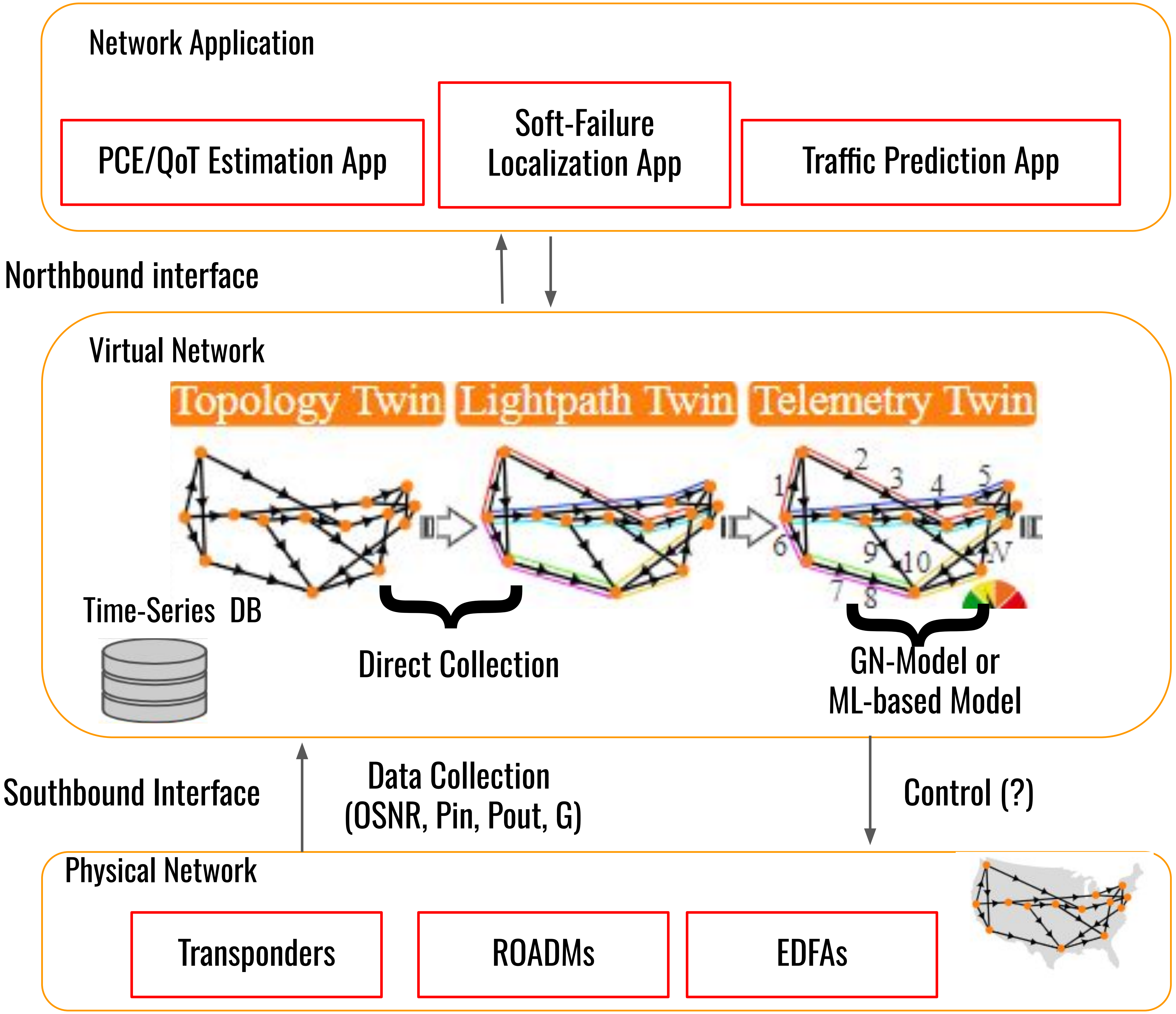
OFC 2022

UNICAMP

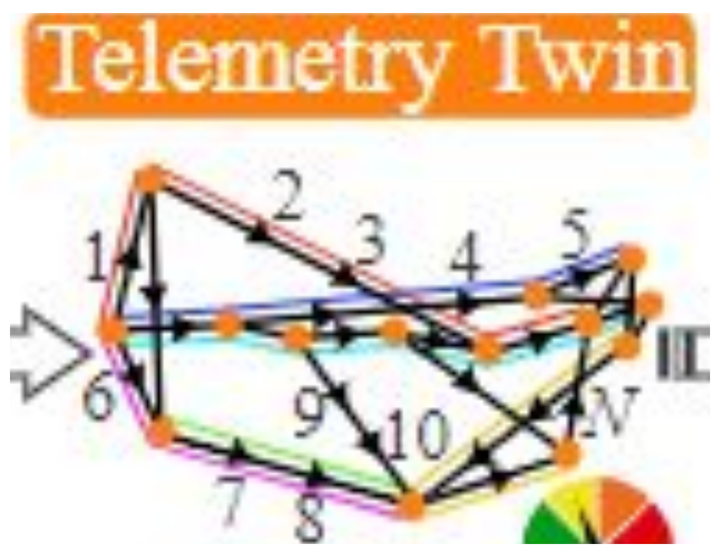
NDT in Optical Networks

```
+----------------------------------------------------------------+
|   +-------+   +-------+          +-------+          Network|
|   | App 1 |   | App 2 |   ...    | App n |      Application|
|   +-------+   +-------+          +-------+                 |
+-------------^--------------------+--------------------------+
              |Capability Exposure|intent input
              |                   |
+-------------------------------v--------------------------------+
|                                        Network Digital Twin|
|  +--------+  +--------------------------+  +--------+     |
|  |        |  | Service Mapping Models   |  |        |     |
|  |        |  |  +------------------+    |  |        |     |
|  | Data   +--->  |Functional Models |   +---> Digital|     |
|  | Repo-  |  |  +-----+-----^------+    |  | Twin   |     |
|  | sitory |  |        |     |           |  | Entity |     |
|  |        |  |  +-----v-----+------+    |  |  Mgmt  |     |
|  |        <---+ |  Basic Models    |  <---+        |     |
|  |        |  |  +------------------+    |  |        |     |
|  +--------+  +--------------------------+  +--------+     |
+--------^-------------------------------------------------------+
         |                             |
         | data collection             | control
+------------------------------------v---------------------------+
|                   Physical Network                          |
|                                                             |
+----------------------------------------------------------------+
```

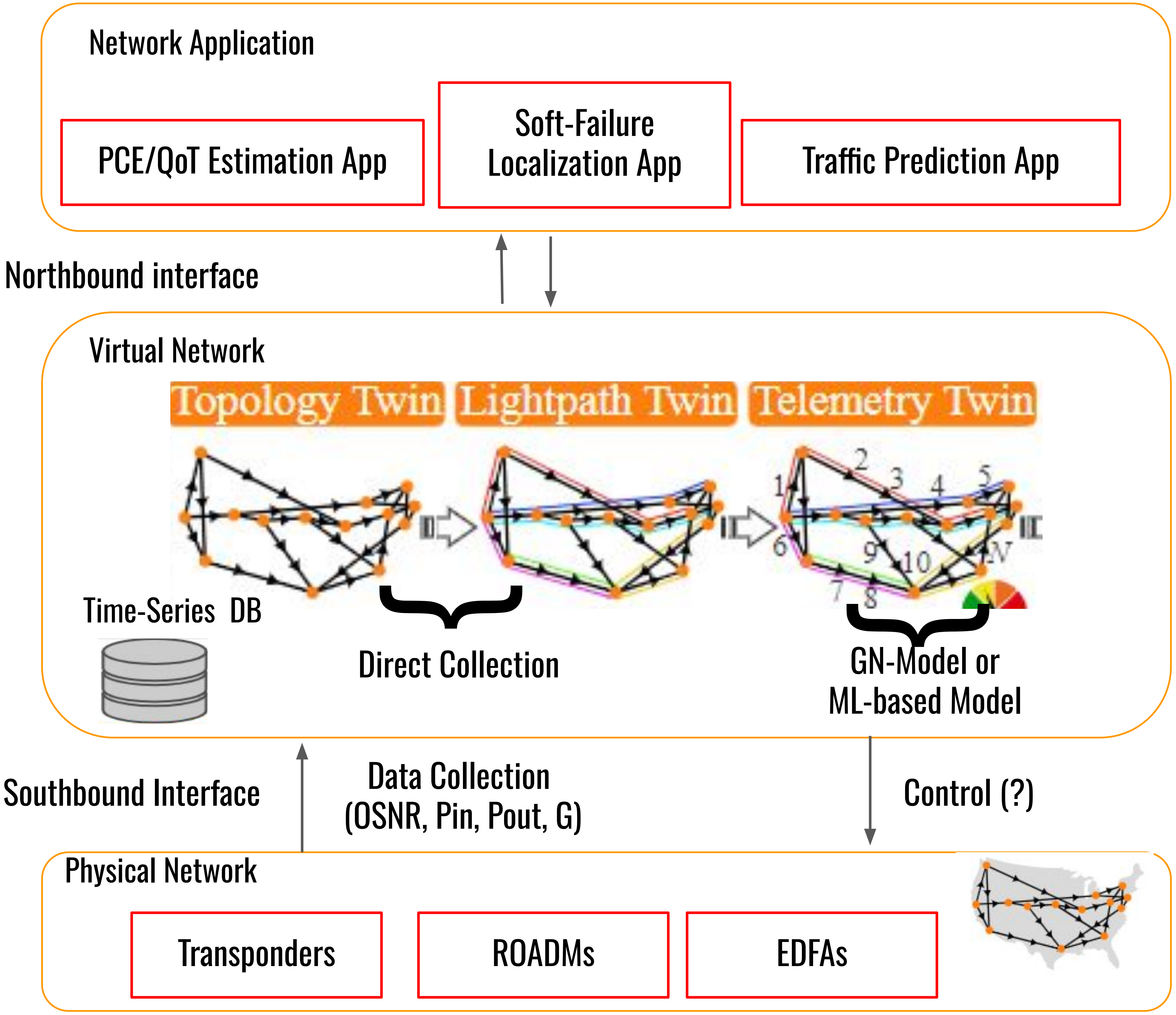
Figure 2: Reference Architecture of Digital Twin Network



NDT in Optical Networks



Parameter	Real Telemetry	Virtual Telemetry
EDFA_1_Pout	10 dBm	10.5 dBm
EDFA_1_Pin	-15 dBm	-14.7 dBm
TRX_2_OSNR	15 dB	17 dB
TRX_3_OSNR	13 dB	13 dB
TRX_5_OSNR	16 dB	14 dB



Issues and Challenges

- How to address inconsistencies between the virtual and the real network?
  - Some parameters can be read directly from the real network (e.g. fiber attenuation)
  - Other parameters must be estimated from the model: e.g. OSNR

- NDT models should be optimized to yield measured parameters
- GN-based models should yield considerable deviations (QoT estimation discussion!)
- ML-based models
- Numerically-optimized models (e.g. gradient descent algorithm)

Issues and Challenges

- How often to update the virtual network?

UNICAMP

Issues and Challenges

- Other issues (IETF)
  - Large-scale challenge (scalability, storage, data compression)
  - Interoperability
  - Data modelling
  - Real-time requirements
  - Security risks

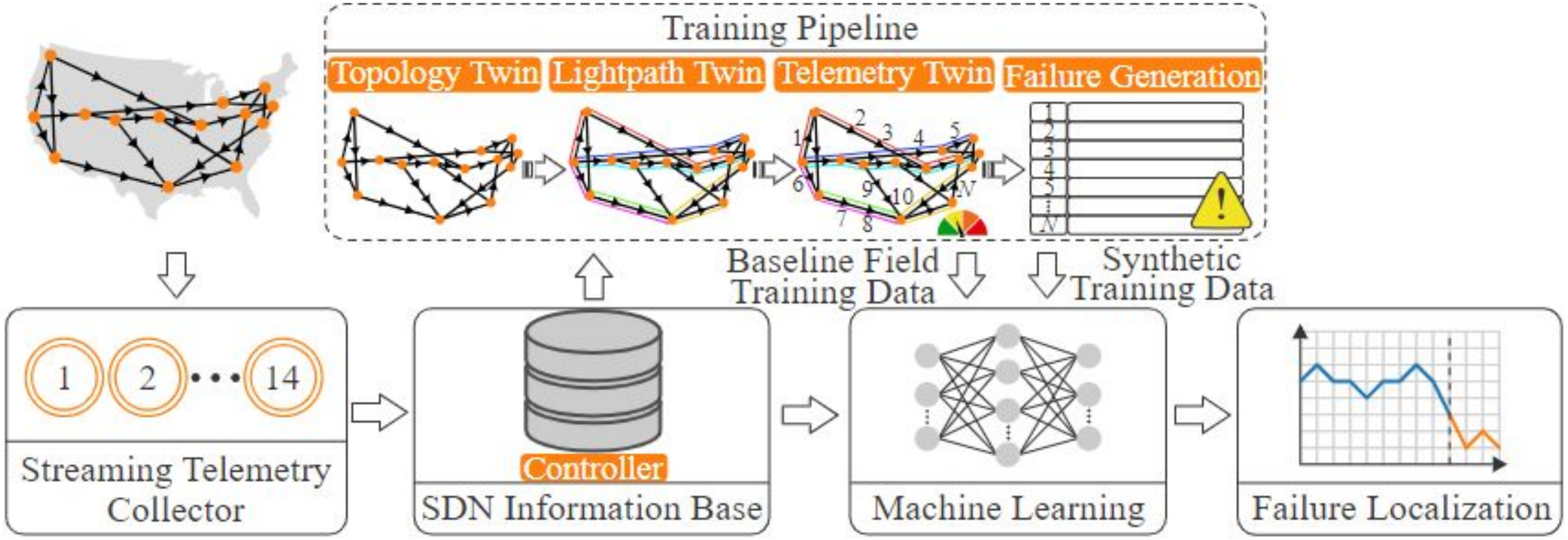
UNICAMP

Case Study: ML-based Soft-Failure Localization

4514 JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 40, NO. 14, JULY 15, 2022

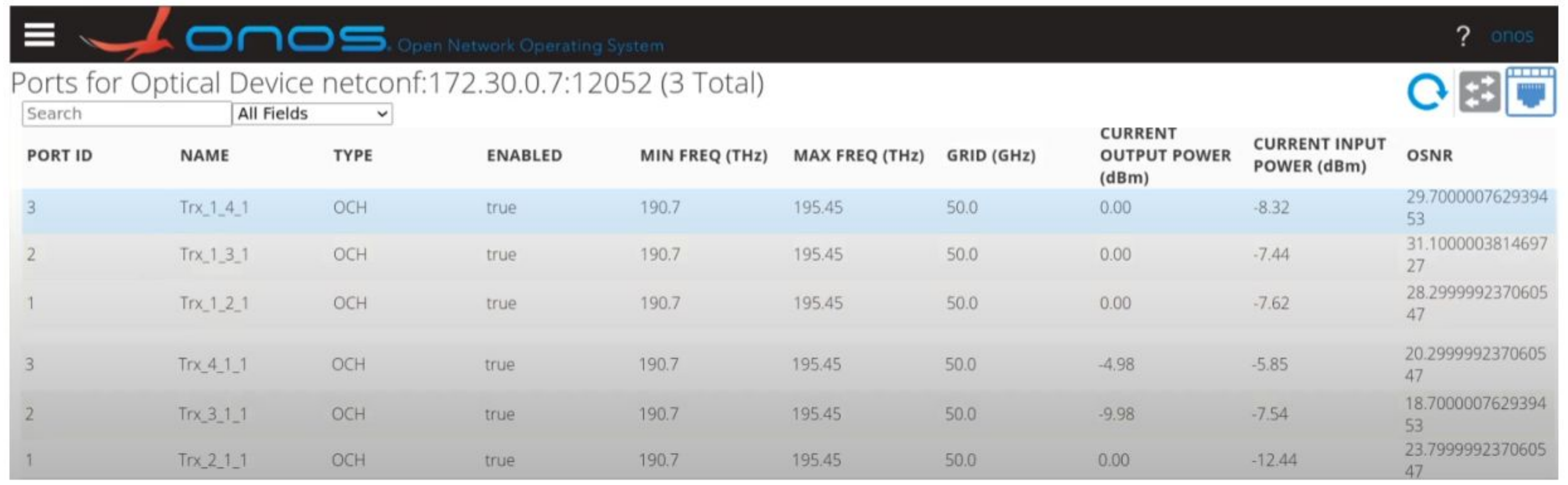
Demonstration of ML-Assisted Soft-Failure Localization Based on Network Digital Twins

Kayol S. Mayer, Rossano P. Pinto, Jonathan A. Soares, Dalton S. Arantes, Christian E. Rothenberg, Vinicius Cavalcante, Leonardo L. Santos, Filipe D. Moraes, and Darli A. A. Mello



- **Lesson learned:** for soft-failure localization, baseline training is the secret for proper algorihtm performance!

Case Study: ML-based Soft-Failure Localization

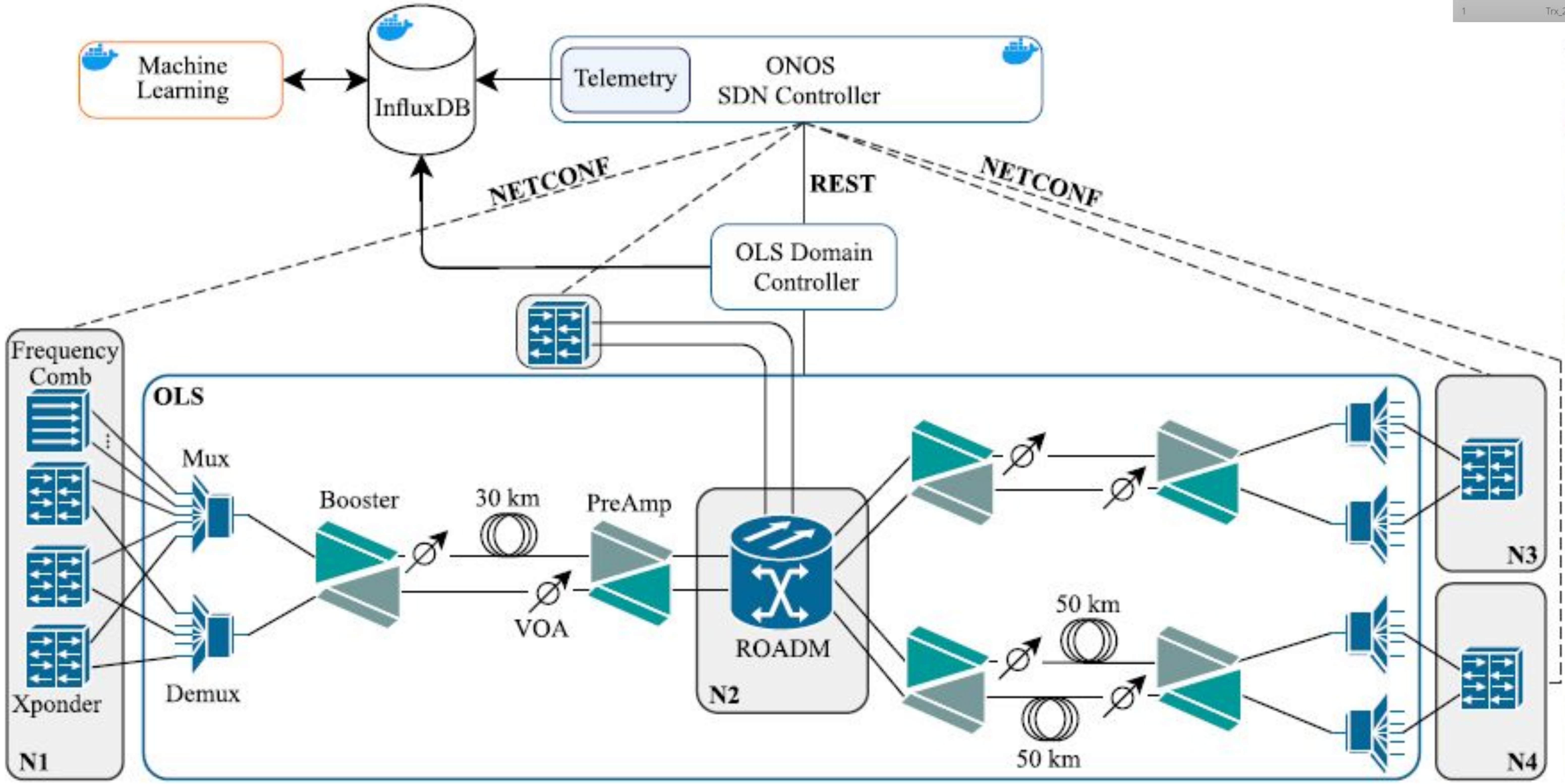


onos Open Network Operating System

Ports for Optical Device netconf:172.30.0.7:12052 (3 Total)

PORT ID	NAME	TYPE	ENABLED	MIN FREQ (THz)	MAX FREQ (THz)	GRID (GHz)	CURRENT OUTPUT POWER (dBm)	CURRENT INPUT POWER (dBm)	OSNR
3	Trx_1_4_1	OCH	true	190.7	195.45	50.0	0.00	-8.32	29.700000762939453
2	Trx_1_3_1	OCH	true	190.7	195.45	50.0	0.00	-7.44	31.100000381469727
1	Trx_1_2_1	OCH	true	190.7	195.45	50.0	0.00	-7.62	28.299999237060547
3	Trx_4_1_1	OCH	true	190.7	195.45	50.0	-4.98	-5.85	20.299999237060547
2	Trx_3_1_1	OCH	true	190.7	195.45	50.0	-9.98	-7.54	18.700000762939453
1	Trx_2_1_1	OCH	true	190.7	195.45	50.0	0.00	-12.44	23.799999237060547

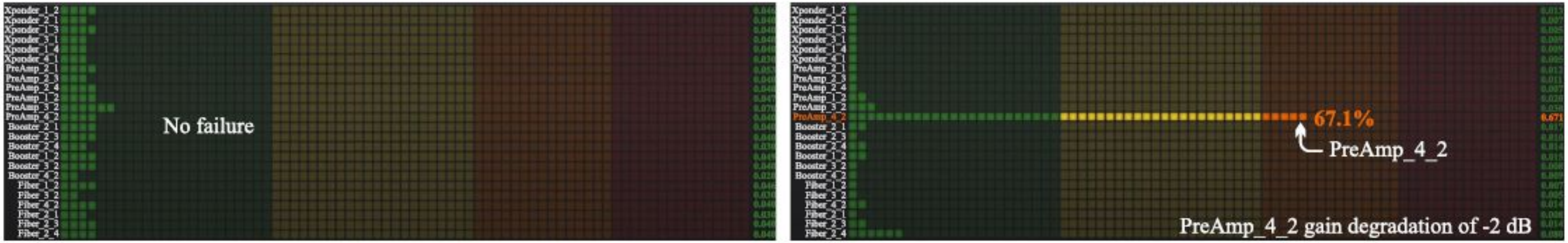
- Experimental setup

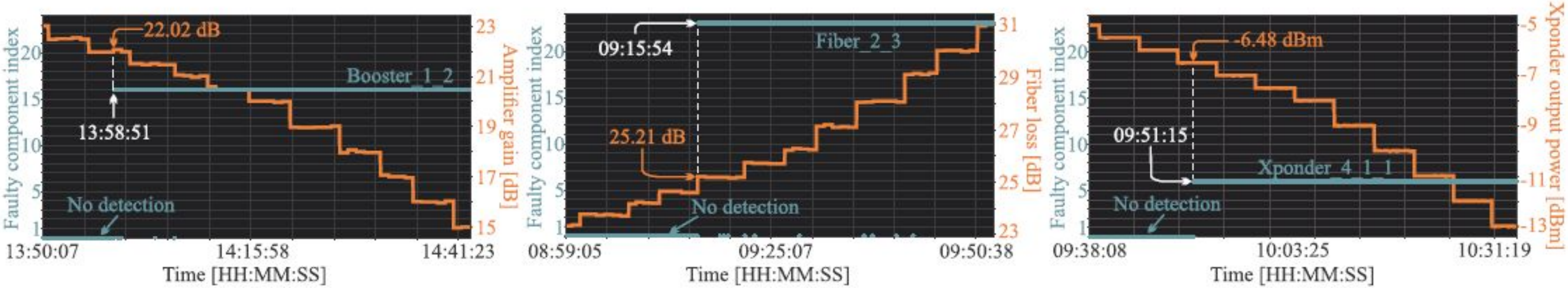


padtec

Case Study: ML-based Soft-Failure Localization

- Experimental soft-failure localization





Case Study: ML-based Soft-Failure Localization

- Soft-failure localization results

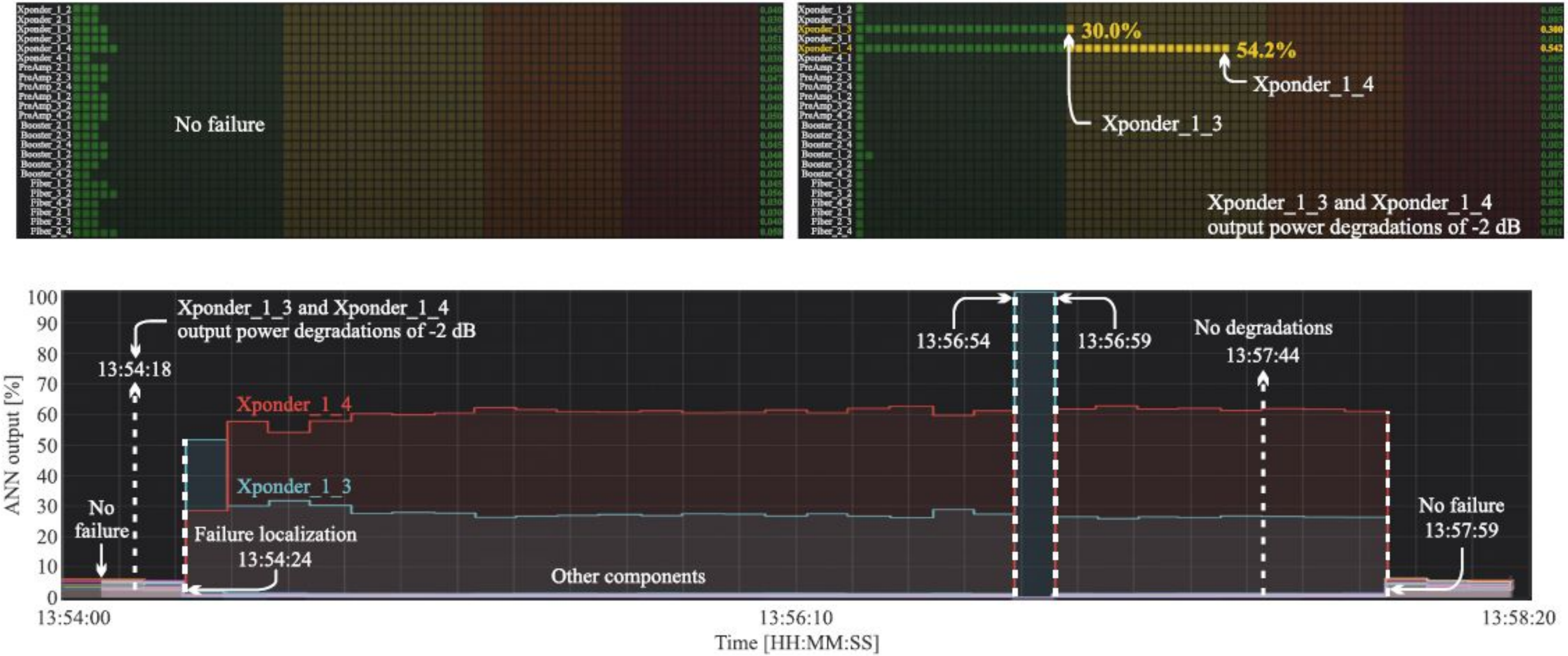
TABLE I
SINGLE-FAILURE LOCALIZATION RESULTS

Component	D_{FL} [dB]	T_{FL} [s]	Component	D_{FL} [dB]	T_{FL} [s]	Component	D_{FL} [dB]	T_{FL} [s]
Booster_1_2	0.98	185	PreAmp_2_3	1.92	80	Fiber_2_4	2.00	4
Booster_2_1	1.95	65	PreAmp_3_2	1.55	17	Fiber_4_2	1.42	5
Booster_2_3	2.13	65	PreAmp_2_4	1.46	107	Xponder_1_2	1.46	8
Booster_3_2	1.72	4	PreAmp_4_2	1.51	272	Xponder_2_1	1.48	5
Booster_2_4	2.54	4	Fiber_1_2	3.73	5	Xponder_1_3	1.97	4
Booster_4_2	1.49	15	Fiber_2_1	1.57	127	Xponder_3_1	1.51	54
PreAmp_1_2	1.49	4	Fiber_2_3	1.89	5	Xponder_1_4	1.48	6
PreAmp_2_1	2.90	5	Fiber_3_2	2.51	4	Xponder_4_1	1.50	122

D_{FL} and T_{FL} are the degradation and time to failure localization, respectively.

Case Study: ML-based Soft-Failure Localization

- Double failure localization



Conclusions

- DTs are gaining several fields of knowledge, and they should also become widespread in optical transmission systems
- NDTs should be of paramount importance for QoT estimation and soft-failure localization
- There are open challenges involving the NDT update behaviour and the consistency between the virtual and the physical network (related with QoT estimation!). Classic numerical optimization and ML-based techniques may be used
- We demonstrated an ML-based soft-failure estimation method based on synthetic failures generated in the virtual network


In-Network AR/CG Traffic Classification Entirely Deployed in the Programmable Data Plane: Unlocking RTP Features and L4S Integration

Alireza Shirmarz, Mateus N. Bragatto, Fábio Luciano Verdi, Suneet Kumar Singh, Christian Rothenberg, Gyanesh Patra, Gergely Pongrácz

Federal University of São Carlos (UFSCar), Sorocaba, SP, Brazil,
Norwegian University of Science and Technology (NTNU), Norway,
Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil,
Ericsson Research, Budapest, Hungary

From Pixels to Packets: Traffic Classification of Augmented Reality and Cloud Gaming

Alireza Shirmarz, Fábio Luciano Verdi
Department of Computer of Science
Federal University of São Carlos (UFSCar)
Sorocaba, Brazil
ashirmarz@ufscar.br, verdi@ufscar.br

Suneet Kumar Singh, Christian Esteve Rothenberg
School of Electrical and Computer Engineering
Universidade Estadual de Campinas (Unicamp)
Campinas, Brazil
ssingh@dca.fee.unicamp.br, chesteve@dca.fee.unicamp.br

SMARTNESS

ML for Networking Use Cases

Selected Publications

on XR & Cloud Gaming identification

Abstract

- **Objective**: Traffic Classification of Augmented Reality (AR) and Cloud Gaming (CG)

- **Importance & Necessity**:
    - AR widespread adoption in retail and industrial applications.
    - AR computational offloading to the remote server.
    - AR more sensitivity to delay
    - Heterogeneous processing units (e.g. CPU, GPU, DPU, FPGA)

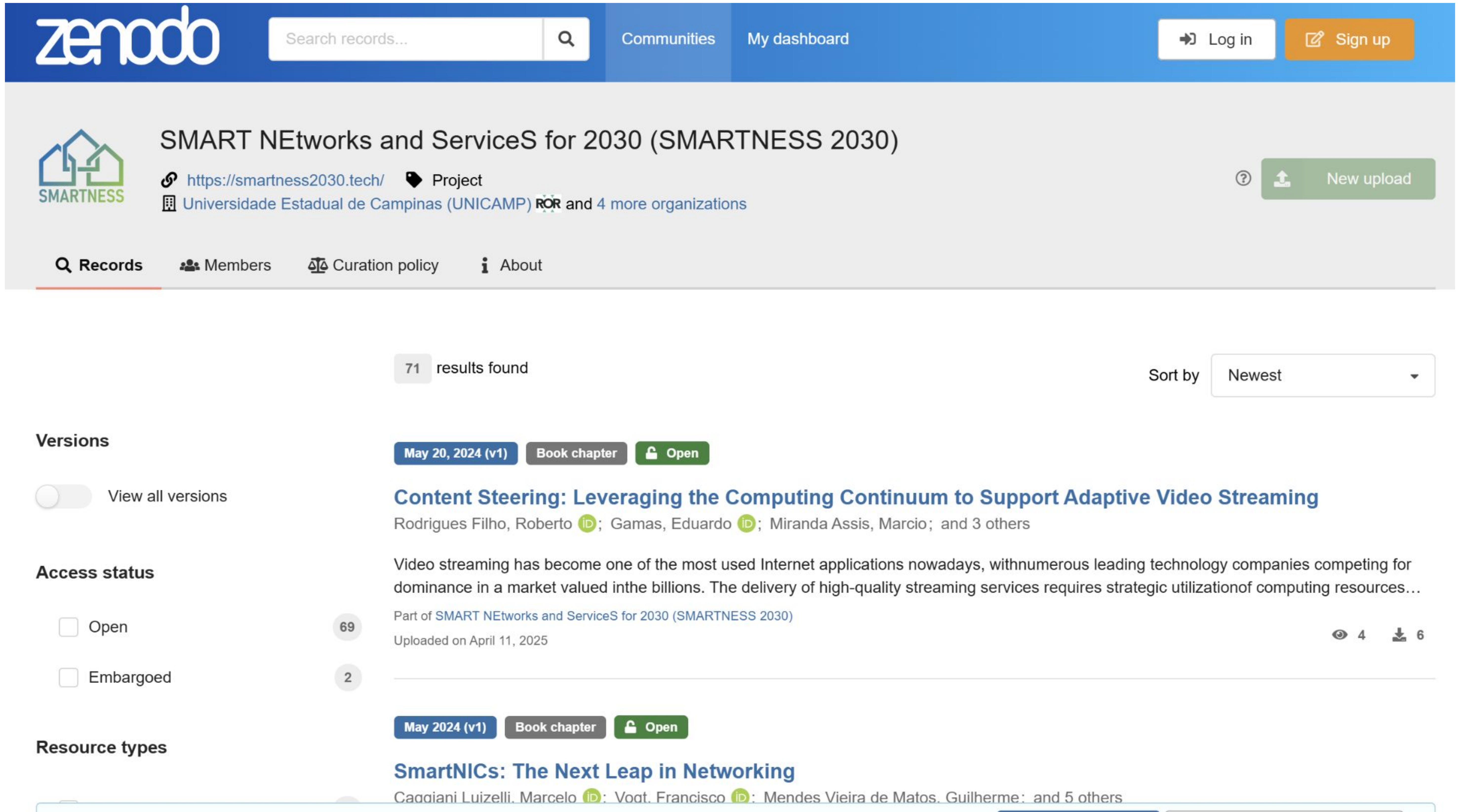
- **Approaches & Findings:**
    - Decision Tree (DT) & Random Forest (RF) models for classification.
    - The accuracy between 88.40% and 94.87% for network traffic classification into ‘AR’, ‘CG’ and ‘other’ classes.
        - ‘Other’ class includes four application types: Video Conference (VC), Video Streaming (VS), Live Video Streaming (LV), and Browsing.
    - Collected datasets for AR & CG
    - Evaluate & verification the models.

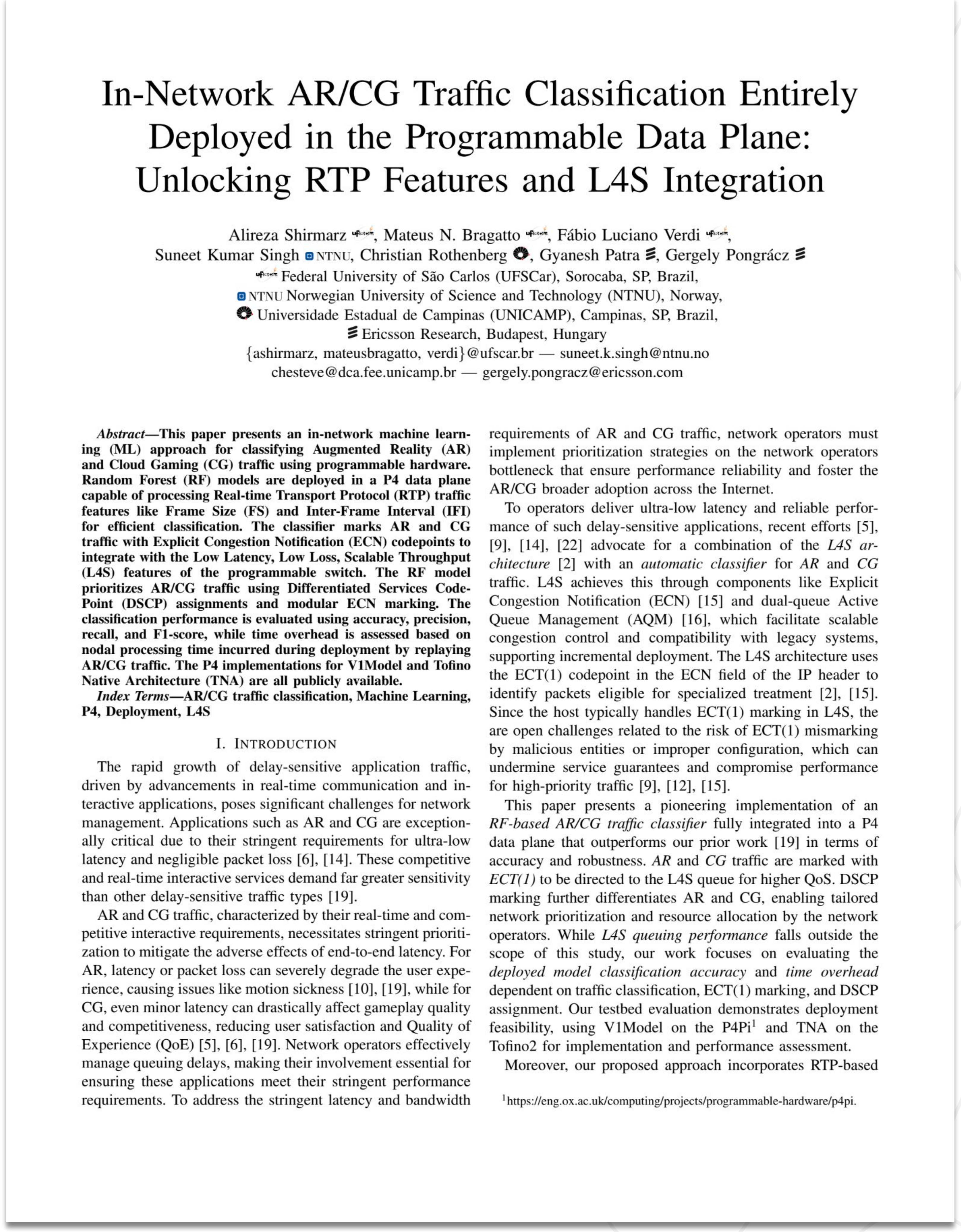
Further readings

Check out and follow SMARTNESS Zenodo

SMARTNESS

https://zenodo.org/communities/smartness2030/





In-Network AR/CG Traffic Classification Entirely Deployed in the Programmable Data Plane: Unlocking RTP Features and L4S Integration

Alireza Shirmarz, Mateus N. Bragatto, Fábio Luciano Verdi, Suneet Kumar Singh, Christian Rothenberg, Gyanesh Patra, Gergely Pongrácz

Federal University of São Carlos (UFSCar), Sorocaba, SP, Brazil,

NTNU Norwegian University of Science and Technology (NTNU), Norway,

Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil,

Ericsson Research, Budapest, Hungary

{ashirmarz, mateusbragatto, verdi}@ufscar.br — suneet.k.singh@ntnu.no

chesteve@dca.fee.unicamp.br — gergely.pongracz@ericsson.com

Abstract—**This paper presents an in-network machine learning (ML) approach for classifying Augmented Reality (AR) and Cloud Gaming (CG) traffic using programmable hardware. Random Forest (RF) models are deployed in a P4 data plane capable of processing Real-time Transport Protocol (RTP) traffic features like Frame Size (FS) and Inter-Frame Interval (IFI) for efficient classification. The classifier marks AR and CG traffic with Explicit Congestion Notification (ECN) codepoints to integrate with the Low Latency, Low Loss, Scalable Throughput (L4S) features of the programmable switch. The RF model prioritizes AR/CG traffic using Differentiated Services Code-Point (DSCP) assignments and modular ECN marking. The classification performance is evaluated using accuracy, precision, recall, and F1-score, while time overhead is assessed based on nodal processing time incurred during deployment by replaying AR/CG traffic. The P4 implementations for V1Model and Tofino Native Architecture (TNA) are all publicly available.**

Index Terms—**AR/CG traffic classification, Machine Learning, P4, Deployment, L4S**

I. INTRODUCTION

The rapid growth of delay-sensitive application traffic, driven by advancements in real-time communication and interactive applications, poses significant challenges for network management. Applications such as AR and CG are exceptionally critical due to their stringent requirements for ultra-low latency and negligible packet loss [6], [14]. These competitive and real-time interactive services demand far greater sensitivity than other delay-sensitive traffic types [19].

AR and CG traffic, characterized by their real-time and competitive interactive requirements, necessitates stringent prioritization to mitigate the adverse effects of end-to-end latency. For AR, latency or packet loss can severely degrade the user experience, causing issues like motion sickness [10], [19], while for CG, even minor latency can drastically affect gameplay quality and competitiveness, reducing user satisfaction and Quality of Experience (QoE) [5], [6], [19]. Network operators effectively manage queuing delays, making their involvement essential for ensuring these applications meet their stringent performance requirements. To address the stringent latency and bandwidth requirements of AR and CG traffic, network operators must implement prioritization strategies on the network operators bottleneck that ensure performance reliability and foster the AR/CG broader adoption across the Internet.

To operators deliver ultra-low latency and reliable performance of such delay-sensitive applications, recent efforts [5], [9], [14], [22] advocate for a combination of the *L4S architecture* [2] with an *automatic classifier* for *AR* and *CG* traffic. L4S achieves this through components like Explicit Congestion Notification (ECN) [15] and dual-queue Active Queue Management (AQM) [16], which facilitate scalable congestion control and compatibility with legacy systems, supporting incremental deployment. The L4S architecture uses the ECT(1) codepoint in the ECN field of the IP header to identify packets eligible for specialized treatment [2], [15]. Since the host typically handles ECT(1) marking in L4S, the are open challenges related to the risk of ECT(1) mismarking by malicious entities or improper configuration, which can undermine service guarantees and compromise performance for high-priority traffic [9], [12], [15].

This paper presents a pioneering implementation of an *RF-based AR/CG traffic classifier* fully integrated into a P4 data plane that outperforms our prior work [19] in terms of accuracy and robustness. *AR* and *CG* traffic are marked with *ECT(1)* to be directed to the L4S queue for higher QoS. DSCP marking further differentiates AR and CG, enabling tailored network prioritization and resource allocation by the network operators. While *L4S queuing performance* falls outside the scope of this study, our work focuses on evaluating the *deployed model classification accuracy* and *time overhead* dependent on traffic classification, ECT(1) marking, and DSCP assignment. Our testbed evaluation demonstrates deployment feasibility, using V1Model on the P4Pi[1] and TNA on the Tofino2 for implementation and performance assessment.

Moreover, our proposed approach incorporates RTP-based

[1] https://eng.ox.ac.uk/computing/projects/programmable-hardware/p4pi.

Conclusiones

Detection of <you name it> from data

Prediction of <you name it> from data

SMARTNESS

- **ML/AI** are **here to stay** and will **keep evolving** and **impacting** networking (operations / OPEX through automation, new revenue streams, etc.)
- **“Data is the new oil”**
  - *Like oil, data is valuable, but if unrefined it cannot really be used*
- Recommendations:
  - Start **collecting data** (the 4 Vs challenge! Veracity, Velocity, Volume, Variety)
  - **Identify** and **Rank** your main “**headaches**” (e.g. cost, risk, dissatisfaction, etc.)
  - **Partner** with ML/AI savvy (real-data hungry) **Research Groups** (e.g. Universities)
  - **Innovate** and take some risks to play out the **use case for ML/AI in your network**

Gracias!

Preguntas?

chesteve@unicamp.br

https://smartness2030.tech/

SMARTNESS

ACKNOWLEDGMENTS & DISCLAIMER

SMARTNESS

This work has been performed within the framework of the FAPESP Engineering Research Center (ERC) Program under FAPESP grant agreement #2021/00199-8 (SMARTNESS).

FAPESP

The information in this document reflects the SMARTNESS ERC's view, but the partner institutions of SMARTNESS are not liable for any use that may be made of any of the information contained therein. The opinions, hypotheses and conclusions or recommendations expressed in this material are the responsibility of the author(s) and do not necessarily reflect FAPESP's vision or the other granting authorities. The views expressed are solely those of the authors and do not necessarily represent Ericsson's official standpoint.

UNICAMP

ERICSSON

INTRIG

INFORMATION & NETWORKING TECHNOLOGIES RESEARCH & INNOVATION GROUP

References

3rd Party Sources / Recommended Reading

- R. Boutaba et al. “**A comprehensive survey on machine learning for networking: evolution, applications and research opportunities**”. J Internet Serv Appl 9, 16, 2018.
- M. A. Ridwan, N. A. M. Radzi, F. Abdullah and Y. E. Jalil, "**Applications of Machine Learning in Networking: A Survey of Current Issues and Future Challenges**," in IEEE Access 2021
- Marios Evangelos Kanakis, Ramin Khalili, and Lin Wang. **Machine Learning for Computer Systems and Networking: A Survey**. ACM Comput. Surv. 55, 4, Article 71, April 2023
- O. Onireti, S. Oteafy, J. Thompson. "**A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks**" IEEE Communications Surveys & Tutorials, 2019
- S. Hussain, R. Saeed, F. Abbas. "**Machine Learning in 5G/B5G Mobile Networks: A Comprehensive Survey**". IEEE Access, 2020
- M. Alsabah, S. Alrabeiah, M. Alwakeel. **Artificial Intelligence in 5G and Beyond: A Survey and Perspective on Enabling Technologies, Applications, and Open Research Challenges**. 2021
- P. Casas “**Two Decades of AI4NETS – Challenges & Research Directions**” in 5th IEEE/IFIP International Workshop on Analytics for Network and Service Management (AnNet), 2020.
- Michael Schapira. “**AI for networking, and networking for AI**”. The Networking Channel, 2023. Available at https://www.youtube.com/watch?v=i6DvbfIUPSg.
- Andrew Ng. “**Machine Learning**”. Stanford Online. Available at https://youtu.be/vStJoetOxJg?si=irOCJjd5n-872uKl

References

Author Publications / Further Reading

- Stadler, R., Pasquini, R., & Fodor, V. (2017). Learning from network device statistics. *Journal of Network and Systems Management*, *25*, 672-698.
- de Araújo, P. L., Carneiro, M. G., Contreras, L. M., & Pasquini, R. (2024). MTP-NT: A Mobile Traffic Predictor Enhanced by Neighboring and Transportation Data. *IEEE Transactions on Network and Service Management*.
- de Almeida, L. C., da Silva, W. R. D., Tavares, T. C., Pasquini, R., Papagianni, C., & Verdi, F. L. (2024). DESiRED—Dynamic, Enhanced, and Smart iRED: A P4-AQM with Deep Reinforcement Learning and In-band Network Telemetry. *Computer Networks*, *244*, 110326.
- Donatti, A., Correa, S. L., Martins, J. S., Abelem, A. J., Both, C. B., de Oliveira Silva, F., ... & Carvalho, T. C. (2023). Survey on Machine Learning-Enabled Network Slicing: Covering the Entire Life Cycle. *IEEE Transactions on Network and Service Management*, *21*(1), 994-1011.
- de Almeida, L. C., da Silva, J. L., Lins, R. P., Maciel Jr, P. D., Pasquini, R., & Verdi, F. L. (2023, May). Wave-um gerador de cargas múltiplas para experimentação em redes de computadores. In *Anais Estendidos do XLI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos* (pp. 9-16). SBC.

References

Author Publications / Further Reading

- “A Framework for QoS and QoE Assessment of Encrypted Video Traffic with 4G and 5G Open Datasets”.
- “Soft Failure Localization Using Machine Learning with SDN-based Network-wide Telemetry”
- "Demonstration of ML-assisted Soft-Failure Localization Based on Network Digital Twins"
- "DNN-based QoT Estimation Using Topological Inputs and Training with Synthetic-Physical Data"
- "Chat-IBN-RASA: Building an Intent Translator for Packet-Optical Networks based on RASA".
- "When Digital Twins Meet Optical Networks Operations".
- “A supervised Machine Learning approach for DASH video QoE prediction in 5G networks"
- "Machine learning assisted real-time DASH video QoE estimation technique for encrypted traffic".
- “Machine Learning Approach to Estimate Video QoE of Encrypted DASH Traffic in 5G Networks”
- "EFFECTOR: DASH QoE and QoS Evaluation Framework For EnCrypTed videO tRaffic"
- “Predicting XR Services QoE with ML: Insights from In-band Encrypted QoS Features in 360-VR”
- "QoEyes: Towards Virtual Reality Streaming QoE Estimation Entirely in the Data Plane"
- "Machine Learning-Assisted Closed-Control Loops for Beyond 5G Multi-Domain ZeroTouch Networks".
- "Harnessing UAVs for Fair 5G Bandwidth Allocation in Vehicular Communication via Deep Reinforcement Learning".
- "Machine Learning for NextGeneration Intelligent Transportation Systems: A Survey"
- “The role of Machine Learning in Fluid Network Control and Data Planes”
- “MTP-NT: A Mobile Traffic Predictor Enhanced by Neighboring Transportation Data”
- “Transfer of Deep Reinforcement Learning for Cloud Service’s Elasticity”

Datasets

- Rafael Pasquini. GitHub. https://github.com/rafaelpasquini
- Patrick L. Araújo. GitHub. https://github.com/PatrickLdA/milan-telecom-analysis
- Ian Resende. GitHub. https://github.com/iansmps/dissertation_repo
- Leandro Almeida. GitHub. https://github.com/leandrocalmeida
- SMARTNESS ERC. GitHub. https://github.com/smartness2030/hackathon5G/tree/main/datasets
- Raza Ul Mustafa, Christian Esteve Rothenberg, Chadi Barakat, “YouTube goes 5G: Benchmarking YouTube in 4G vs 5G”, IEEE Dataport, doi: https://dx.doi.org/10.21227/h00h-ew92. Dec., 2022.

Credits

Pedro Casas (slide courtesy + talk recordings) [Alumni UdelaR]: https://bigdama.ait.ac.at/pcasas/

- **DEEP in the NET – Deep Learning for Network Monitoring and Analysis** PhD Course, 10th TMA PhD School, the Netherlands 2022.
  - **1 – Introduction to Deep Learning for TMA**
  - **2 – Deep Learning Basics**
  - **3 – Introduction to Graph Neural Networks**
- **AI4NETS – AI/ML for** Tutorial, 38th IFIP Performanc
  - **1 –**
  - **2**
  - **3**
  - **4**
  - **5 – Other Learning Topics**
- https://www.youtube.com/watch?v=B5PmsGYOZ04
- https://www.youtube.com/watch?v=rjkEl9c5jtE
- https://www.youtube.com/watch?v=rjkEl9c5jtE

https://bigdama.ait.ac.at/ *http://mobiqoe.ait.ac.at/*

pedro.casas@ait.ac.at
http://pcasas.info

Credits

Raza Ul Mustafa, PhD @ INTRIG/UNICAMP: "Machine Learning Assisted DASH Video QoE Inference Through Network QoS Features in 4G and 5G scenarios", Dec. 2022.

Now, Assistant Professor at Loyola University New Orleans

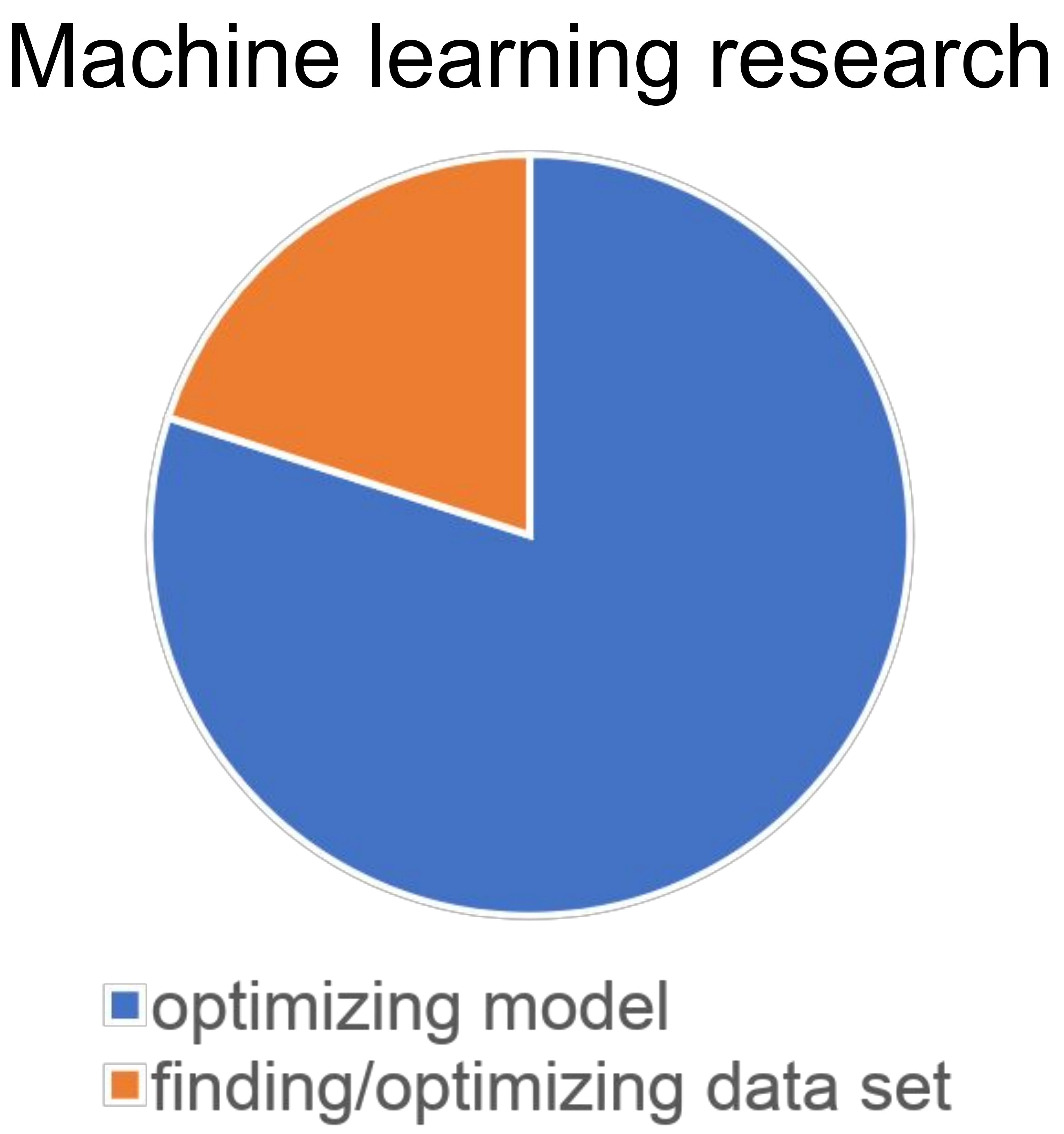
https://www.loyno.edu/academics/faculty-and-staff-directory/dr-raza-ul-mustafa

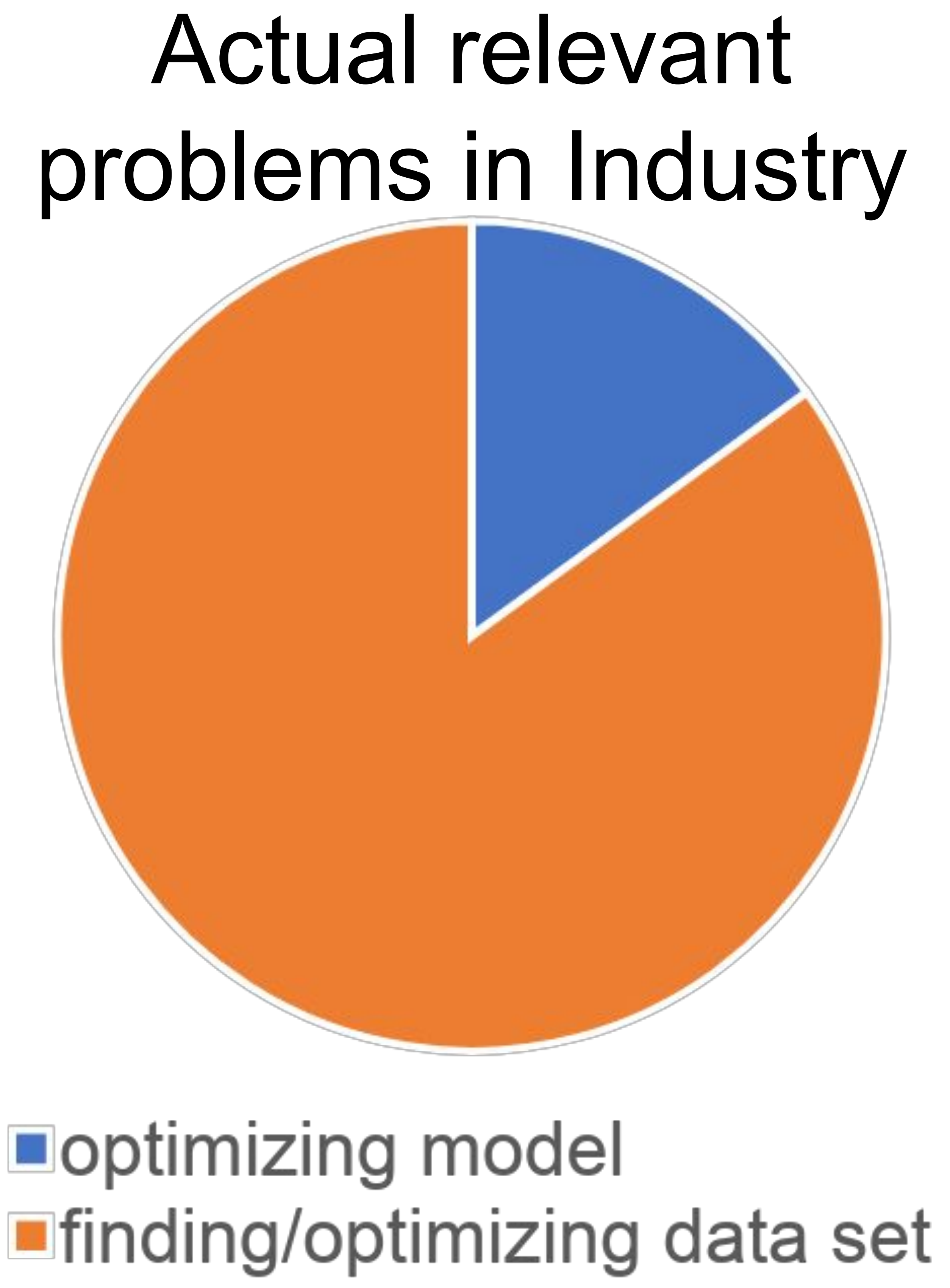
-

SMARTNESS

BACKUP

It’s more about the data set

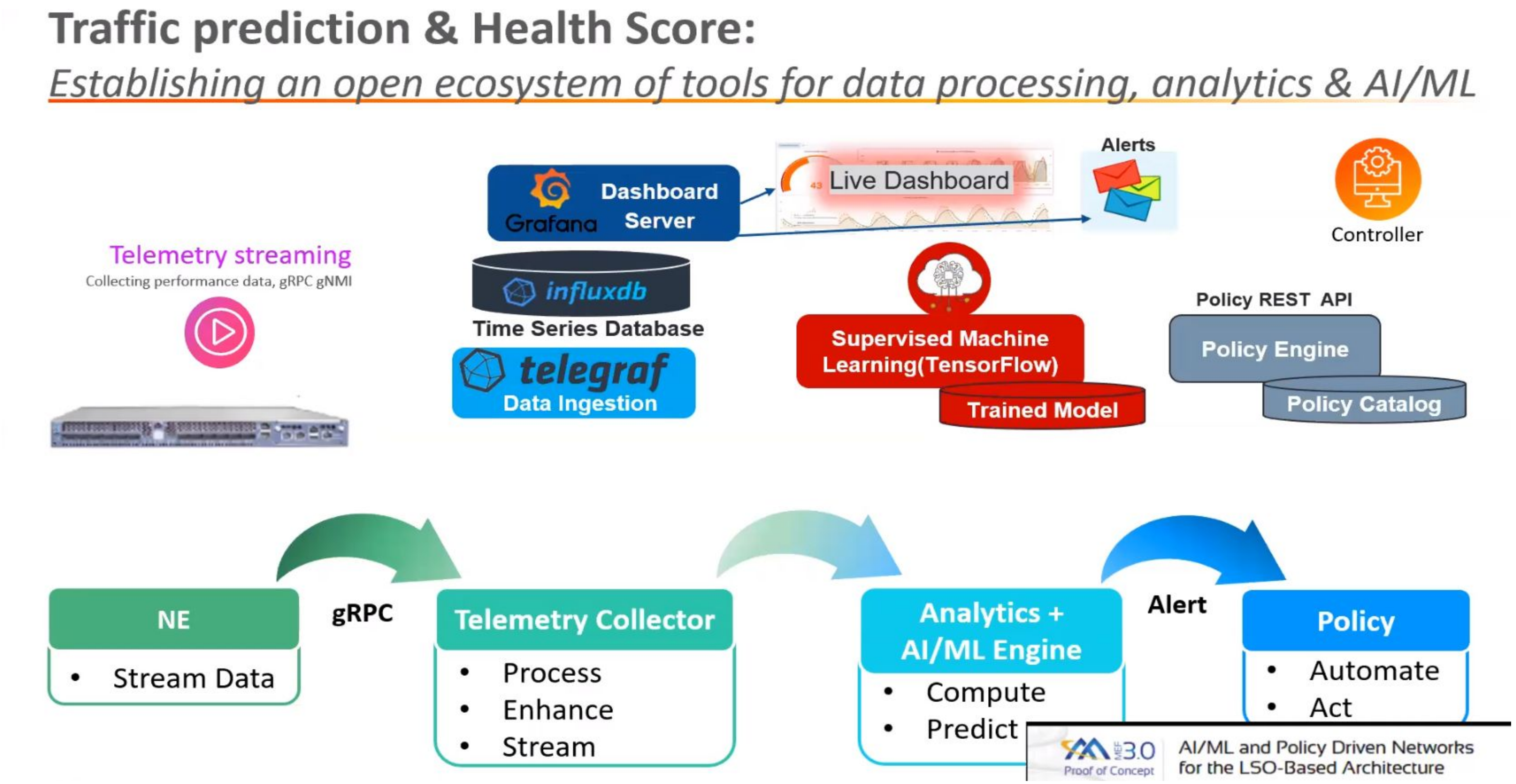




- Most engineering applications use simple ML on domain-specific data sets
- Standardized open dataset for optical networks?
- Customer privacy/ company confidentiality / rarity of failure scenarios

Optical Performance Monitoring (OPM)

H. Bock, Infinera, OSA PC, 2020



Century Link, Infinera, Telia Company

- SDN-based telemetry can provide a “health check” on connections and components

Quality of Transmission (QoT) Estimation

HUAWEI | EUROPEAN RESEARCH CENTER

Why is QoT estimation still inaccurate?

Y. Pointurier, Huawei, ICTON, 2020

- **Inaccurate modeling:**
  - Physics is not well mastered (example: spectral hole burning)
  - Physics is well mastered but modeling is too slow (example: split step Fourier, coherent Gaussian noise model)
- **Inaccurate parameters:**
  - Components characterized in the lab, but behavior varies from sample to sample (example: EDFA gain vs. load)
  - Full lines (OMS) characterized in the lab, but impossible to characterize all lines of a network
  - Behavior of components/lines change when moved from lab to the field (example: TRX temperature)
  - Type of components known but not characterized (example: fiber)
  - Type of component not known (example: fiber!)
  - Events not recorded or monitored (example: splice)

Quality of Transmission (QoT) Estimation

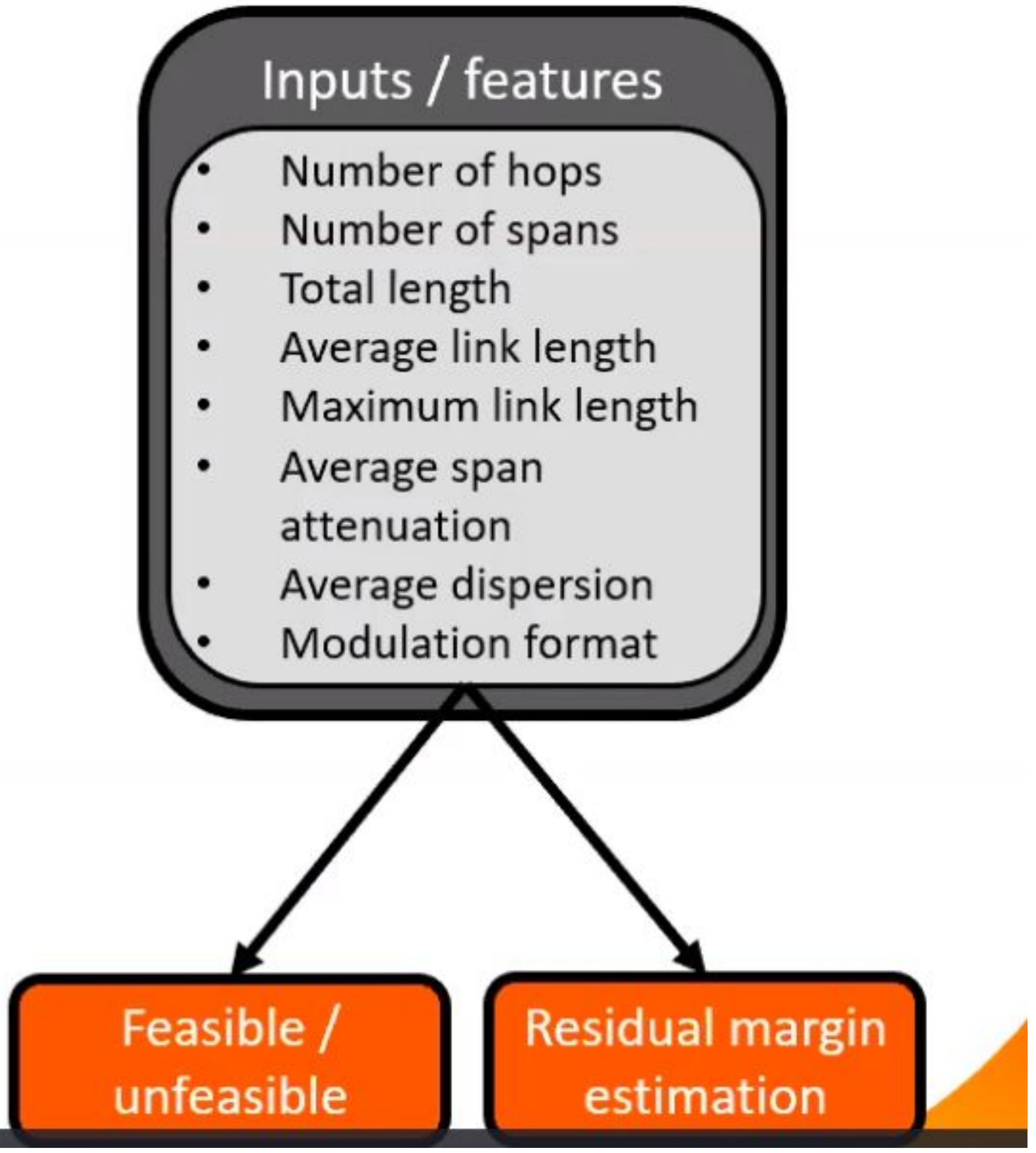
QoT Estimation: Labeled Data Gathering

H. Bock, Infinera, OSA PC, 2020

- Due to the lack of real network data, synthetic data generated for 4 reference networks
- The **networks have different characteristics** in terms of number of nodes, links, fiber types, and fiber span length
- The **Gaussian noise (GN) model** was used to calculate lightpath QoT for line interfaces operating at: **QPSK (100Gbit/s), 8QAM (150Gbit/s), 16QAM (200Gbit/s), 32QAM (250Gbit/s), 64QAM (300Gbit/s)**
- End-of-life performance was considered, i.e. considering fully loaded links

Networks characteristics

Network	Nodes	Links	Av. Nodal Degree	Av. Link Length [km]	Fiber Types	Number of Paths
GBN	12	20	3.3	243.0	SSMF	4,099
TIM	44	71	3.2	174.3	SSMF LEAF	1,134,272
SPARKLE	49	72	2.9	388.0	SSMF LEAF	241,115
CORONET	75	82	2.2	396.0	SSMF	105,320



UNICAMP

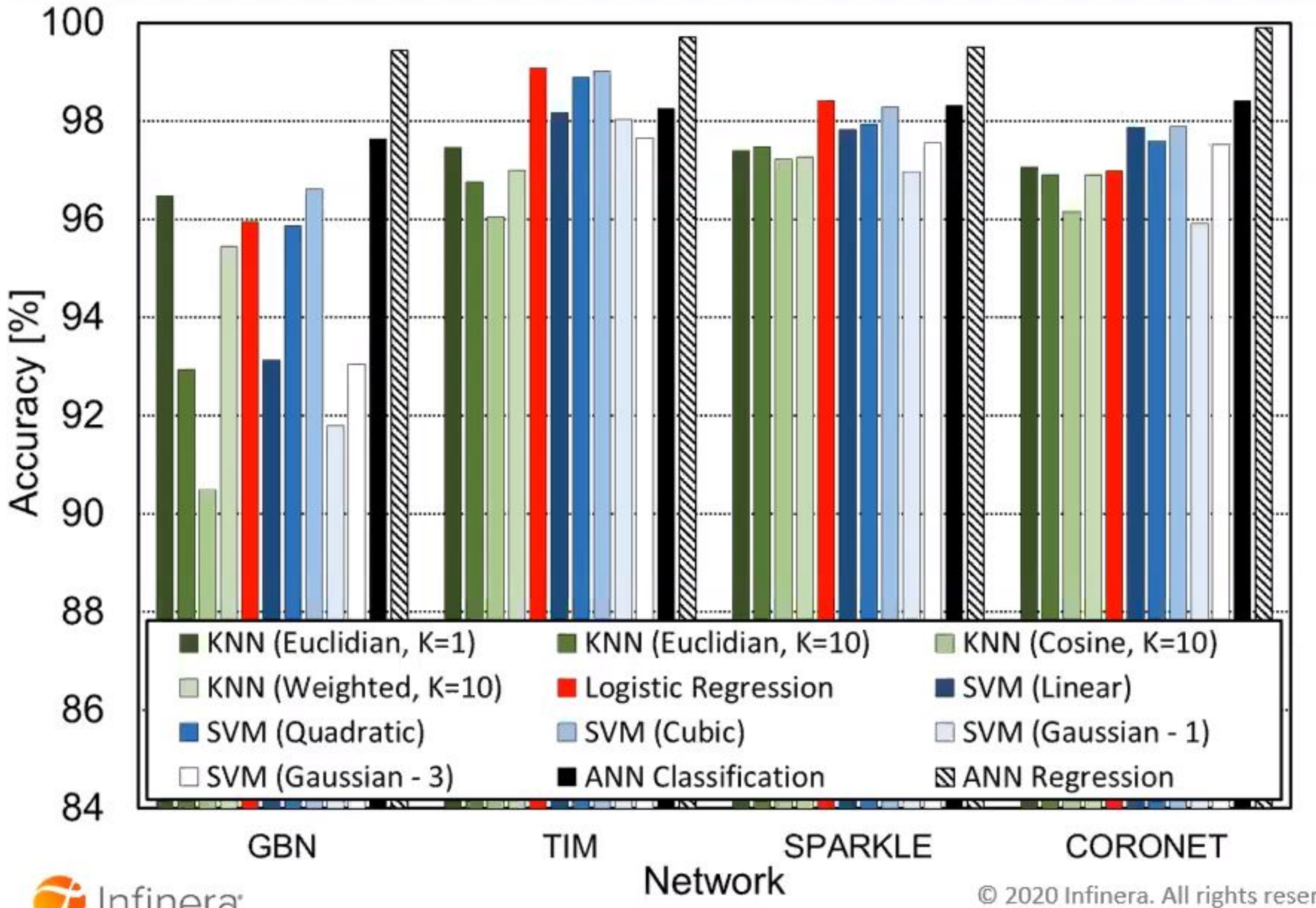
Quality of Transmission (QoT) Estimation

H. Bock, Infinera, OSA PC, 2020

Results: Accuracy Considering the Minimum Training Set

- **ANN for regression** is the model with the highest accuracy for all networks
- ANN for classification and logistic regression present second-best results
- For **simpler networks ANN for classification** obtains the second-best result, whereas for networks with a more **complex fiber mix logistic regression** performs better

Accuracy [%]



Number of training and testing examples in the data sets, and best obtained accuracy for each considered network

Network	Number of paths in training set	Number of paths in testing set	Best accuracy obtained
GBN	189	20 495	99.44%
TIM	5 318	5 671 360	99.71%
SPARKLE	1 161	1 205 575	99.49%
CORONET	530	526 600	99.90%

Infinera

[10] R. Morais, *et al.*, Machine learning models for estimating quality of transmission in DWDM networks , in *IEEE/OSA JOCN*, 2018.

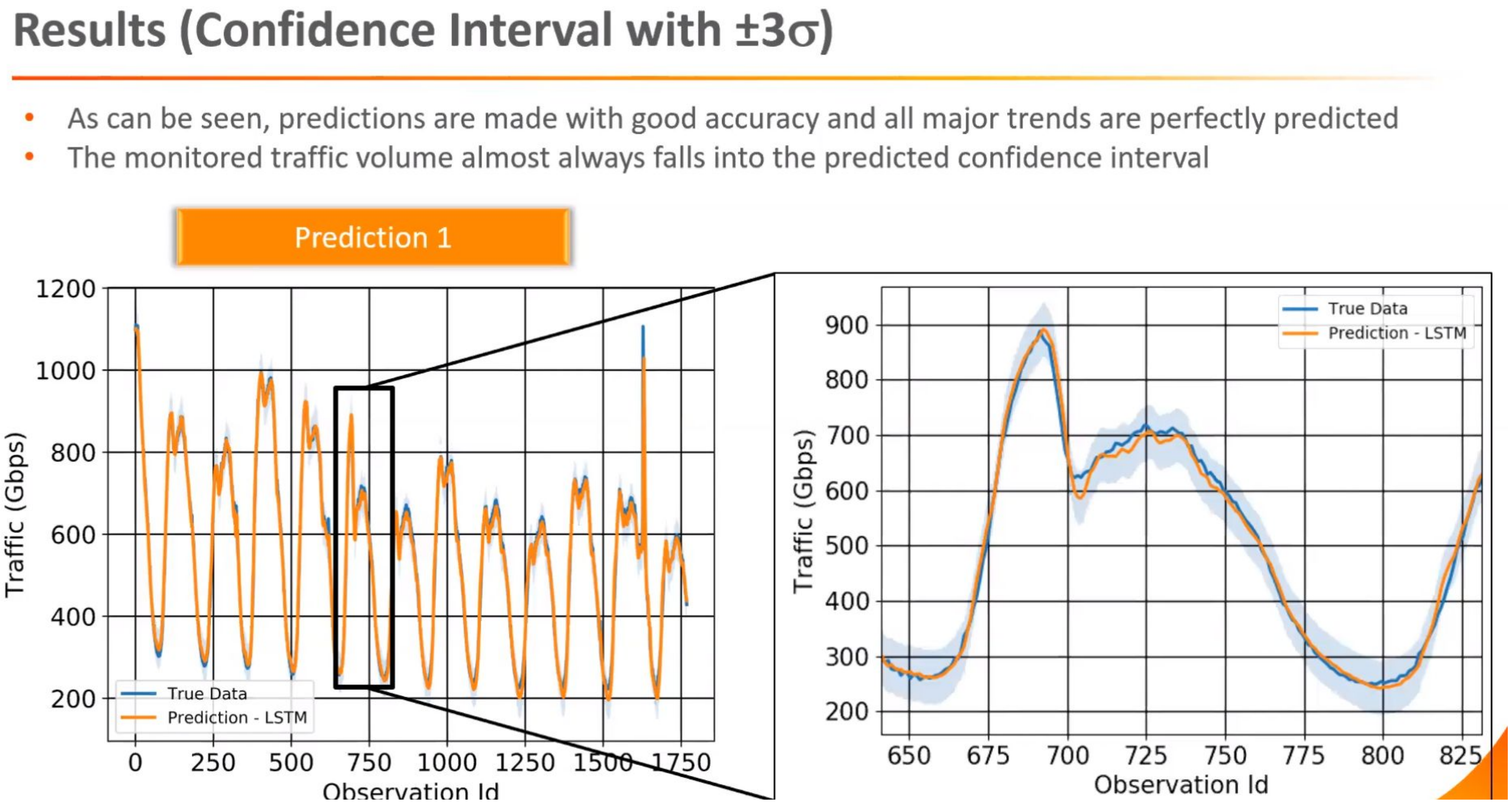
15

UNICAMP

Traffic Prediction

- Traffic predictions for load balancing

H. Bock, Infinera, OSA PC, 2020

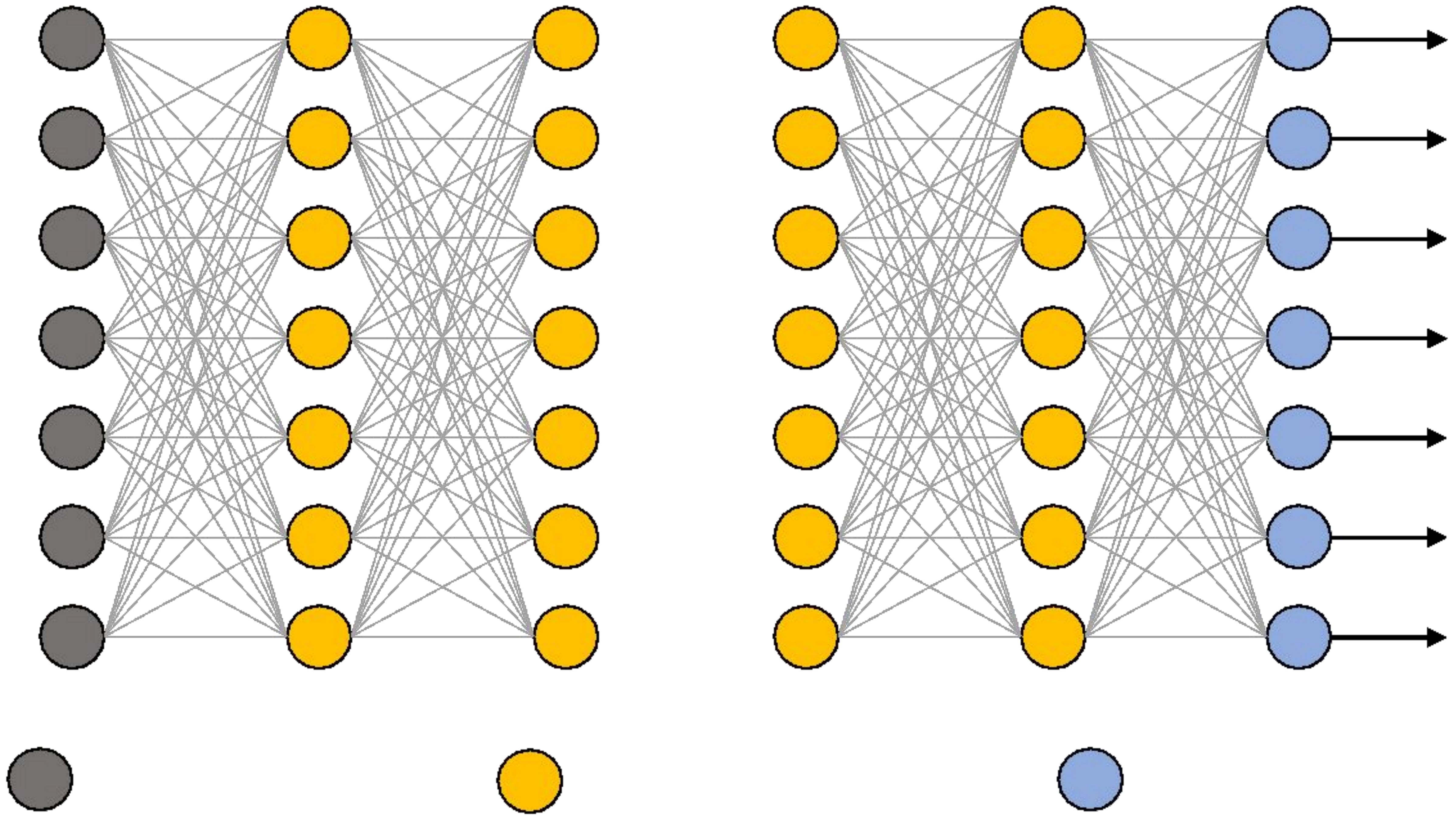


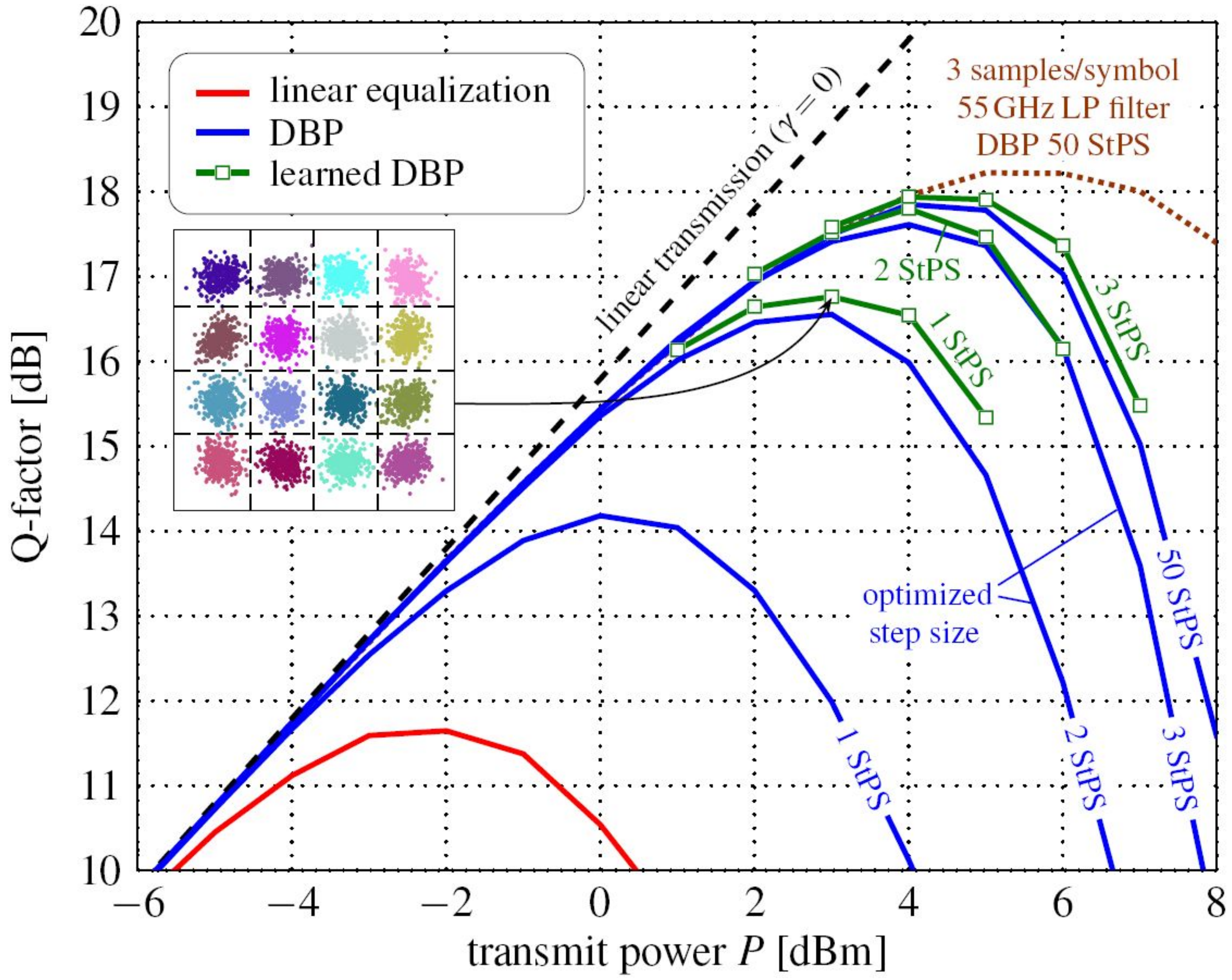
Nonlinearity Compensation

Fiber nonlinearity compensation using deep learning

- Digital BackPropagation(DBP)'s sequence of interleaved linear and nonlinear operations

$\mathbf{W}^{-1}\,\sigma^{-1}(\mathbf{W}^{-1}\,\sigma^{-1}(\dots.))$ can be modeled as a DNN structure





- Parameters of **W** and nonlinear function σ can be learned via BP algorithm
- Learned DBP performs similarly to standard DBP but is computationally simpler

C. Häger and H. D. Pfister, "Nonlinear Interference Mitigation via Deep Neural Networks," 2018 Optical Fiber Communications Conference and Exposition (OFC), San Diego, CA, 2018, pp. 1-3.

Failure Management

JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 37, NO. 16, AUGUST 15, 2019 4125

A Tutorial on Machine Learning for Failure Management in Optical Networks

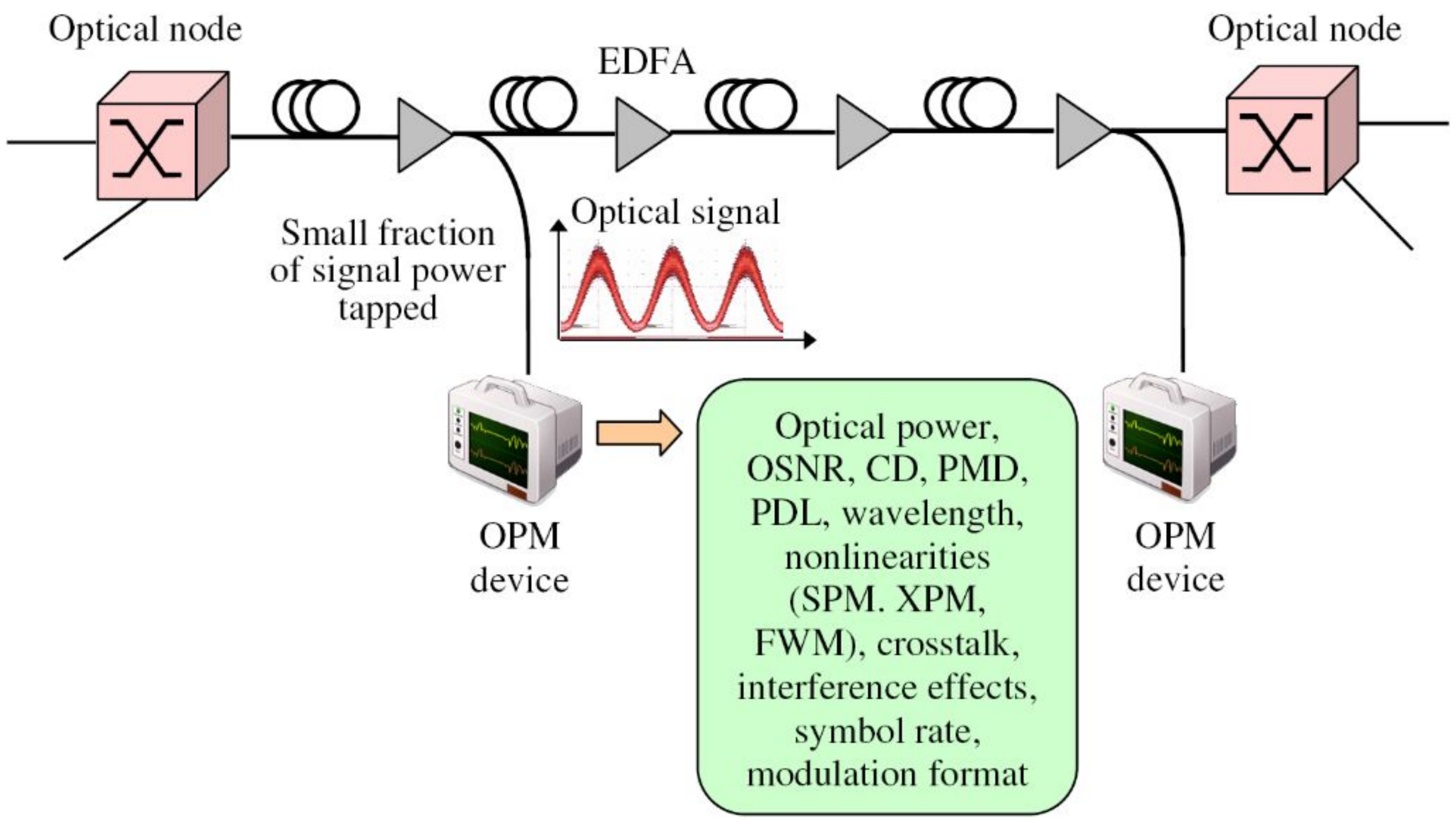
Francesco Musumeci, Cristina Rottondi, Giorgio Corani, Shahin Shahkarami, Filippo Cugini, and Massimo Tornatore

(Invited Tutorial)

Algorithm	Task	Description	Ref.
Random Forests	Detection	BER anomaly detection	[33]
	Identification	equiment failure type identification	[34]
Artificial Neural Networks	Monitoring	OSNR monitoring	[35]–[37]
	Monitoring	eye diagram monitoring	[38]–[43]
	Monitoring	phase portrait monitoring	[44]
	Prediction/Identification	equipment failure prediction	[45], [46]
	Detection/Identification	BER anomaly detection and identification	[33]
Support Vector Machines	Prediction	equipment failure prediction	[47]
	Detection	BER anomaly detection	[33]
	Localization/Identification	filter failure identification and localization	[34]
Gaussian Processes	Monitoring	OSNR monitoring	[48]
	Localization/Identification	link failure identification and localization	[49]
Bayesian Networks	Localization/Identification	localization and identification of tight filter-ing anc inter-channel interference	[50]
	Identification	failure diagnosis	[51]–[54]
Network Kriging	Localization	link failure localization	[55]

UNICAMP

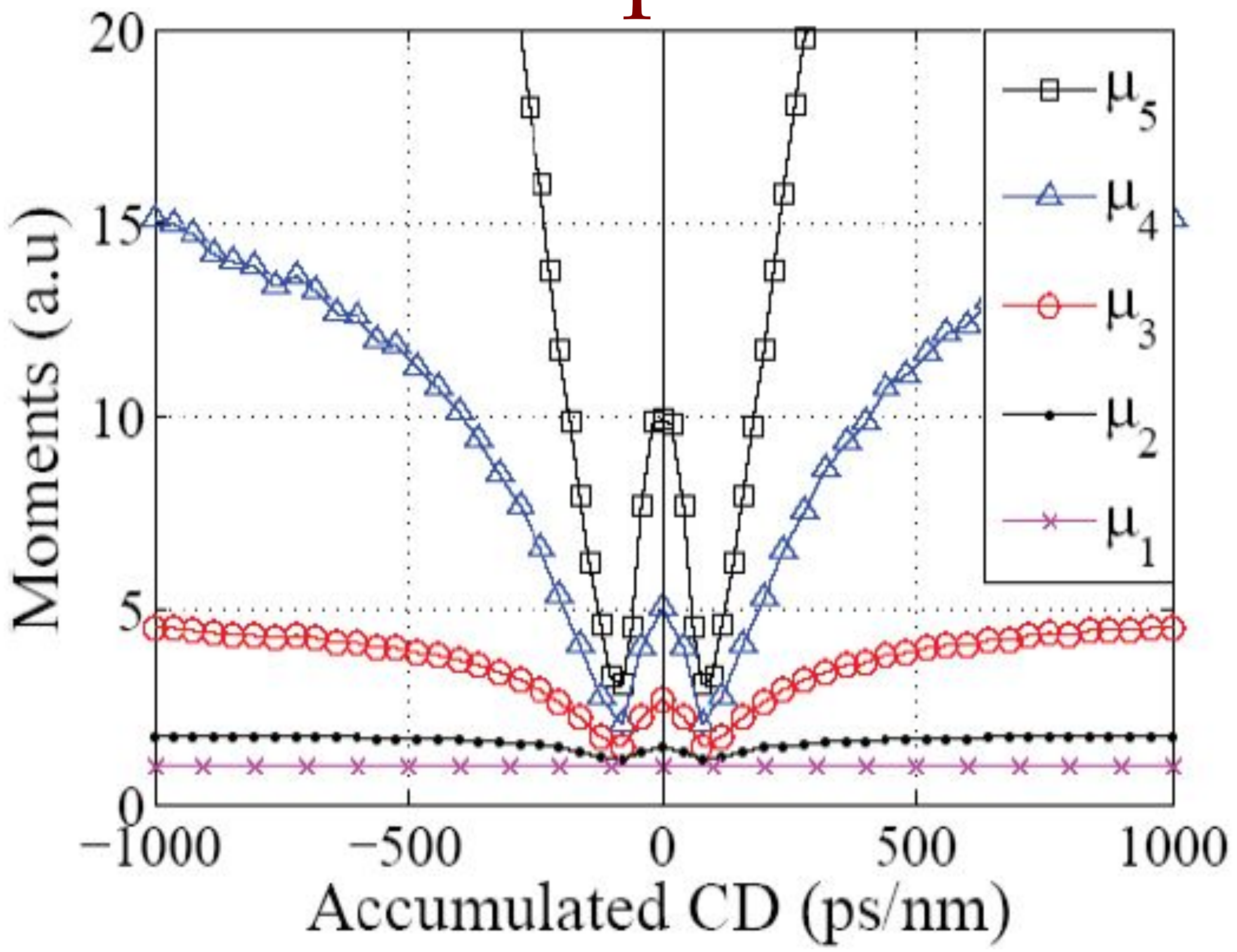
Optical Performance Monitoring (OPM)



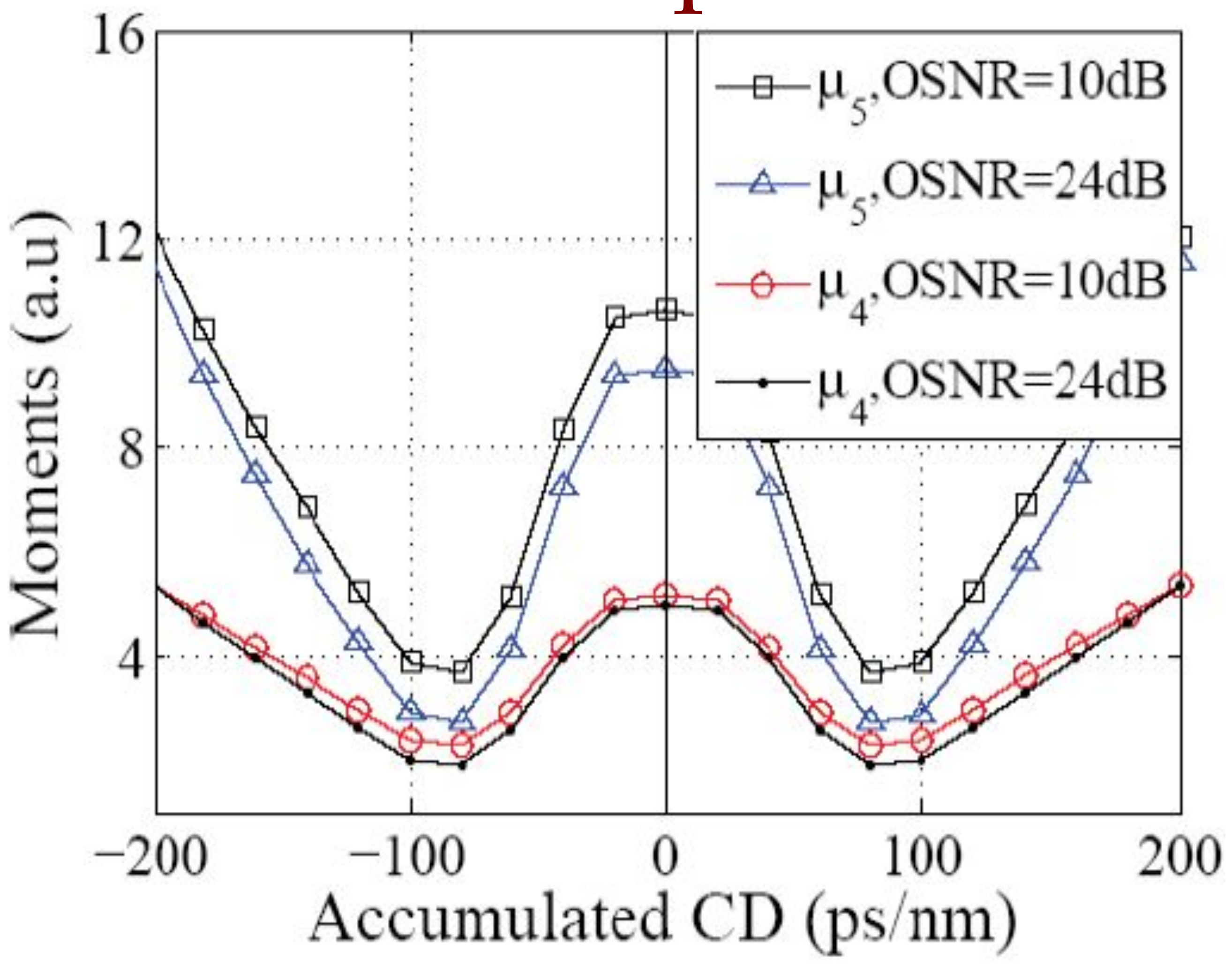
- OPM is a set of measurements performed (@ Tx/Rx or intermediate nodes) on an optical signal to estimate physical parameters e.g. OSNR, CD etc. of an optical channel
- Reliable/efficient operation of optical networks requires real-time physical links information
- OPM is also becoming a key component to enable impairment-aware SDN

Optical Performance Monitoring (OPM)

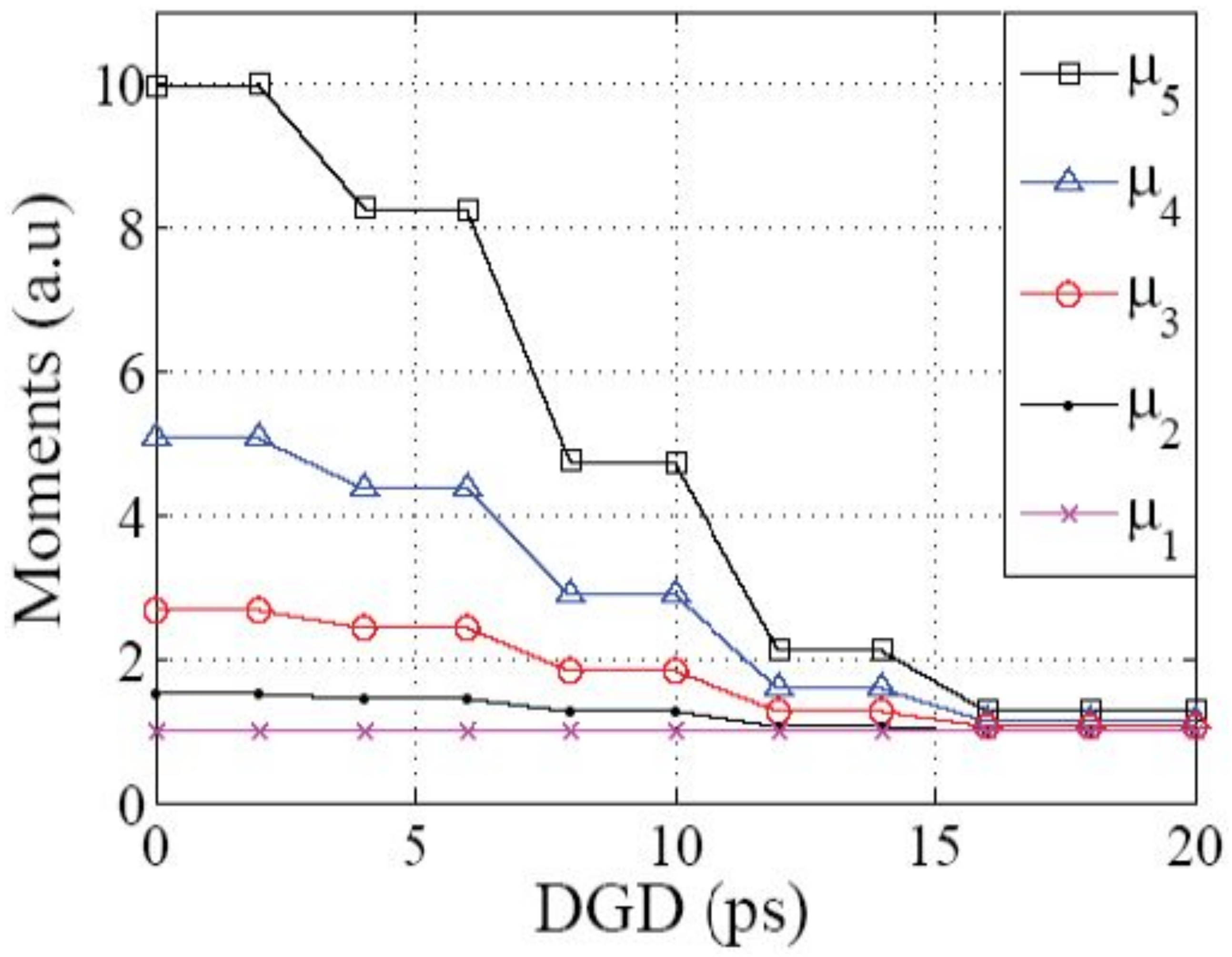
CD dependence



OSNR dependence

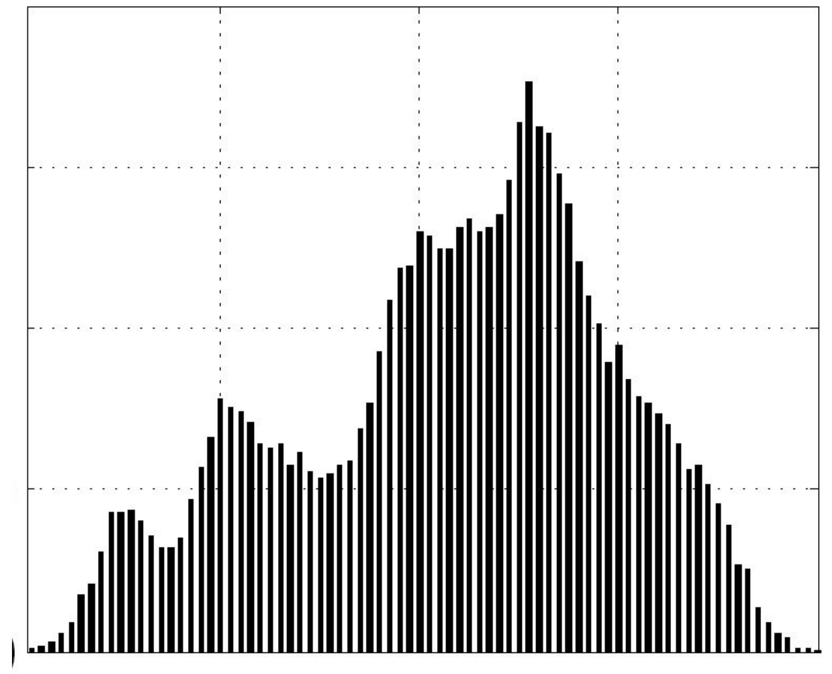


DGD dependence



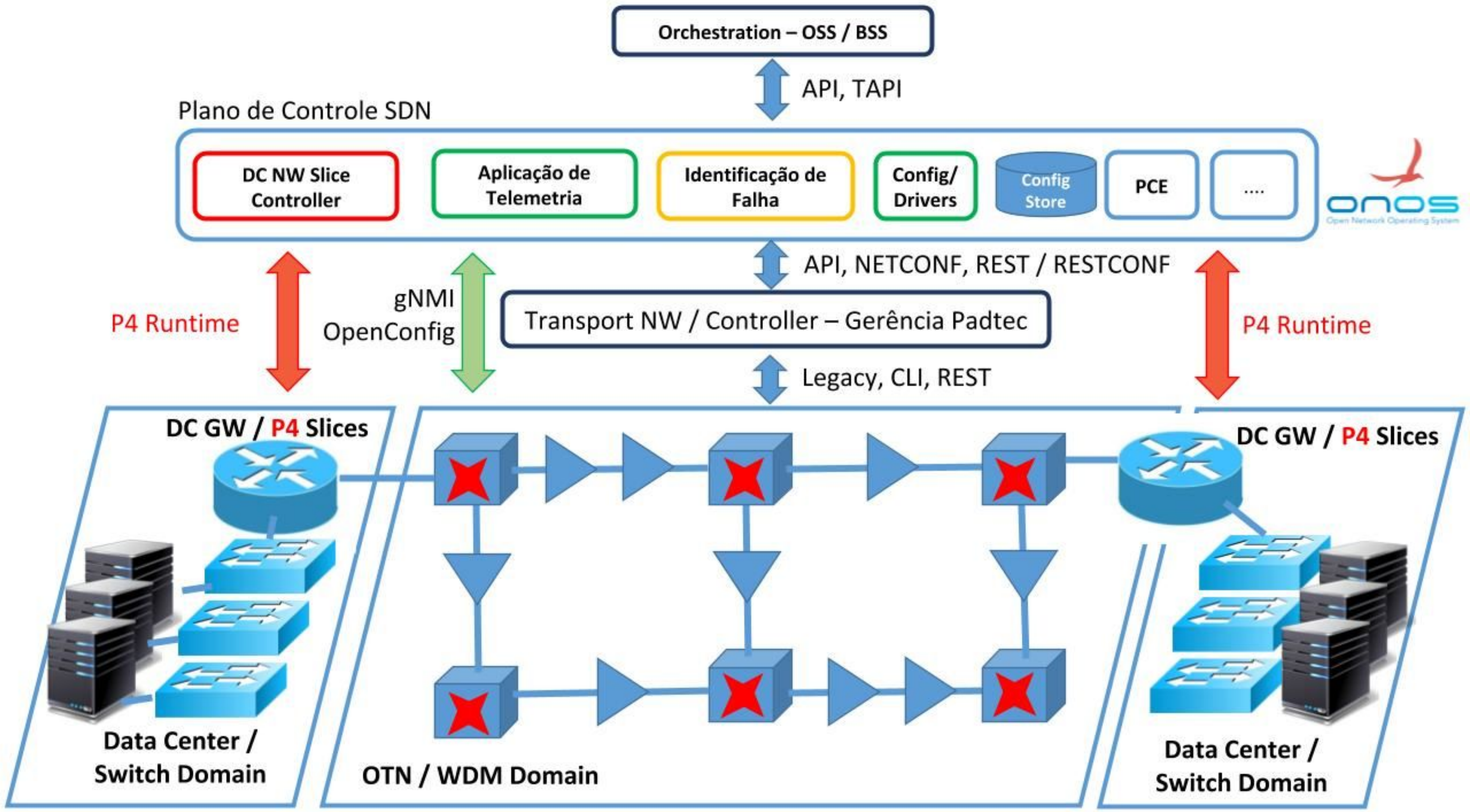
- Moments of μ_1, μ_2, μ_3 o of the he amplitude histogram can provide information about physical effects affecting transmission

Amplitude distributions

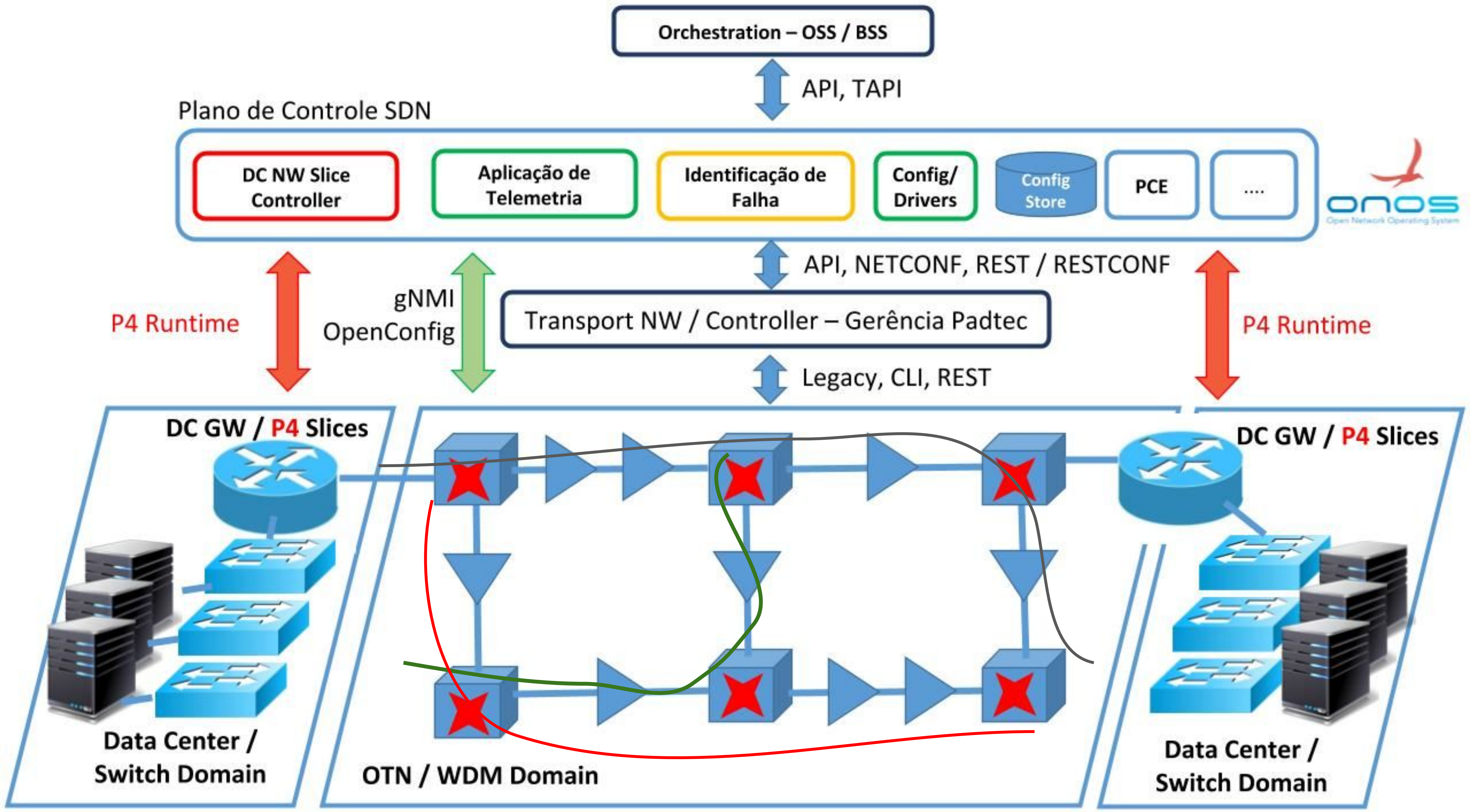


F.N. Khan *et al.*, *IEEE Photonics Technology Letters*, Jun. 2012.

Failure Management



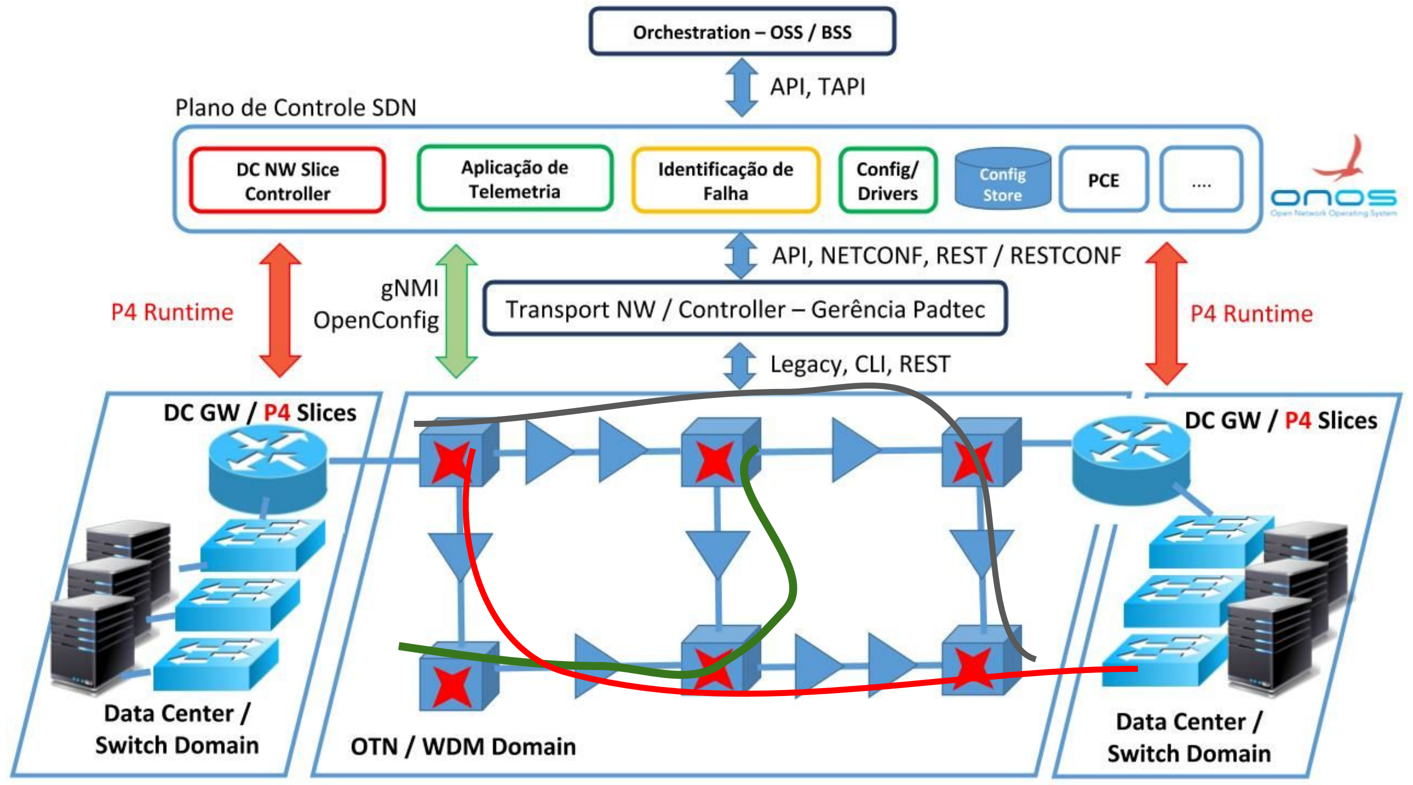
Failure Management

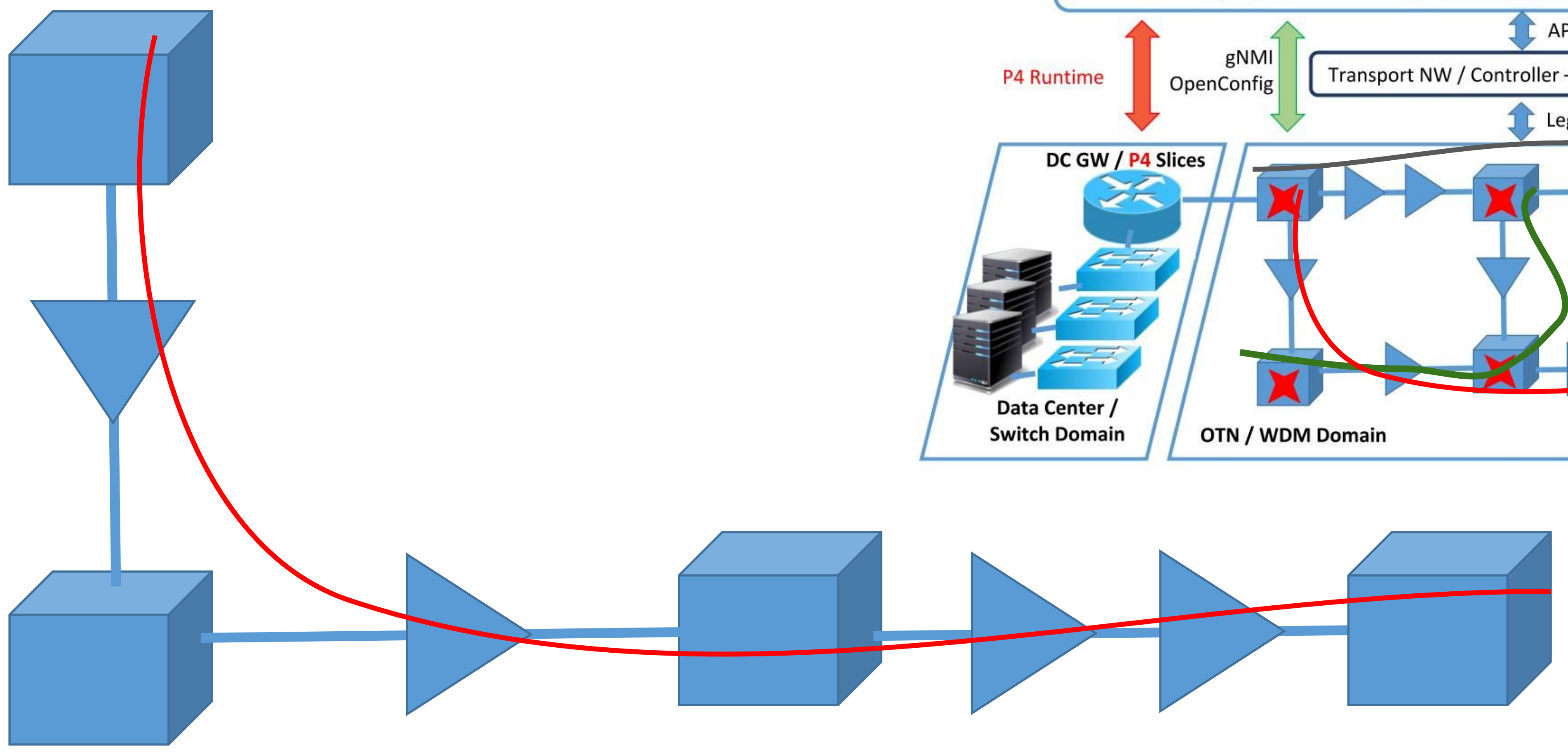


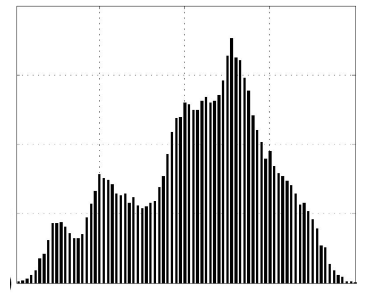
Failure Management

Per lightpath direct failure identification

- Interesting for detecting of soft degradations with slow evolution
- Helps identifying the physical effects underlying the failure (e.g. filtering, nonlinearities)
- In case of hard failures of amplifiers and fibers, several lightpaths are affected, and a correlation algorithm is required for failure isolation







Estimation of

- Q-Factor
- SNR
- CD
- Filtering
- Etc.

Failure Management

Th1F.2.pdf OFC 2020 © OSA 2020

Soft-Failure Localization and Device Working Parameters Estimation in Disaggregated Scenarios

S. Barzegar[1], E. Virgillito[2], M. Ruiz[1], A. Ferrari[2], A. Napoli[3], V. Curri[2], and L. Velasco[1*]

[1] Optical Communications Group (GCO), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
[2] Politecnico di Torino, Turin, Italy [3]Infinera, Munich, Germany
e-mail: lvelasco@ac.upc.edu

- Uses clustering of ligthpaths with similar SNR
- If there are outliers, tries to find common resources among them
- Uses two types of monitors, for lightpaths and for devices
- Uses GNPy planning as a reference of the SNR
- The accuracy of the algorithms is not statistically evaluated

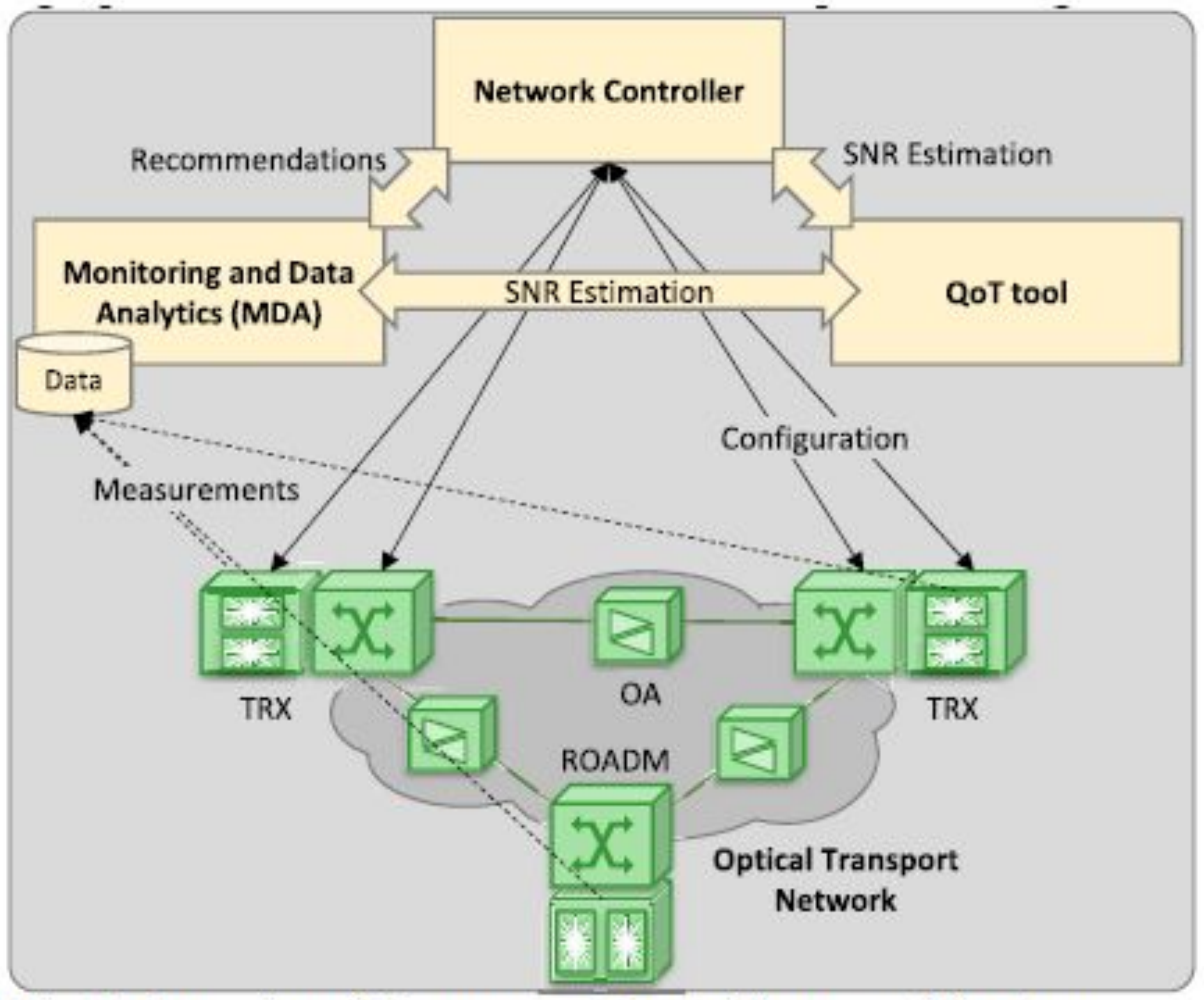


Fig. 1. Overview of the proposed surveillance architecture

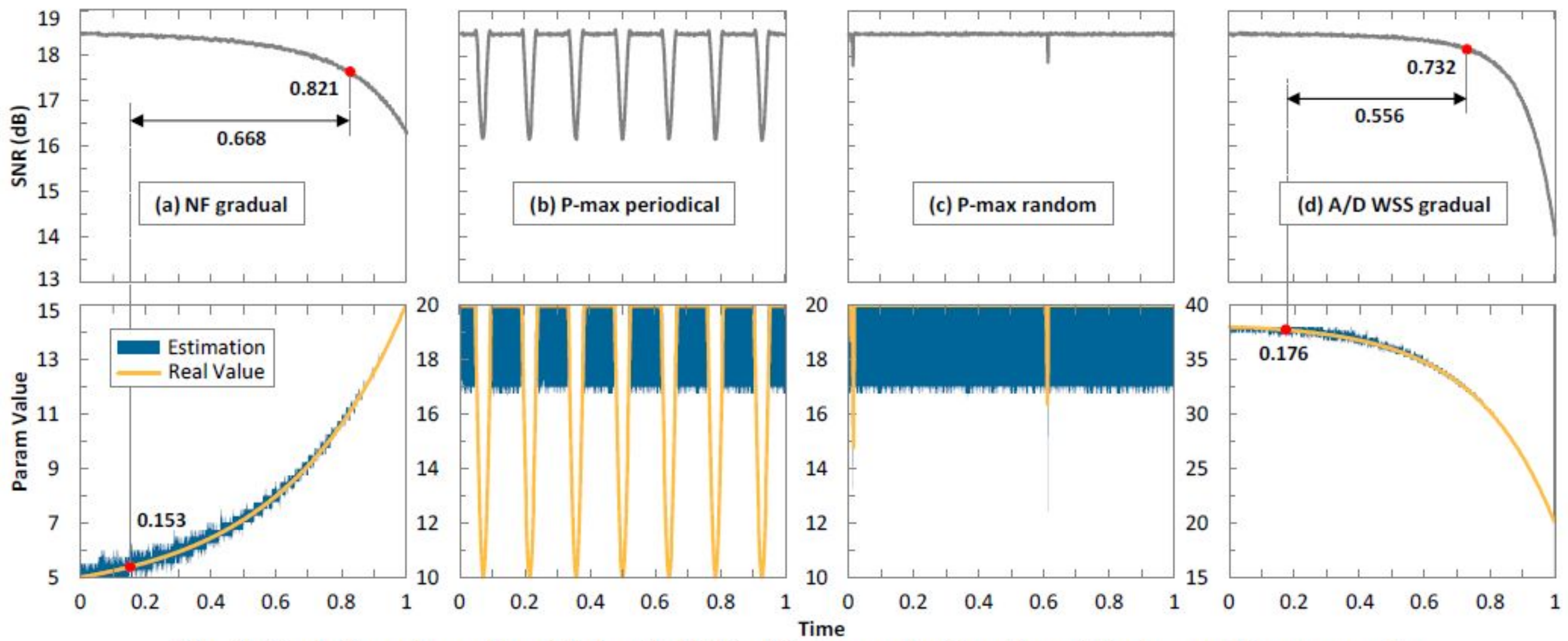


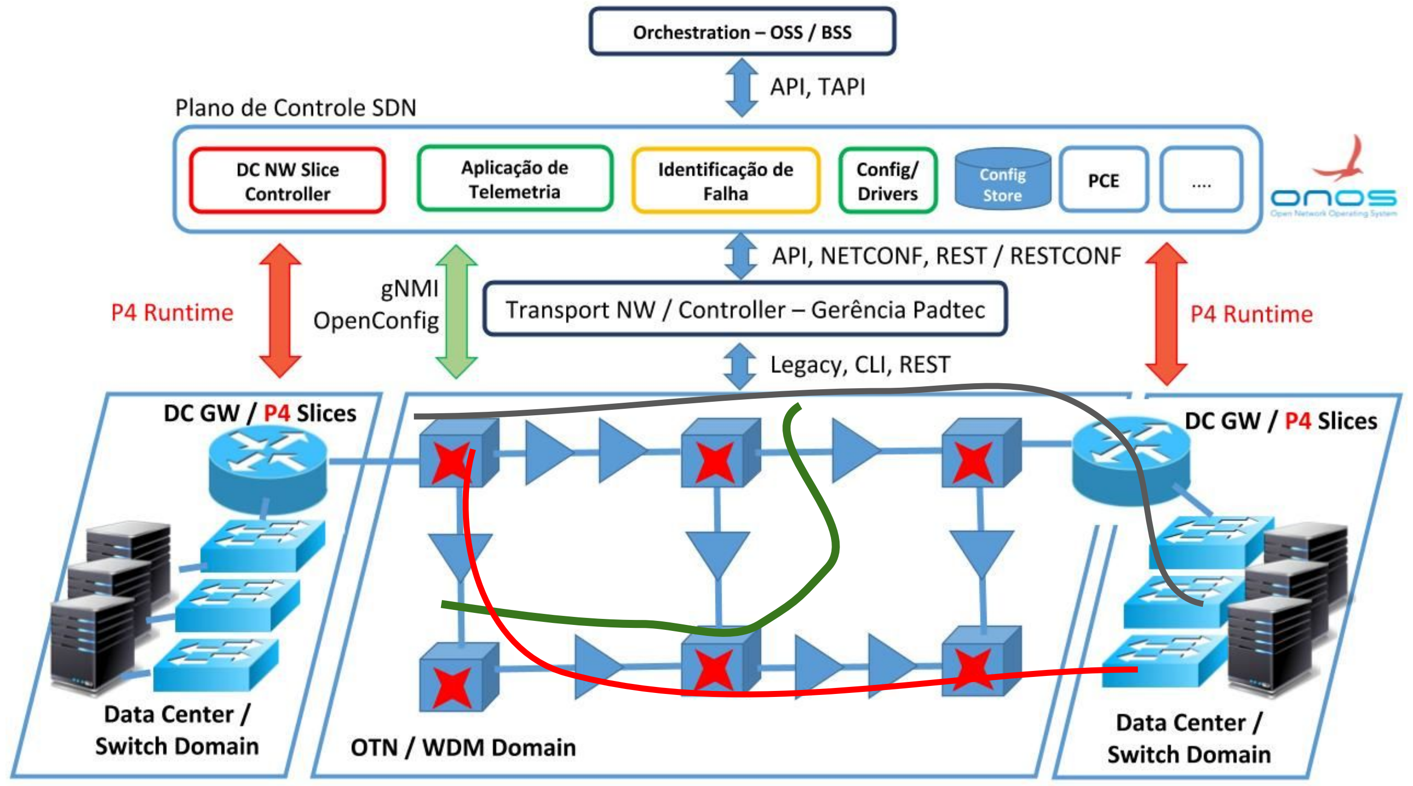
Fig. 2. Evolution of monitored lighpath SNR with time and estimation of device working parameters.

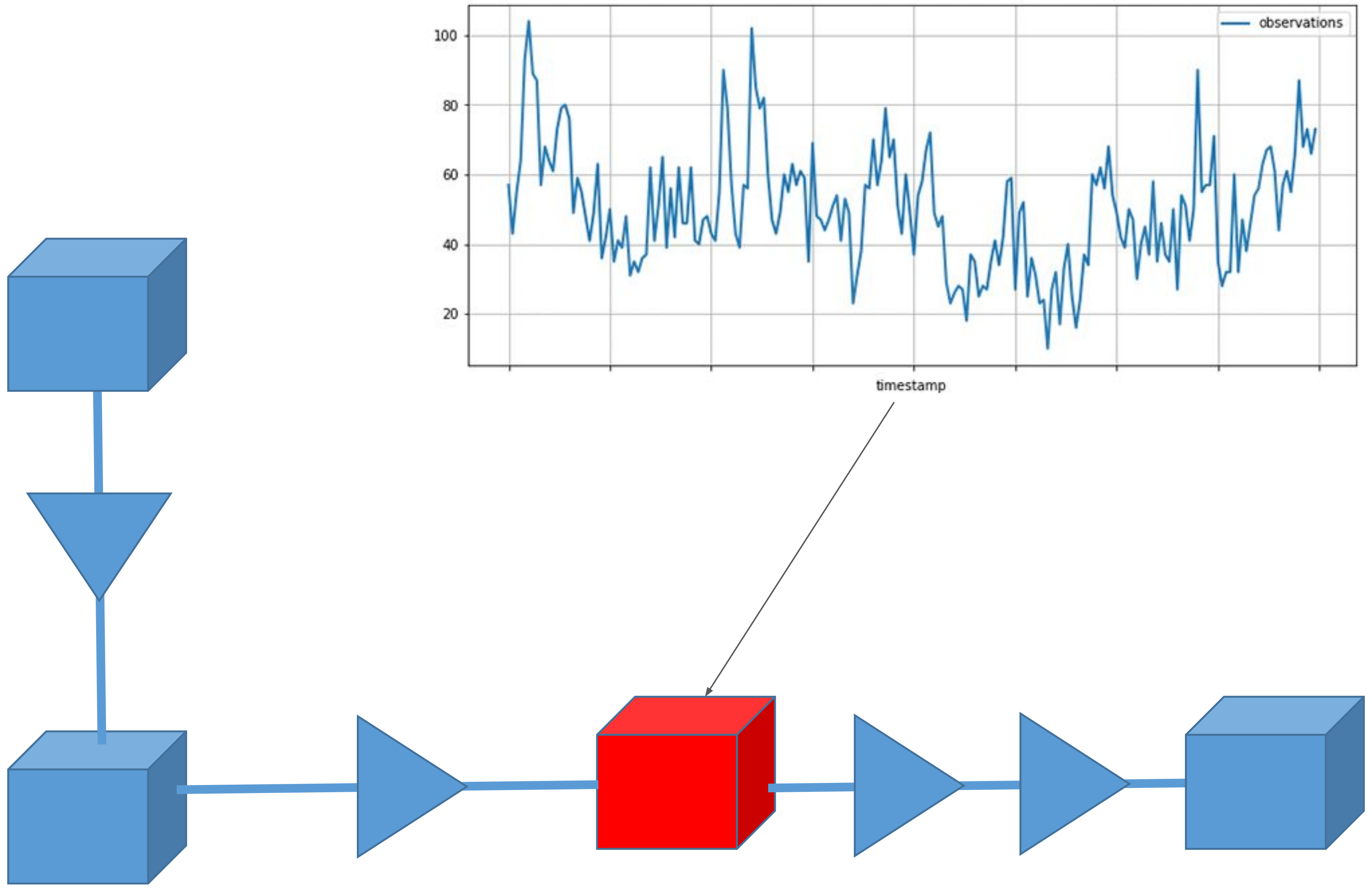
UNICAMP

Failure Management

Per device direct failure identification

- Monitors boards and devices (e.g. lasers)
- Works with time-series
- Ignores the network-wide effect of the failure





Failure Management

JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 36, NO. 7, APRIL 1, 2018 1443

Cognitive Assurance Architecture for Optical Network Fault Management

Danish Rafique, Thomas Szyrkowiec, Helmut Grießer, Achim Autenrieth, and Jörg-Peter Elbers

- Detects different types of failures (ramp, single event, small variations)
- Pre-analysis based on extreme studentized deviate tests
- Assurance based on a neural network with a 7-input layer
- Monitors different equipments individually
- Does not perform network-wide evaluation

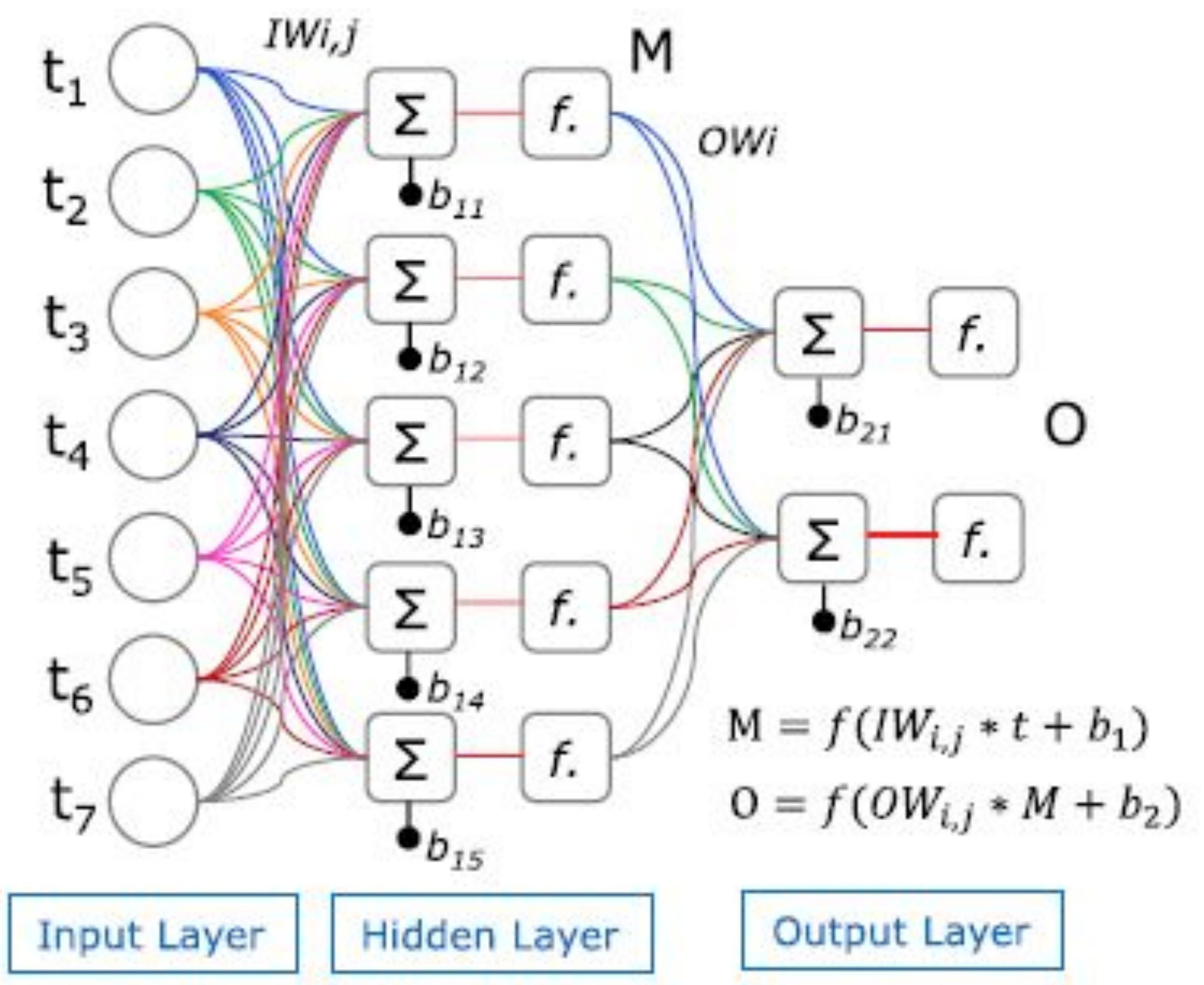


Fig. 6. Neural network architecture. t inputs, $IW_{i,j}$ are the input weights, OW_i are the output weights, b are the bias values, and f is the activation function.

Failure Management

JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 36, NO. 7, APRIL 1, 2018 1443

Cognitive Assurance Architecture for Optical Network Fault Management

Danish Rafique, Thomas Szyrkowiec, Helmut Grießer, Achim Autenrieth, and Jörg-Peter Elbers

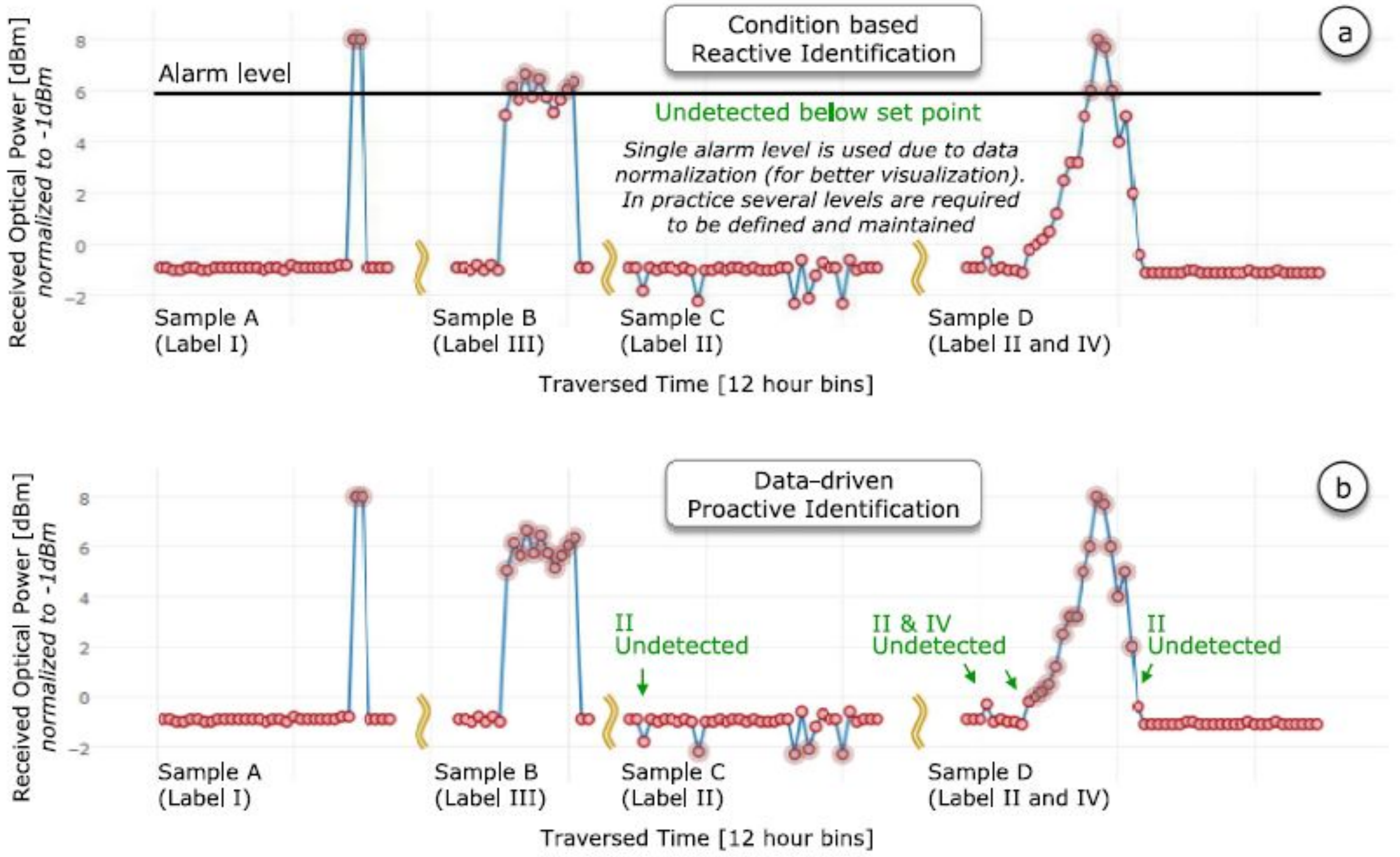


TABLE I
FAULT TYPES TYPICALLY ENCOUNTERED IN COMMERCIAL OPTICAL NETWORKING SYSTEMS

Fault Label	Description
I	Point abnormalities due to random flash events and may lead to abrupt device damage
II	Local abnormalities indicating potential flaws with potential long-term impact on service performance
III	Steady abnormalities due to preceding system configuration changes, and may lead to damage and/or consistent performance loss
IV	Ramp abnormalities representing gradual system and/or service distortion possibilities

Failure Management

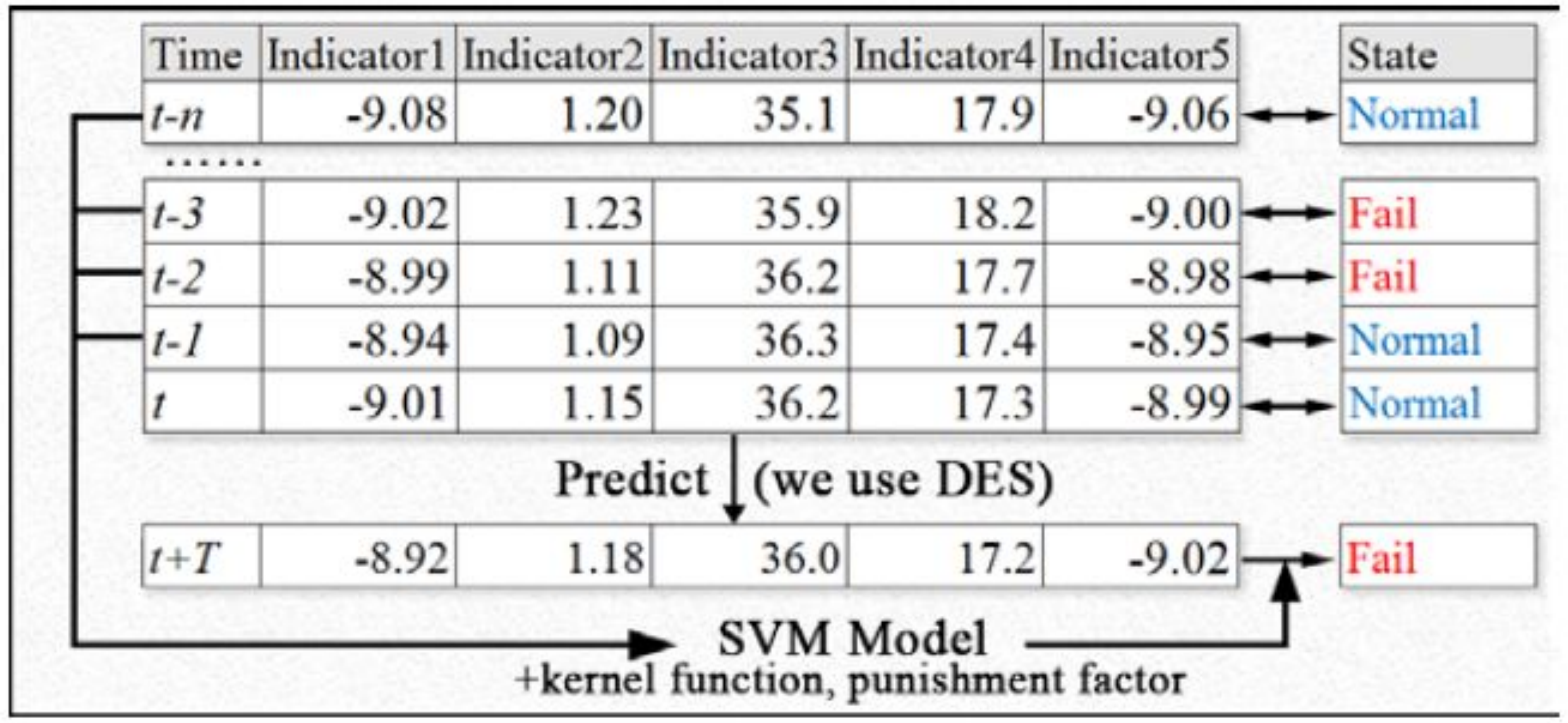
Research Article Vol. 25, No. 16 | 7 Aug 2017 | OPTICS EXPRESS 18553
Optics EXPRESS

Failure prediction using machine learning and time series in optical network

ZHILONG WANG,[1] MIN ZHANG,[1,*] DANSHI WANG,[1] CHUANG SONG,[1] MIN LIU,[1] JIN LI,[1] LIQI LOU,[1] AND ZHUO LIU[2]

[1]*State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*China Mobile Communications Corporation, Beijing 100033, China*
[*]*mzhang@bupt.edu.cn*

Indicator Name	Units
Input Optical Power (IOP)	dBm
Laser Bias Current (LBC)	mA
Laser Temperature Offset (LTO)	°C
Output Optical Power (OOP)	dBm
Environmental Temperature (ET)	°C
Unusable Time	s



- Monitors boards of China Telecom to predict failures in the next day
- Uses SVM+Time series processing
- Identifies the parameters (laser bias, laser temperature, environmental temperature) most related to the failures
- Achieves high prediction levels

UNICAMP

Failure Management

Research Article Vol. 25, No. 16 | 7 Aug 2017 | OPTICS EXPRESS 18553

Optics EXPRESS

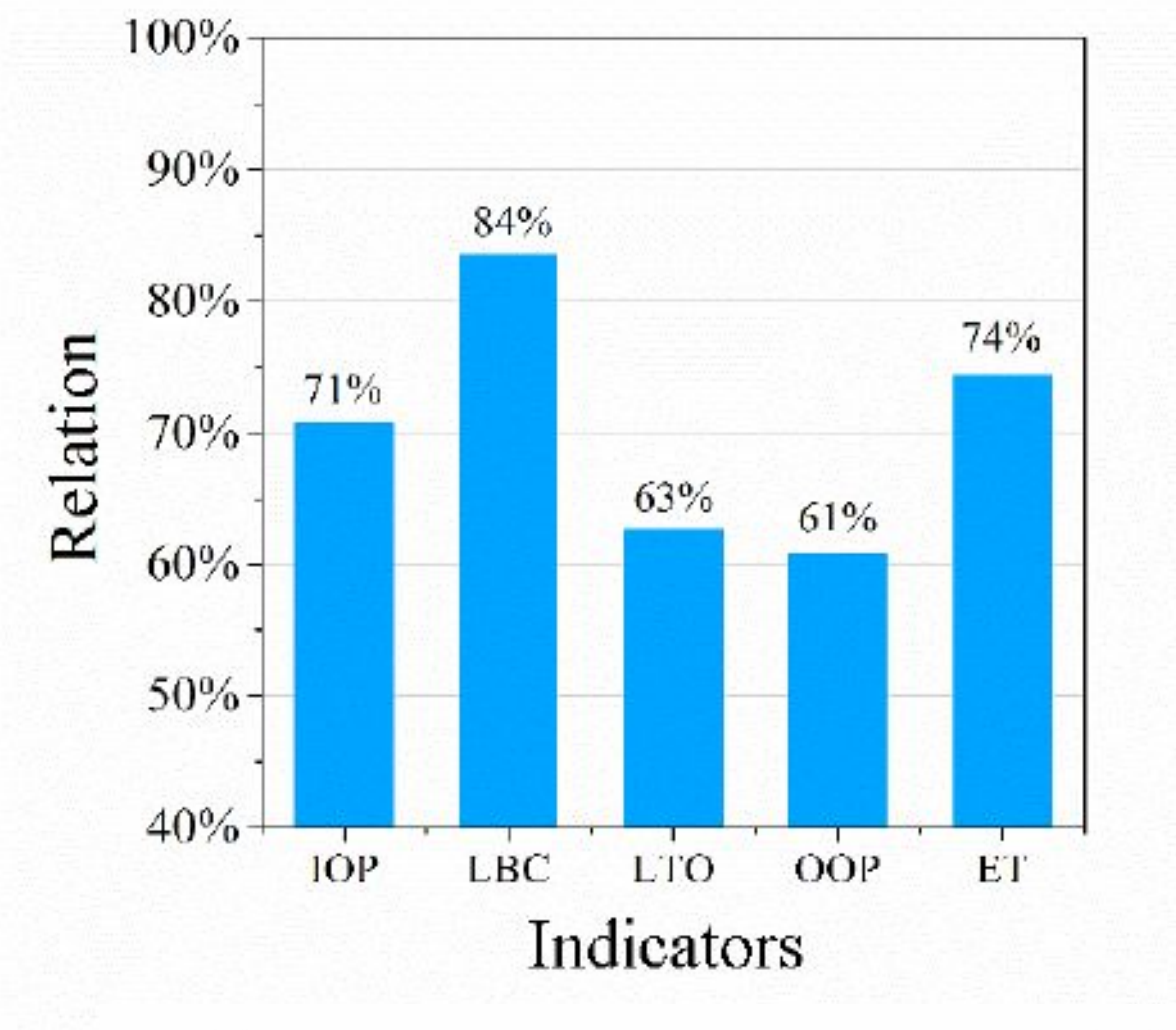
Failure prediction using machine learning and time series in optical network

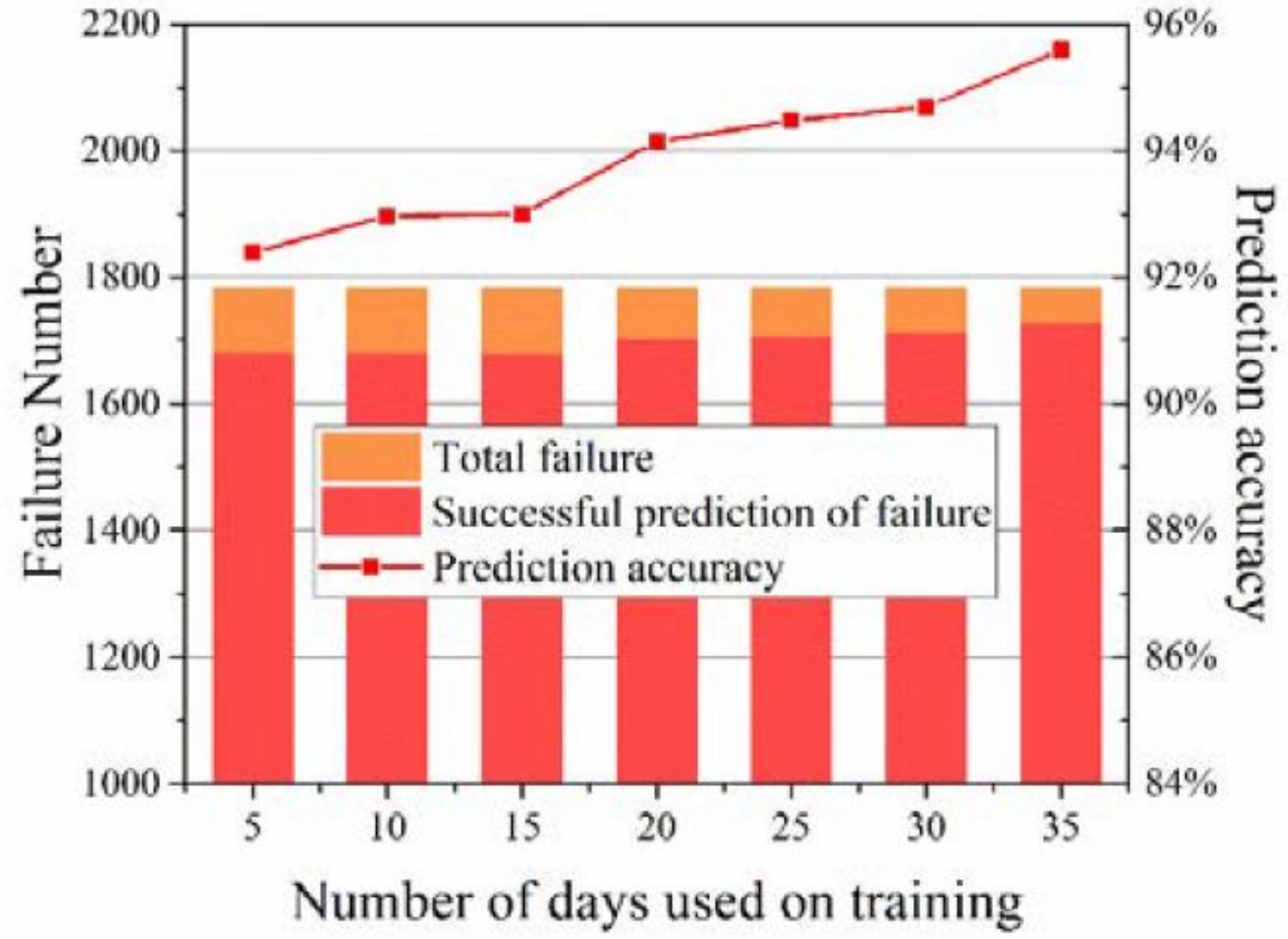
ZHILONG WANG,[1] MIN ZHANG,[1,*] DANSHI WANG,[1] CHUANG SONG,[1] MIN LIU,[1] JIN LI,[1] LIQI LOU,[1] AND ZHUO LIU[2]

[1]*State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*China Mobile Communications Corporation, Beijing 100033, China*
[*]*mzhang@bupt.edu.cn*

Indicator Name	Units
Input Optical Power (IOP)	dBm
Laser Bias Current (LBC)	mA
Laser Temperature Offset (LTO)	°C
Output Optical Power (OOP)	dBm
Environmental Temperature (ET)	°C
Unusable Time	s

- Monitors boards of China Telecom to predict failures in the next day
- Uses SVM+Time series processing
- Identifies the parameters (laser bias, laser temperature, environmental temperature) most related to the failures
- Achieves high prediction levels

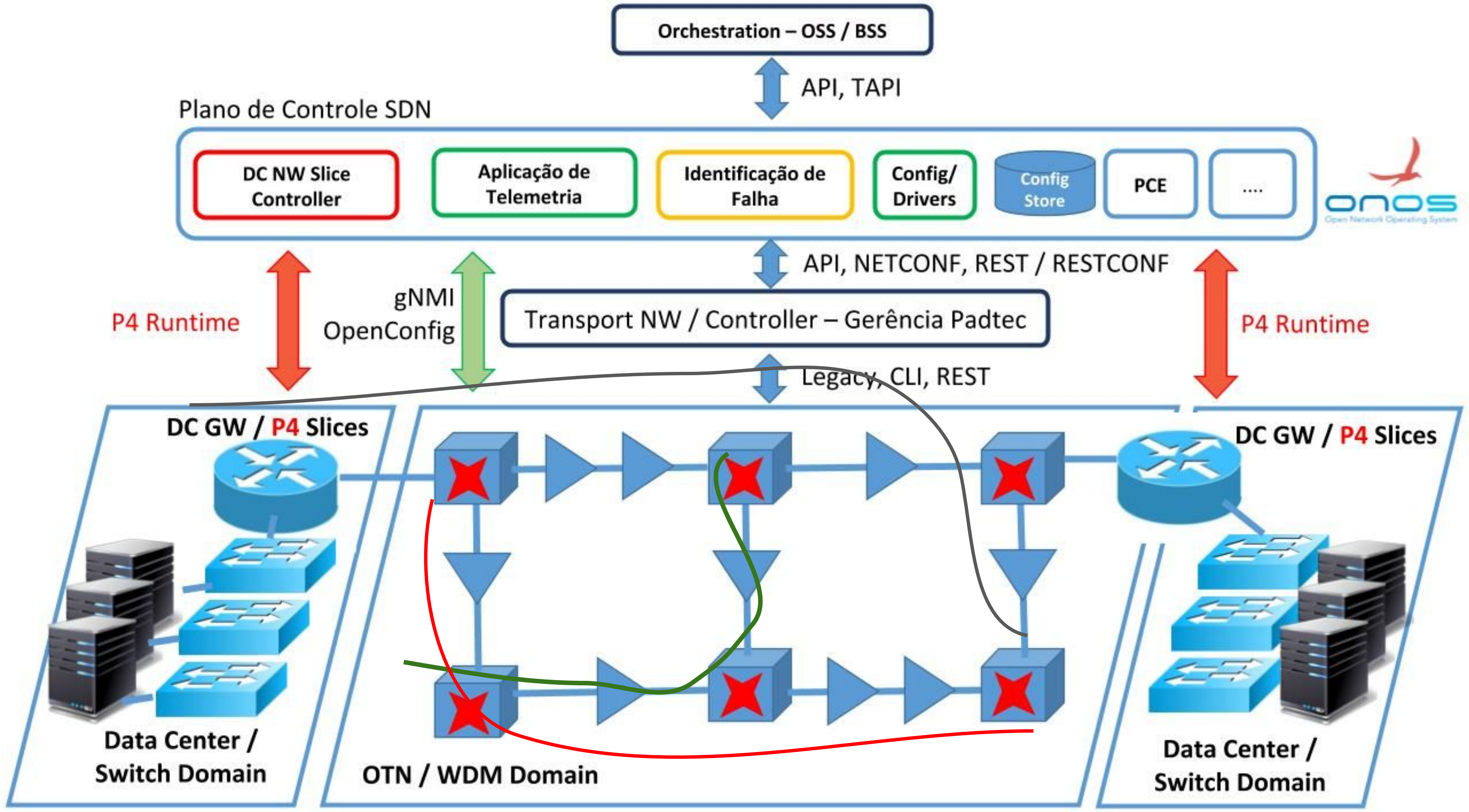




Failure Management

Network-wide failure identification (our approach)

- Identifies and isolates failures based on telemetry carried out over the whole network
- Currently, works with the instantaneous network status
- Identifies problems in current status compared to a training status
- Currently, we are evaluating the capacity of generalization of the algorithm with respect to different types/levels of failures

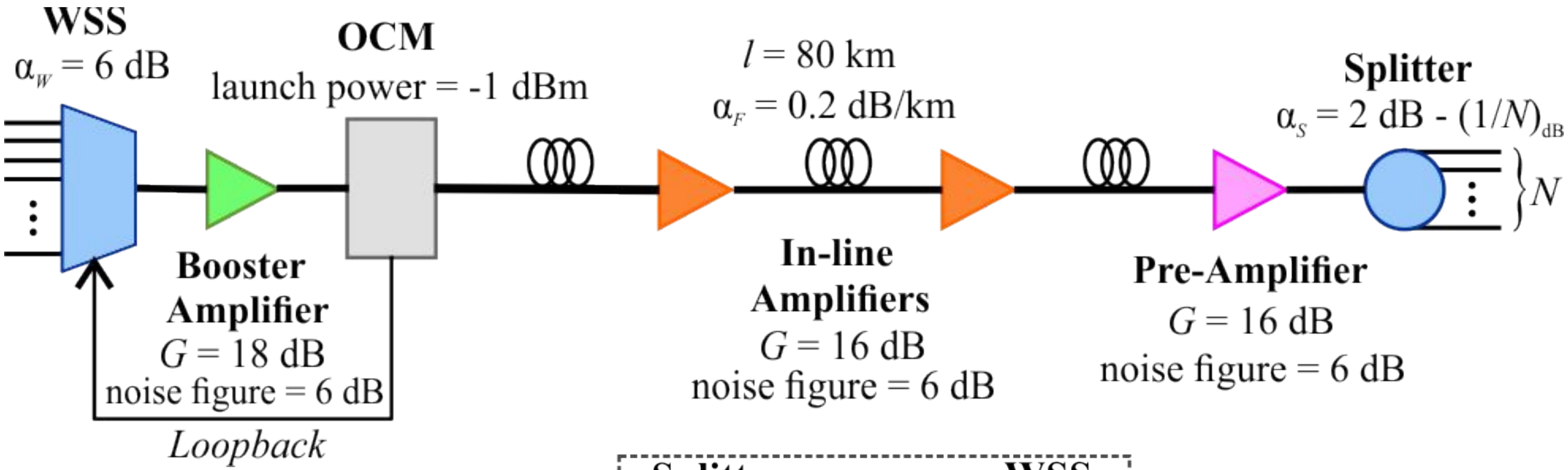


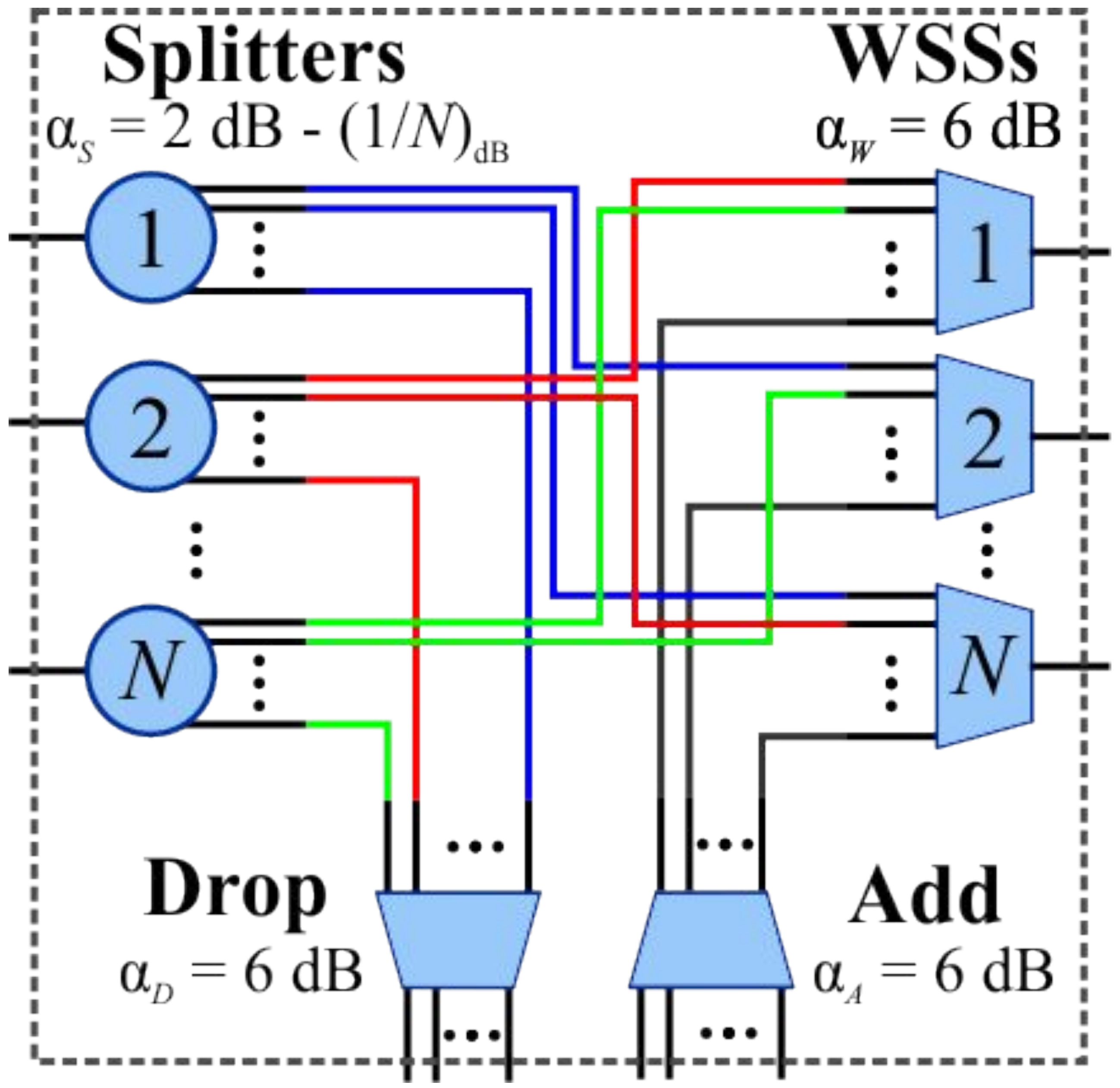
Load Simulator

Components:
Transceivers;
Fibers;
EDFA amplifiers;
ROADMs:

- Splitters;
- WSSs;
- Adds & Drops;
- OCMs.

Algorithms:
Python
NetworkX - Network model;
Dijkstra - Routing;
First-fit - Spectrum allocation.





Transceiver power = -1dBm

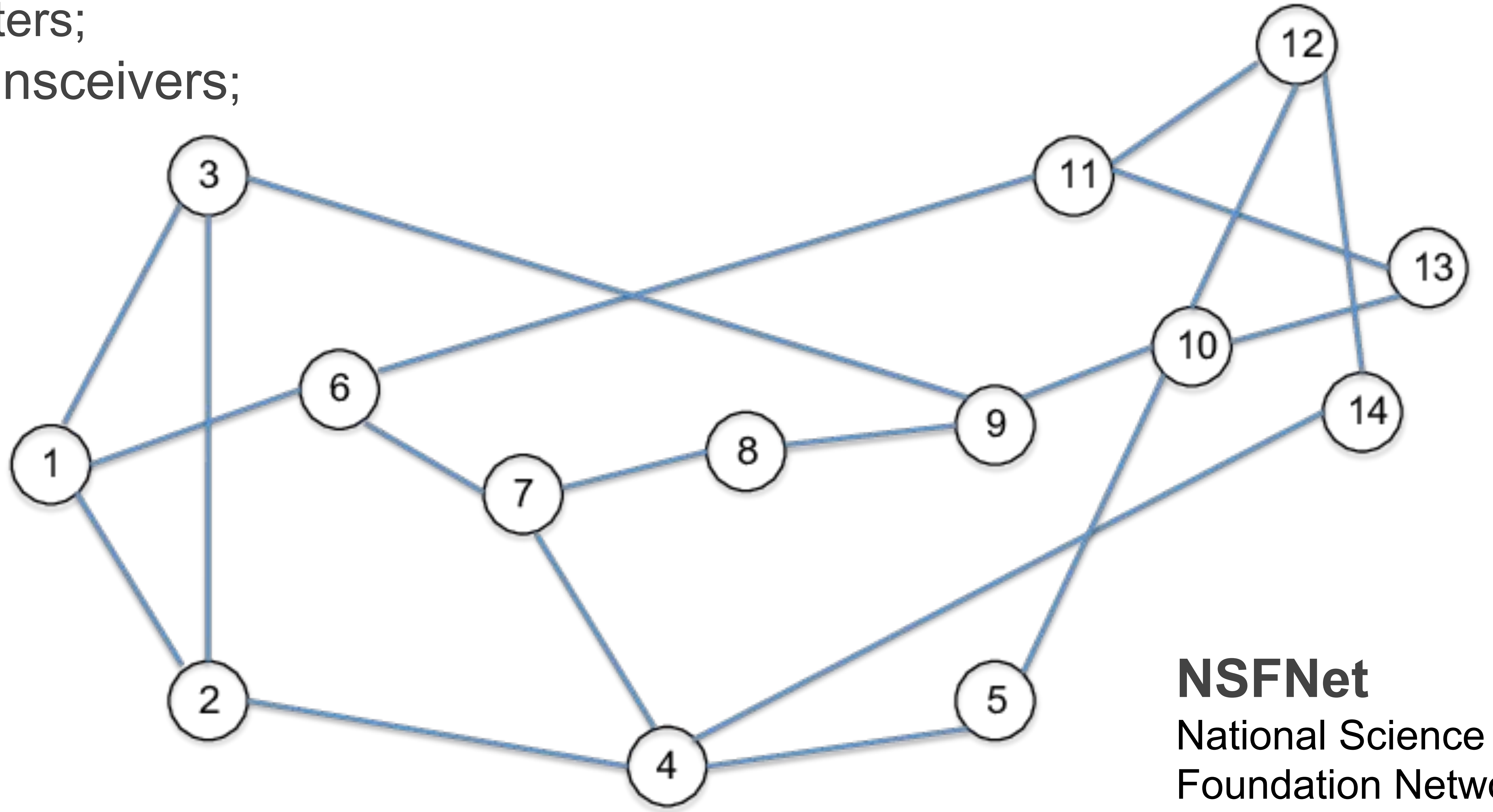
Load Simulator

2136 Components:

- 580 Unidirectional fiber spam;
- 624 Amplifiers;
- 42 OCMs;
- 72 WSSs;
- 44 Splitters;
- 772 Transceivers;

Requested services:

- 1000 demands;
- 386 accepted;



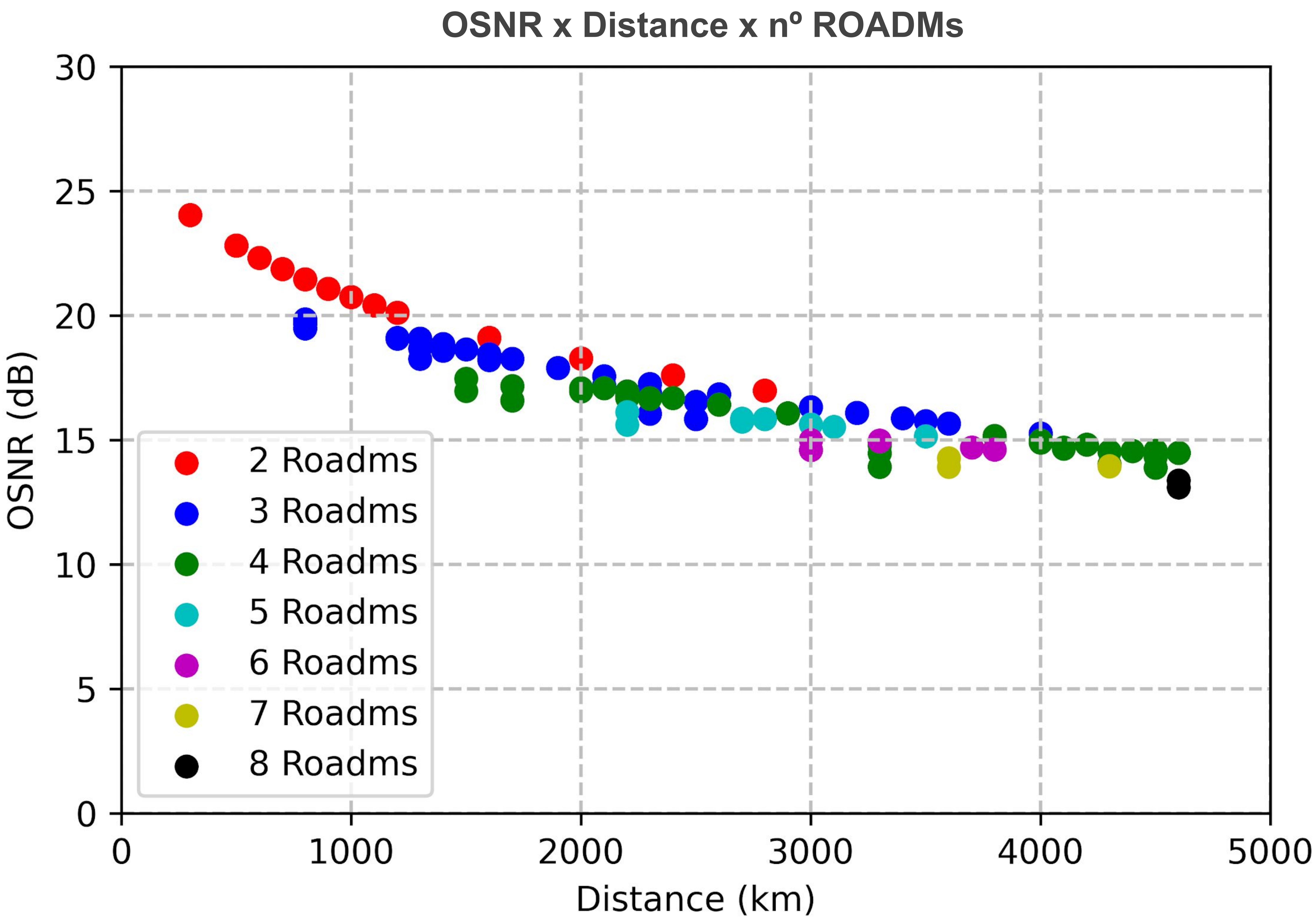
NSFNet
National Science
Foundation Network
USA

Links

Source	Destination	Distance (km)
Seattle (WA)	Palo Alto (CA)	1100
Seattle (WA)	San Diego (CA)	1600
Seattle (WA)	Champaing (IL)	2800
Palo Alto (CA)	Salt Lake City (UT)	1000
Palo Alto (CA)	San Diego (CA)	600
San Diego (CA)	Houston (TX)	2000
Salt Lake City (UT)	Boulder (CO)	600
Salt Lake City (UT)	Ann Arbour (MI)	2400
Boulder (CO)	Lincoln (NE)	800
Boulder (CO)	Houston (TX)	1100
Lincoln (NE)	Champaing (IL)	700
Champaing (IL)	Pittsburg (PA)	700
Houston (TX)	Atlanta (GA)	1200
Houston (TX)	College Pk (MD)	2000
Atlanta (GA)	Pittsburg (PA)	900
Pittsburg (PA)	Ithaca (NY)	500
Pittsburg (PA)	Princeton (NJ)	500
College Pk (MD)	Ithaca (NY)	500
College Pk (MD)	Princeton (NJ)	300
Ann Arbour (MI)	Ithaca (NY)	800
Ann Arbour (MI)	Princeton (NJ)	1000

UNICAMP

Load Simulator



Load Simulator

Monitorable parameters

Component		Monitored parameters	Equipment	Total monitored parameters
Transceivers	Pin	3	772	2316
	Pout			
	OSNR			
OCMs	Pout/Channel	96	42	4032
Amplifiers	Pin	2	624	1248
	Pout			
				7596

UNICAMP

Fault Simulator

Device faults table generated in the simulation

Component	Fault	Quantity
Transceivers	Pout = 0 W	772
Amplifiers	Gain = 0 dB	624
Fibers	Pout = 0 W	582
	TOTAL	1978

Load and Fault Simulator

Output files

Simulador de carga	
trx_telemetry.csv amp_telemetry.csv ocm_telemetry.csv	Telemetry files
noise_power_out.csv noise_power_in.csv signal_power_out.csv signal_power_in.csv	Auxiliary files
frequency_slots_per_fiber.csv blocked_lightpaths.csv accepted_lightpaths.csv frequency_slots_per_lightpath.csv lightpaths.csv	Network files

Simulador de falha
output_data.csv input_data.csv

Load and Fault Simulator

ML - Input data

	OSNR_Trx_14_1	OSNR_Trx_13_1	OSNR_Trx_13_2	OSNR_Trx_2_1	...	PowerIn_Amp_13_9_9	PowerIn_Amp_13_9_10	PowerIn_Amp_13_9_11	PowerIn_Amp_13_9_12
0	24,037	24,037	13,954	13,940	...	-1,857	-1,857	-1,857	-1,857
1	24,037	-inf	13,954	13,940	...	-1,857	-1,857	-1,857	-1,857
2	-inf	24,037	13,954	13,940	...	-1,857	-1,857	-1,857	-1,857
3	24,037	24,037	13,954	-inf	...	-1,857	-1,857	-1,857	-1,857
4	24,037	24,037	-inf	13,940	...	-1,857	-1,857	-1,857	-1,857
...	...	...	...	...	...	...	...	...	...
1975	24,037	24,037	13,954	13,940	...	-inf	-inf	-inf	-inf
1976	24,037	24,037	13,954	13,940	...	-1,857	-inf	-inf	-inf
1977	24,037	24,037	13,954	13,940	...	-1,857	-1,857	-inf	-inf
1978	24,037	24,037	13,954	13,940	...	-1,857	-1,857	-1,857	-inf

ML - Output data

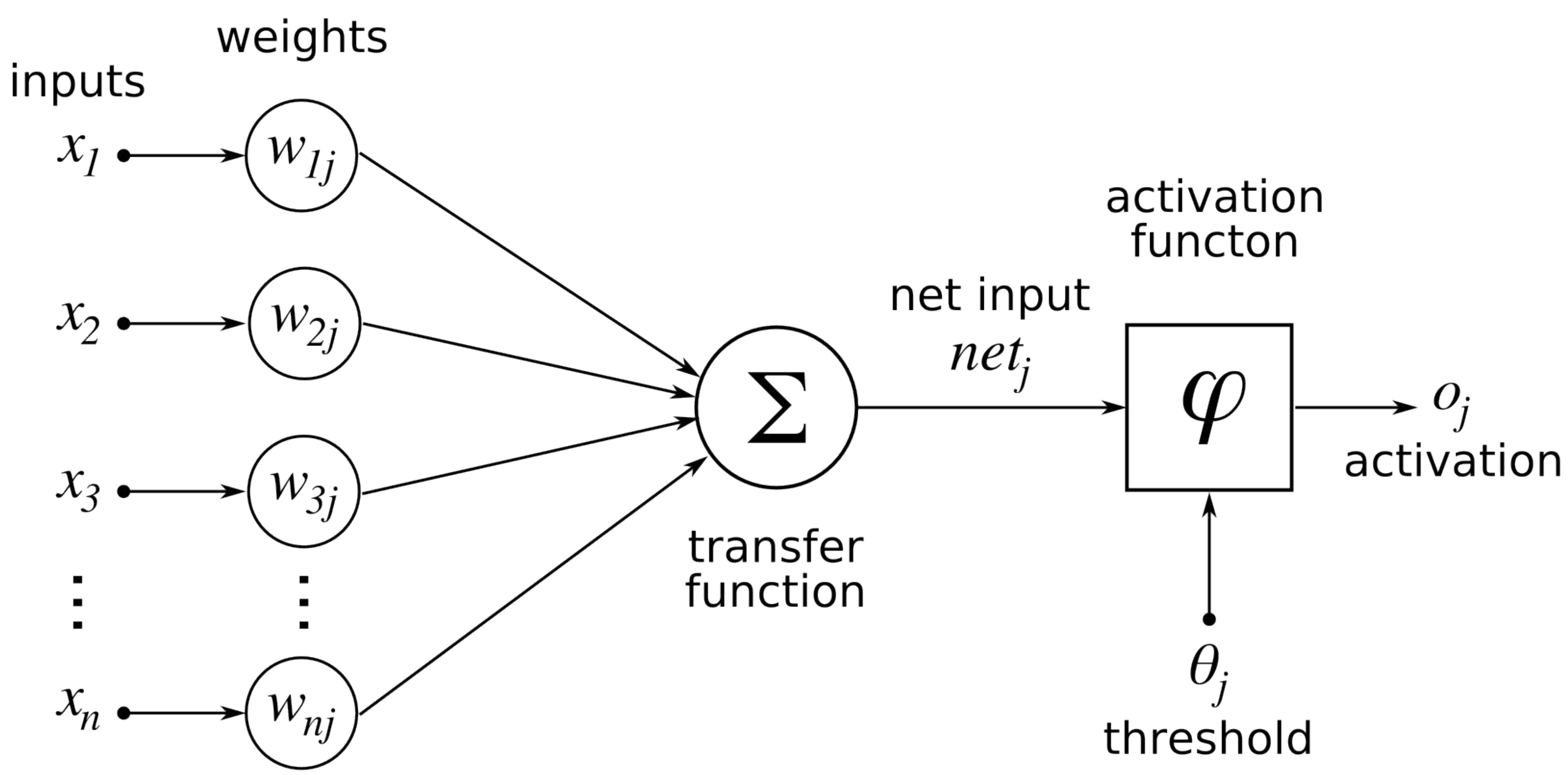
	Trx_14_1	Trx_13_1	Trx_13_2	Trx_2_1	...	Fiber_Amp_13_9_9	Fiber_Amp_13_9_10	Fiber_Amp_13_9_11	Fiber_Amp_13_9_12
0	0	0	0	0	...	0	0	0	0
1	1	0	0	0	...	0	0	0	0
2	0	1	0	0	...	0	0	0	0
3	0	0	1	0	...	0	0	0	0
4	0	0	0	1	...	0	0	0	0
...	...	...	...	...	...	...	...	...	...
1975	0	0	0	0	...	1	0	0	0
1976	0	0	0	0	...	0	1	0	0
1977	0	0	0	0	...	0	0	1	0
1978	0	0	0	0	...	0	0	0	1

UNICAMP

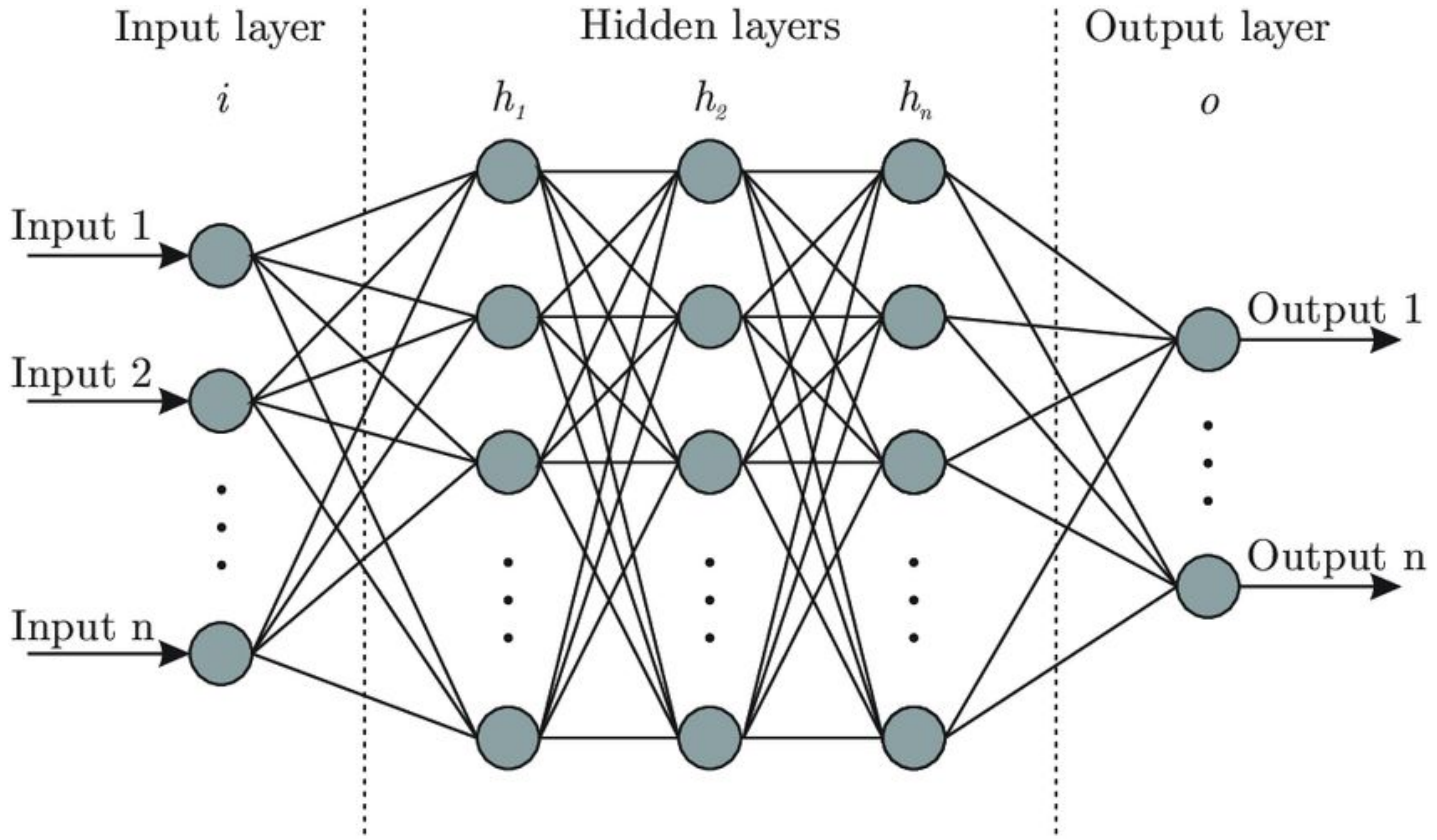
Neural Networks

- Nonlinear filters;
- Universal approximators of any continuous function on compact (closed and bounded) subsets of *n*-dimensional Euclidian space.

Artificial neuron



Artificial neural network



Normalization

- Does not change the type of distribution;

$$\hat{X} = aX + b \rightarrow f_{\hat{X}}(x) = \frac{1}{|a|} f\left(\frac{x-b}{a}\right)$$

- Improves the numerical stability of the model;
- May speed up the training process;
- Large input values saturate activation functions (e.g., sigmoid and ReLu).

Z-score normalization

$$\hat{X} = \frac{X - \mu_X}{\sigma_X}$$

$$\mu_{\hat{X}} = 0$$

$$\sigma_{\hat{X}} = 1$$

Nonlinear Activation Functions

- Without a nonlinear activation function in the artificial neural network, no matter how many layers it had, it would behave just like a linear single-layer perceptron;
- Allow the model to create complex mappings between the network's inputs and outputs.

Softmax

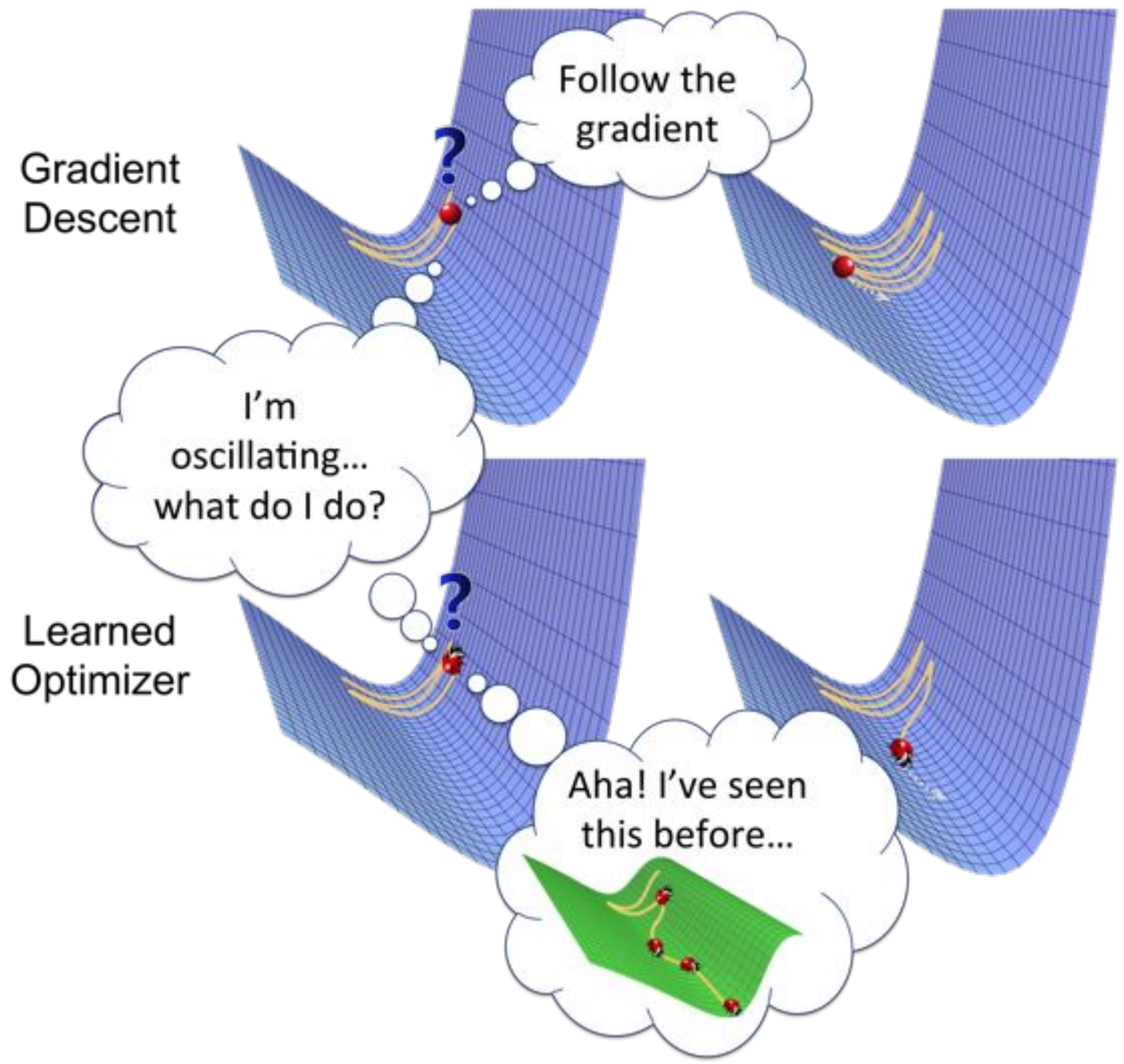
$$o_j = \frac{net_j}{\sum_{n=1}^{N} net_n}$$

Optimizers

- Are algorithms or methods used to change the attributes of neural networks such as weights, learning rate, and momentum in order to reduce losses;

  Adam (Adaptive Moment Estimation):
  Uses estimations of first and second moments of gradient to adapt the learning rate for each weight of the neural network.

  Adamax:
  Based on Adam, Adamax replaces the second-order moment with the infinite order moment.



Loss Functions

- Are functions that map an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event;

  Regression loss functions: continuous values (e.g., MSE, MAE);
  Classification loss functions: finite categorical values (e.g., cross-entropy, categorical cross-entropy);

  Cross-entropy: $$CE = -\sum_{n=1}^{N} d_n \log(o_n)$$

  Categorical cross-entropy: $$CCE = -\sum_{n=1}^{N} d_n \log\left(\frac{e^{o_n}}{\sum_{i=1}^{N} e^{o_i}}\right)$$

Accuracy Metrics

- Accuracy is a metric for evaluating classification models. Basically, accuracy is the fraction of predictions our model got right.

  Categorical accuracy: checks to see if the index of the maximal true value is equal to the index of the maximal predicted value.

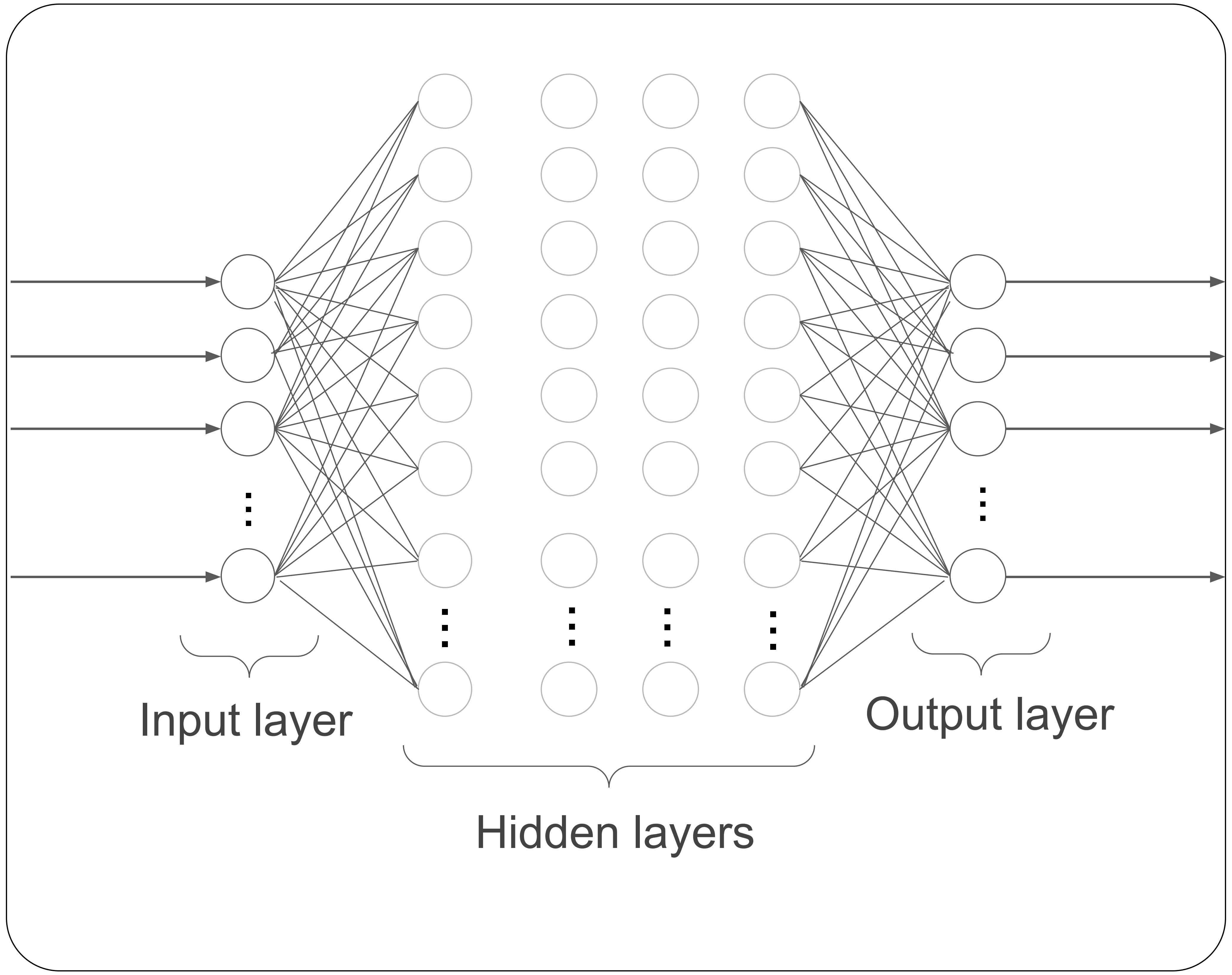
ML module

Neural Network Parameters

Input layer	Hidden layers	Output layer
7596 monitorable parameters of the network.	Configurable. In this case, one layer with 1000 neurons.	1978 equipment that may fail on the network.

Normalization of the input using "*Zscore*".
Probabilistic output with "*Softmax*".
Training optimization with "*Adamax*".
Loss metric "*categorical_crossentropy*"
Accuracy metric "*categorical_accuracy*".

Graphical representation of the Neural Network



ML module

Results

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 1000)              7597000
_________________________________________________________________
dropout (Dropout)            (None, 1000)              0
_________________________________________________________________
dense_1 (Dense)              (None, 1978)              1979978
=================================================================
Total params: 9,576,978
Trainable params: 9,576,978
Non-trainable params: 0
```

ML module

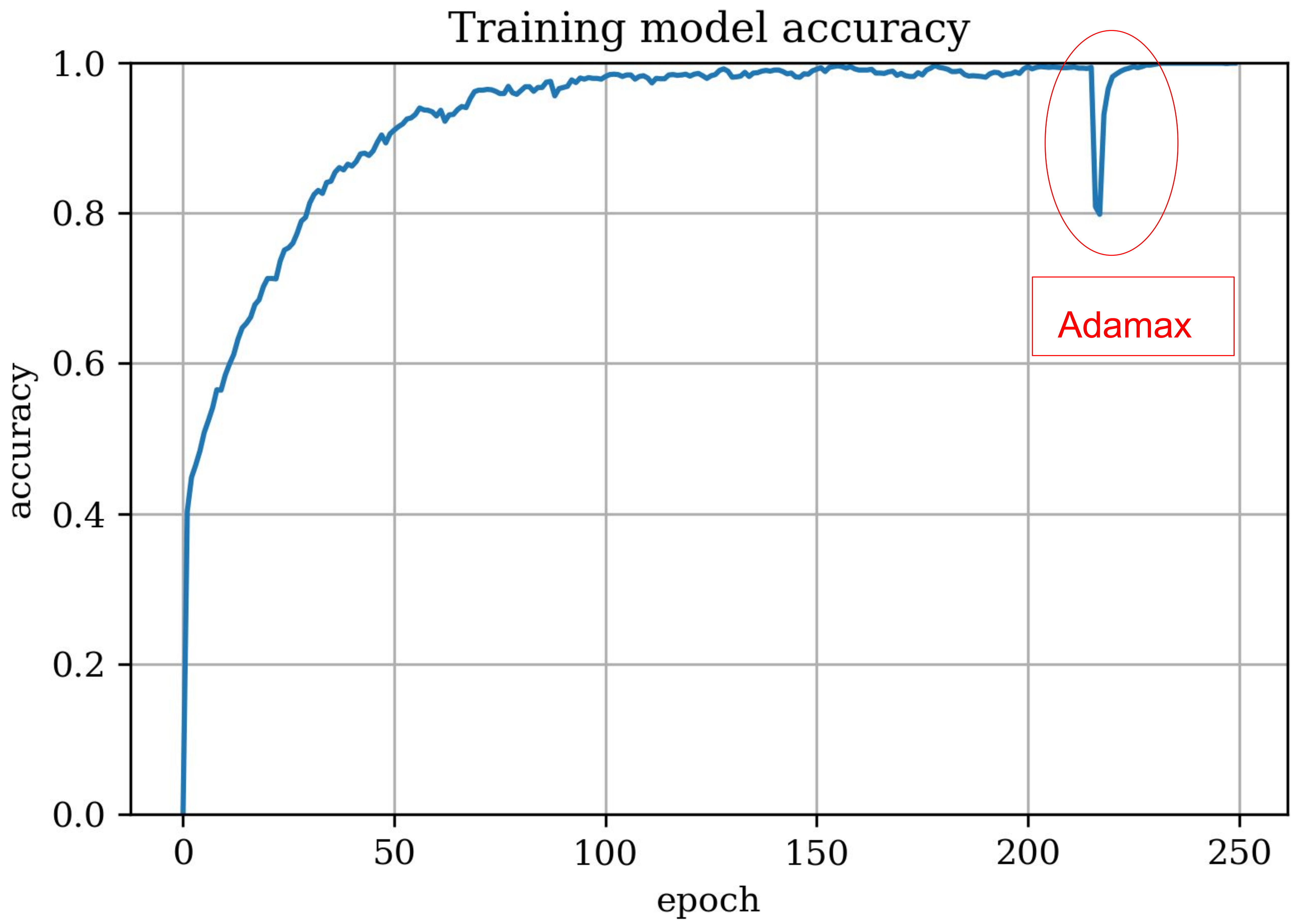
Results

Training Epochs

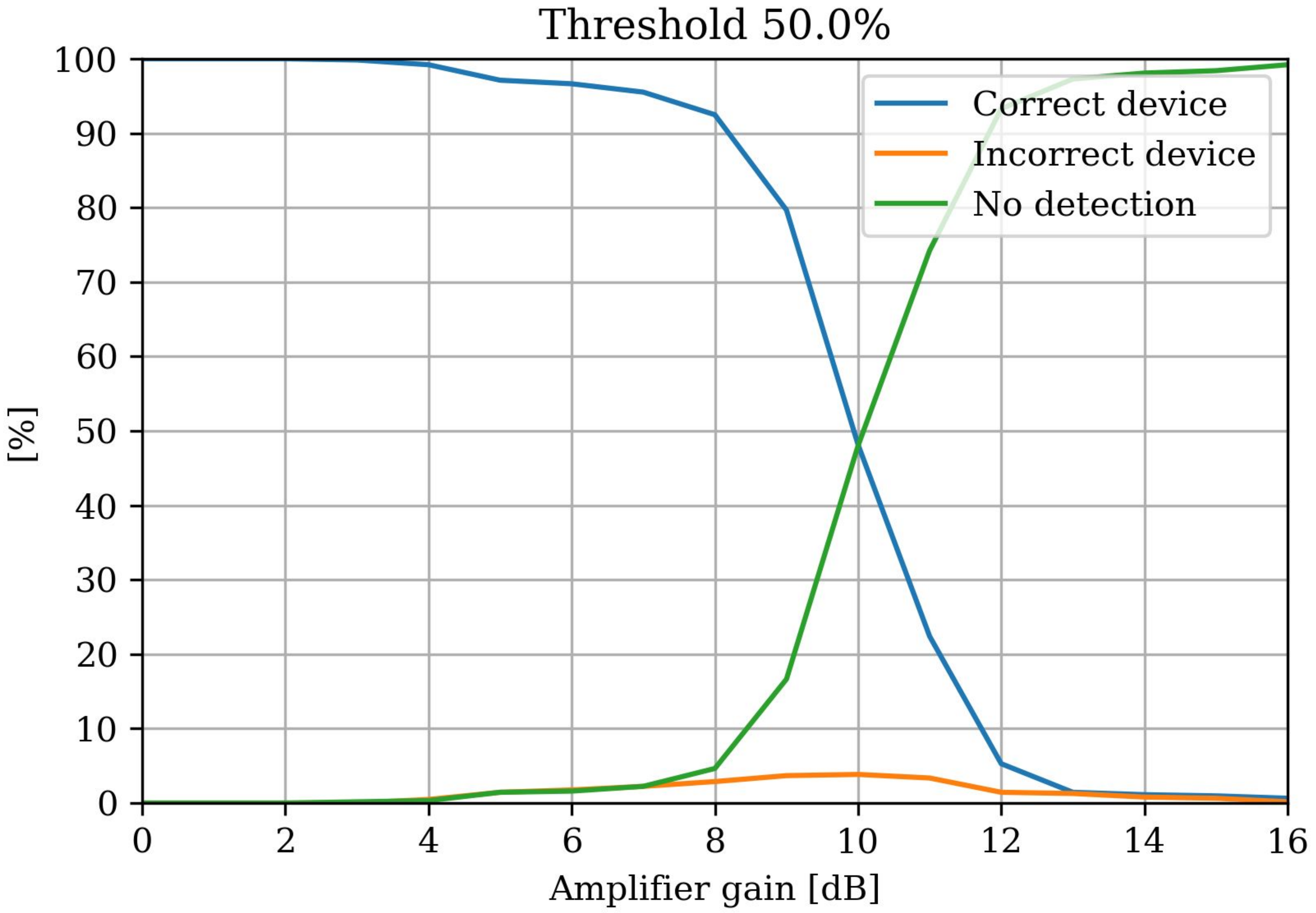
Epoch 245/250
1978/1978 [==============================] - 3s 68ms/step - loss: 0.0175 - categorical_accuracy: 0.9995
Epoch 246/250
1978/1978 [==============================] - 3s 68ms/step - loss: 0.0172 - categorical_accuracy: 0.9995
Epoch 247/250
1978/1978 [==============================] - 3s 67ms/step - loss: 0.0171 - categorical_accuracy: 0.9995
Epoch 248/250
1978/1978 [==============================] - 3s 65ms/step - loss: 0.0171 - categorical_accuracy: 0.9995
Epoch 249/250
1978/1978 [==============================] - 3s 65ms/step - loss: 0.0168 - categorical_accuracy: 0.9995
Epoch 250/250
1978/1978 [==============================] - 3s 69ms/step - loss: 0.0161 - categorical_accuracy: 0.9995

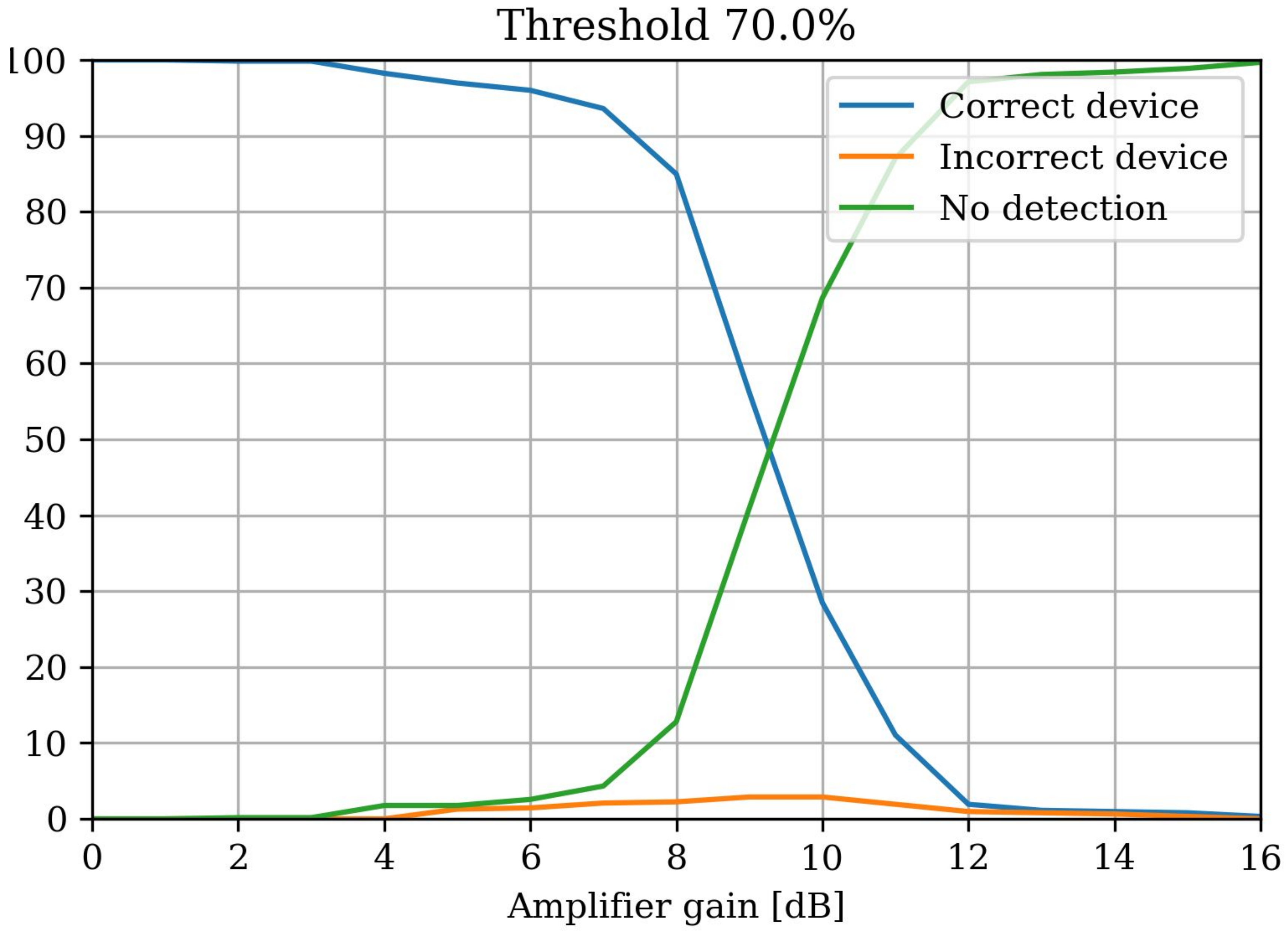
ANN test accuracy: 99.95%

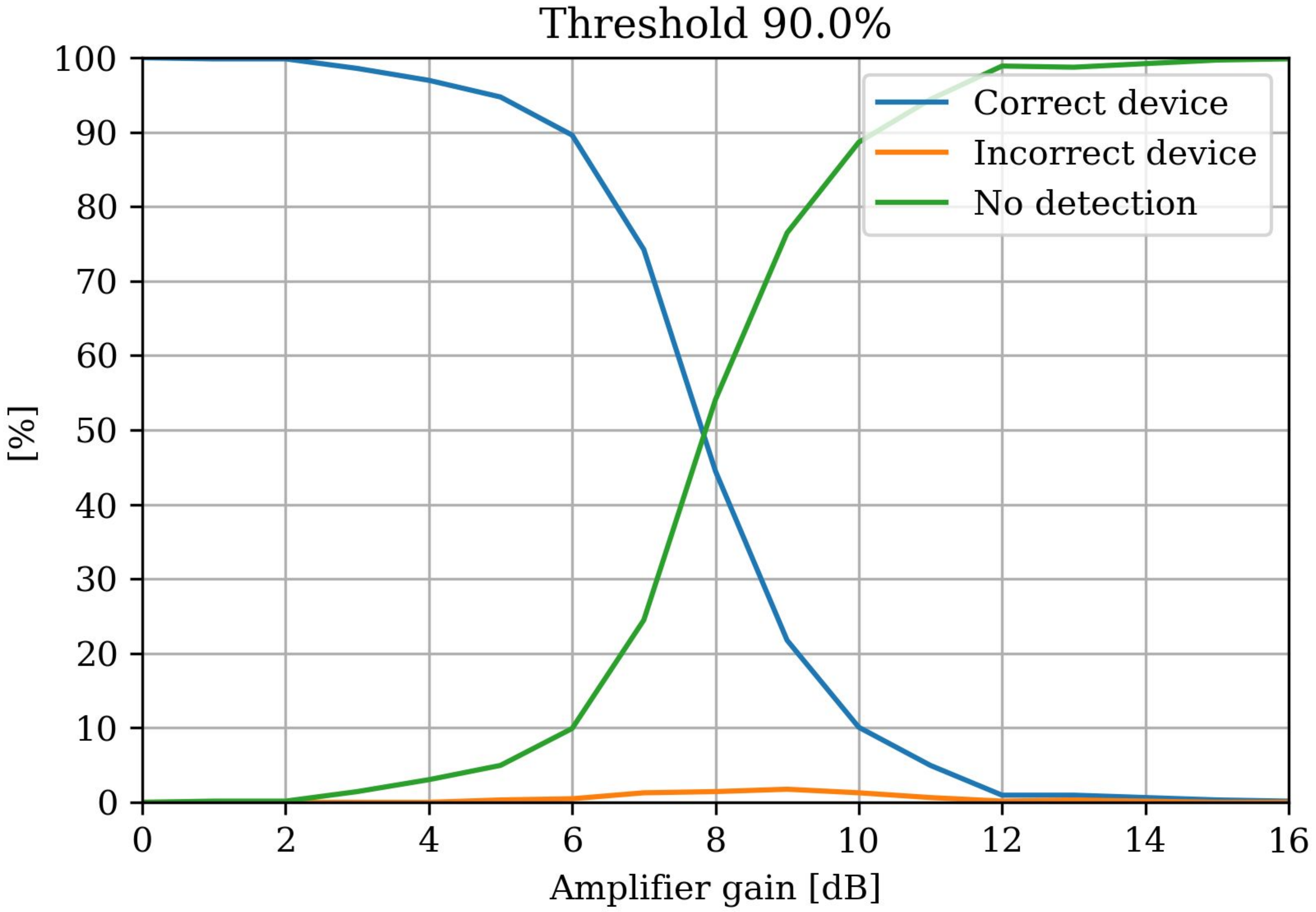
Total training time ~7.5 min

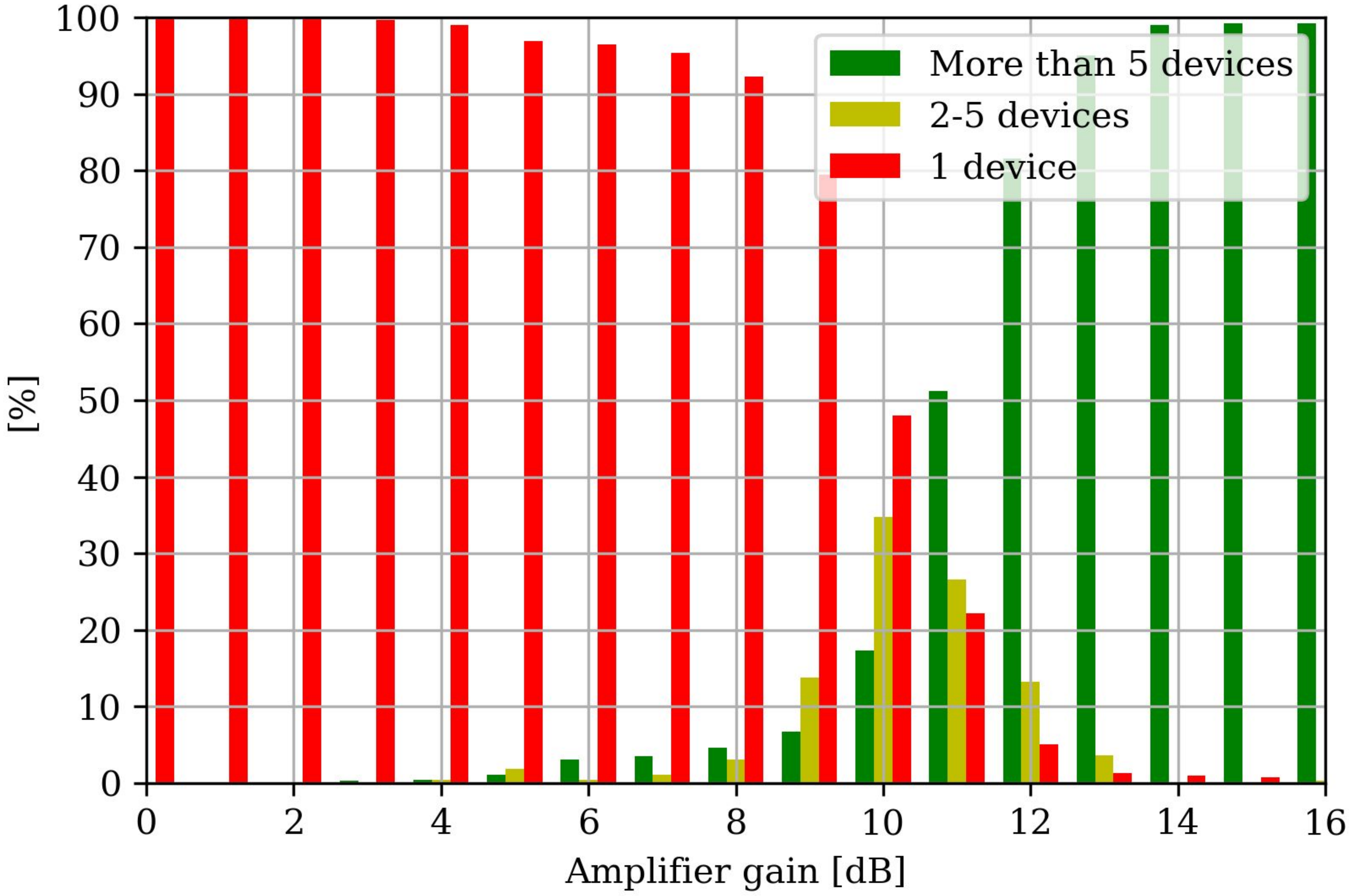


ML module

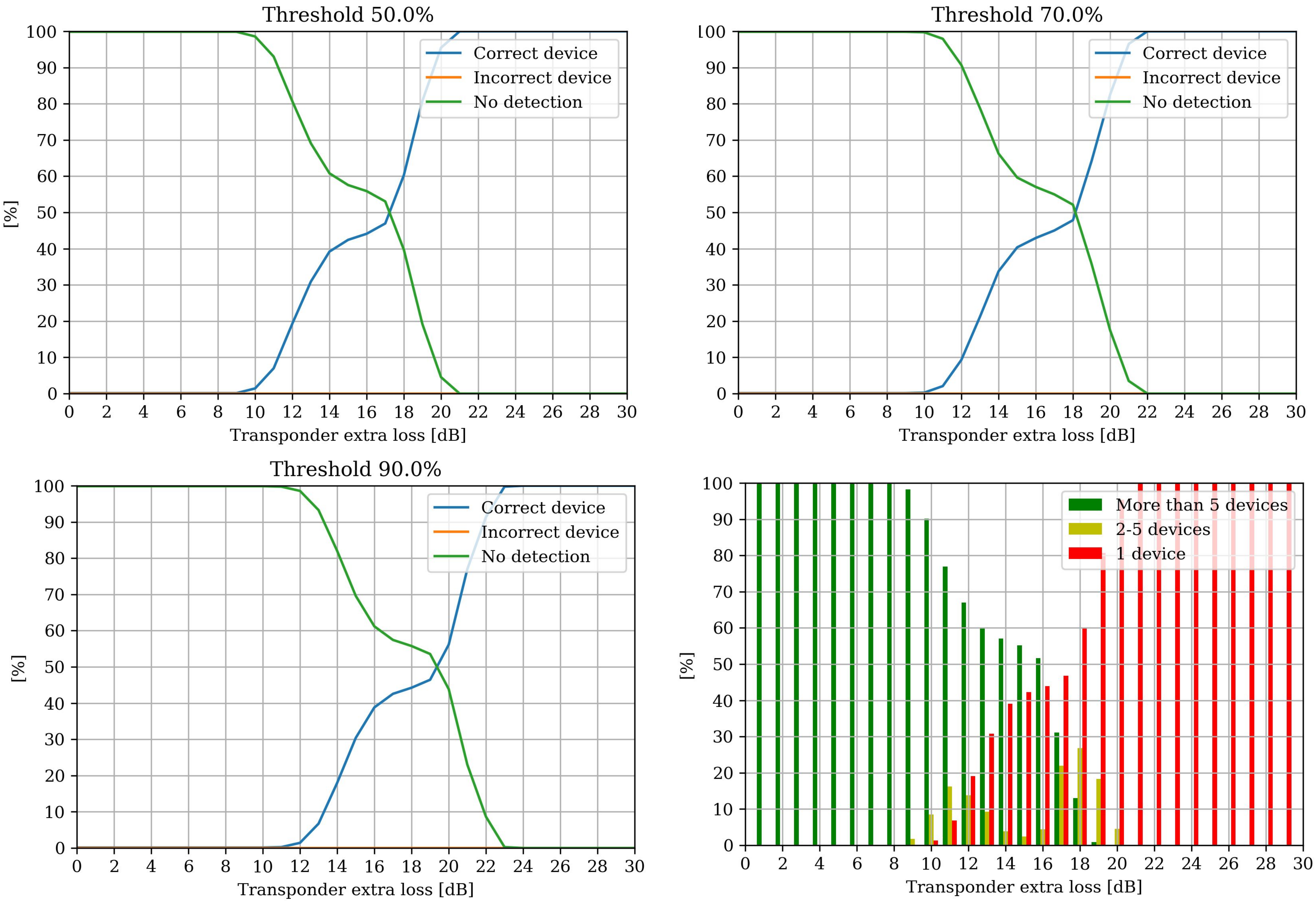

Threshold 50.0%
Correct device
Incorrect device
No detection
[%]
Amplifier gain [dB]


Threshold 70.0%
Correct device
Incorrect device
No detection
Amplifier gain [dB]

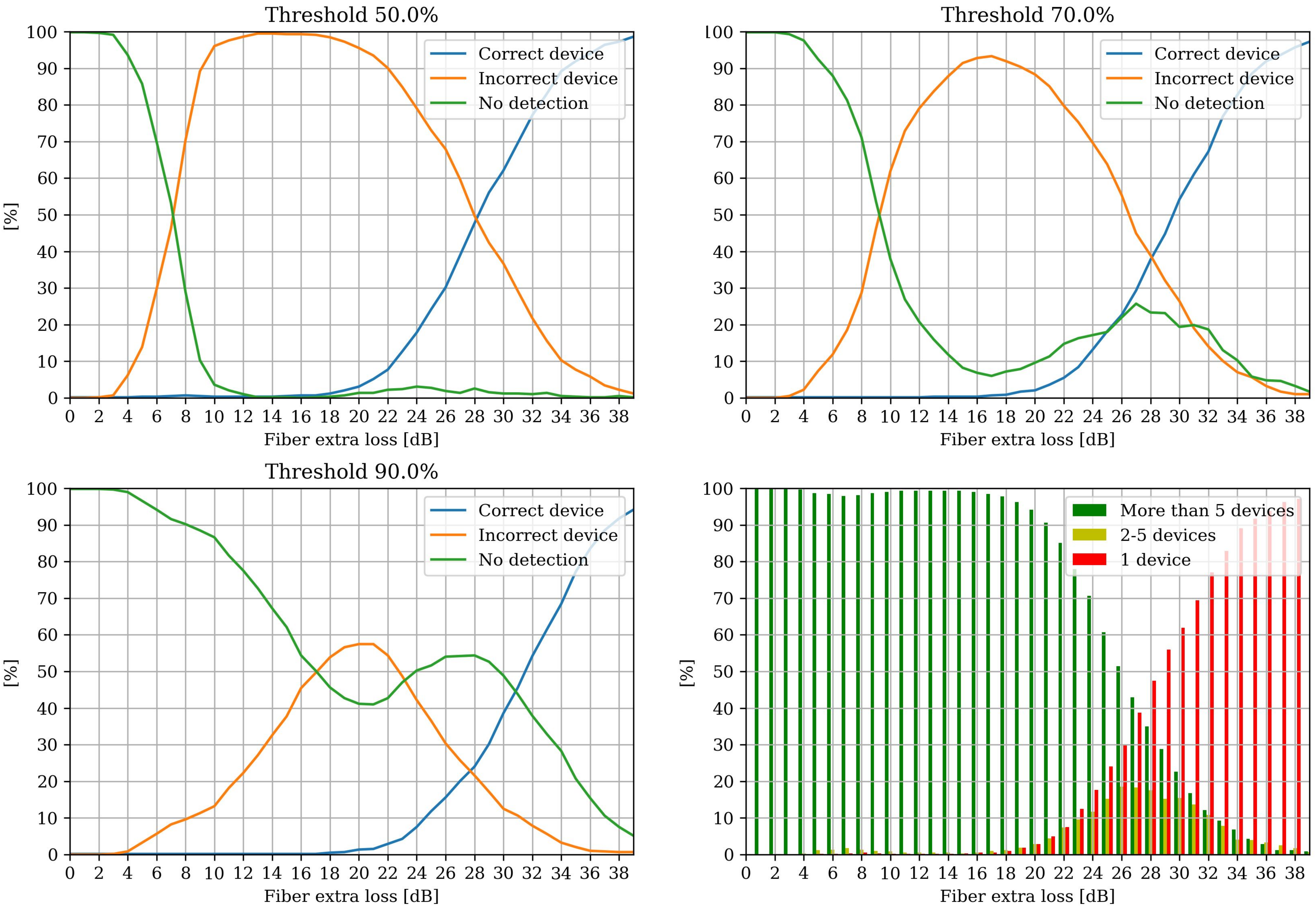

Threshold 90.0%
Correct device
Incorrect device
No detection
[%]
Amplifier gain [dB]


More than 5 devices
2-5 devices
1 device
[%]
Amplifier gain [dB]

ML module

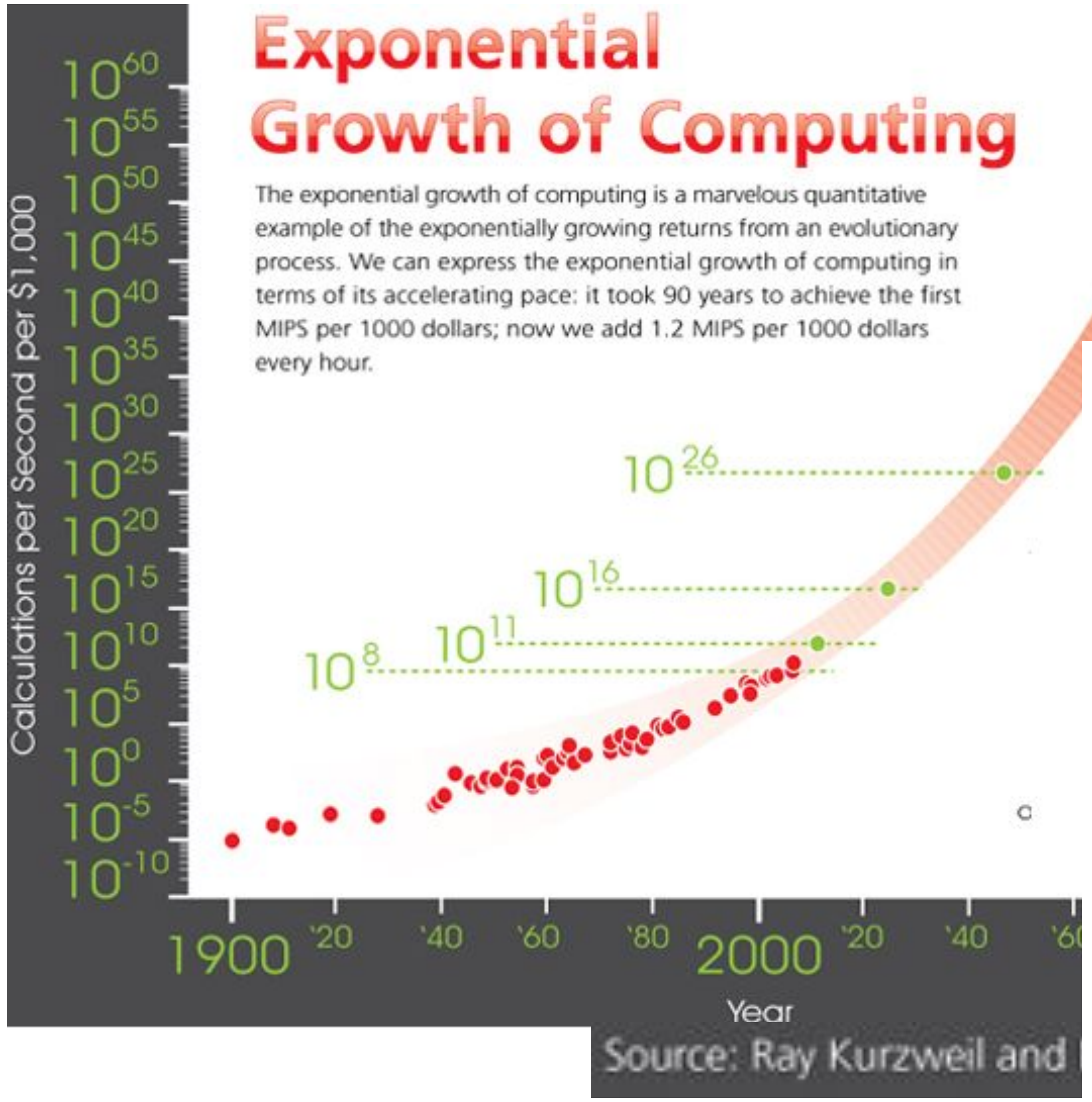

Threshold 50.0%
Threshold 70.0%
Threshold 90.0%
Correct device
Incorrect device
No detection
More than 5 devices
2-5 devices
1 device
[%]
Transponder extra loss [dB]

ML module

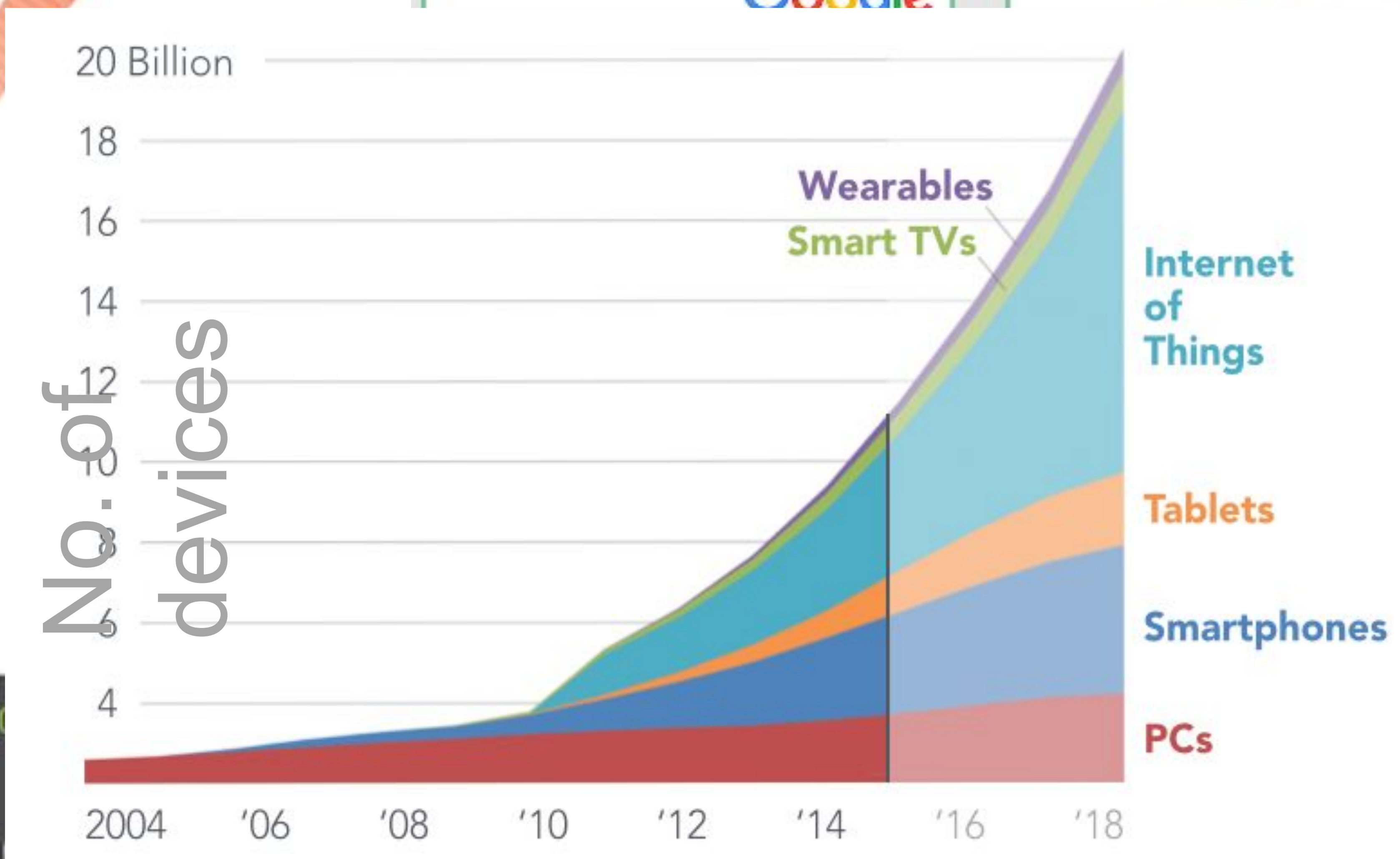

Threshold 50.0%
Threshold 70.0%
Threshold 90.0%
Correct device
Incorrect device
No detection
More than 5 devices
2-5 devices
1 device
[%]
Fiber extra loss [dB]

Why Data Analytics?

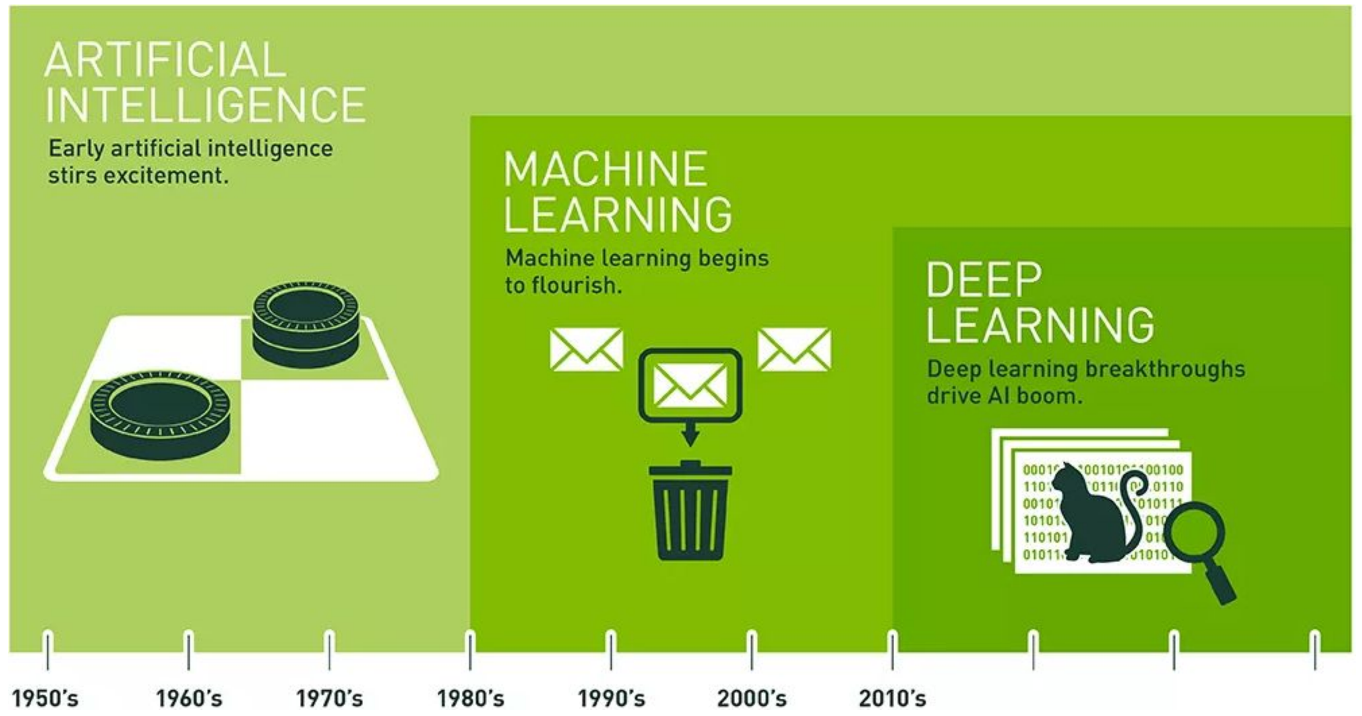
- Increasing computational power
- Increasing number of network devices for analysis, beyond human capabilities
- Growing progress in algorithms & theory developed by researchers
- Advanced telemetry capabilities (streaming)
- Operators look for capabilities beyond throughput (traffic prediction, failure management, advanced planning, performance monitoring, low margins and costs)



Caffe
NVIDIA Caffe
IBM Caffe
torch
TensorFlow Google
theano
Chainer



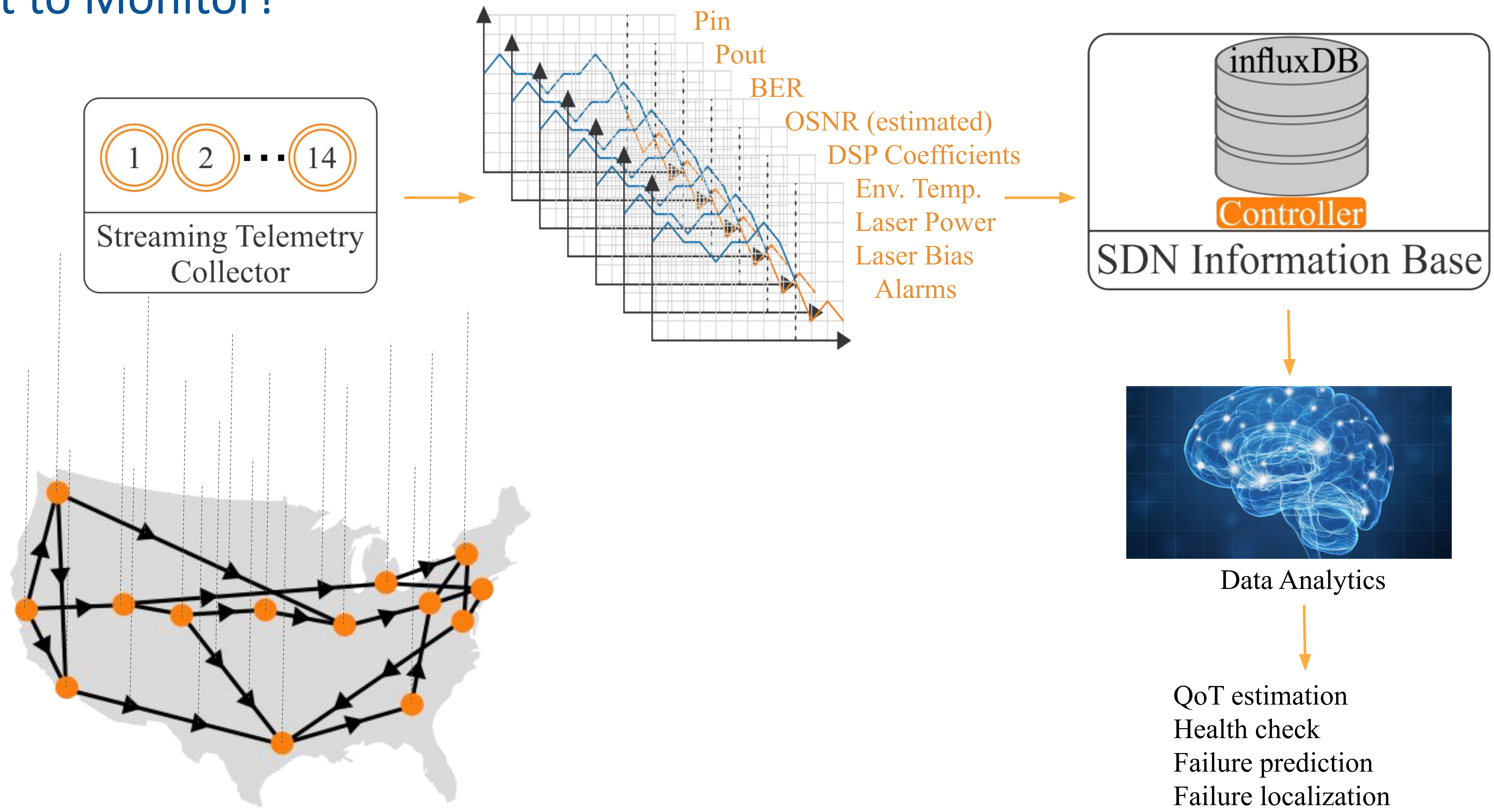
Source: Gartner, IDC, Strategy Analytics, Machina Research, company filings, BII estimates (http://forecastjoy.com/wp-content/uploads/2014/03/deviceforecast.png)



Alan Lau
OFC'2020

UNICAMP

What to Monitor?



Data Analytics Applications

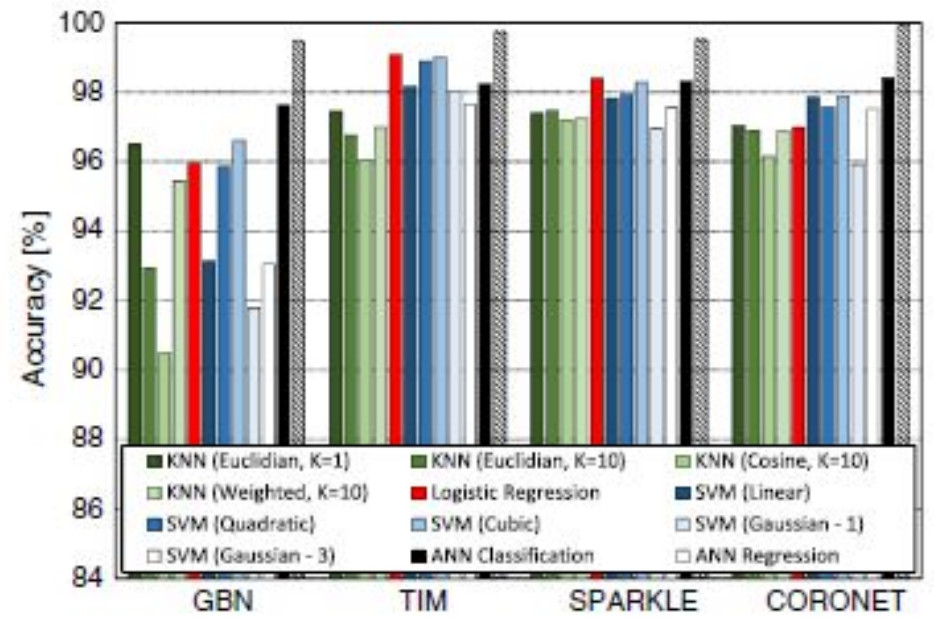
QoT Estimation

- Network planning is becoming increasingly complex because of the multitude of operation parameters
  - In network planning, which is the QoT of an unestablished lightpath?

Network Planning

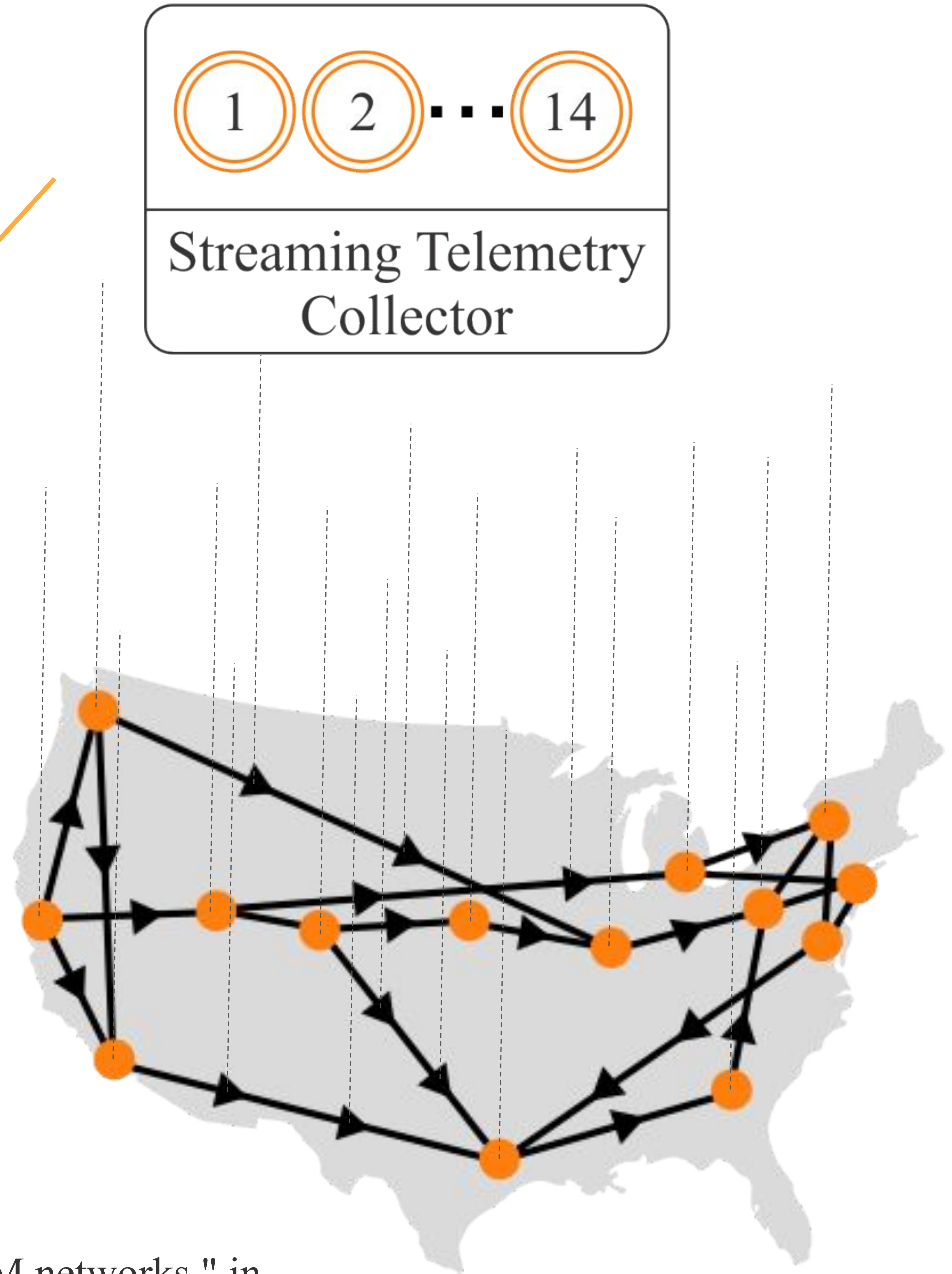
Number of hops
Number of spans
Total length
Average link length
Maximum link length
Average span attenuation
Average dispersion
Modulation format

Feasible/Unfeasible
Margin [dB]



Data Analytics

Network Operation



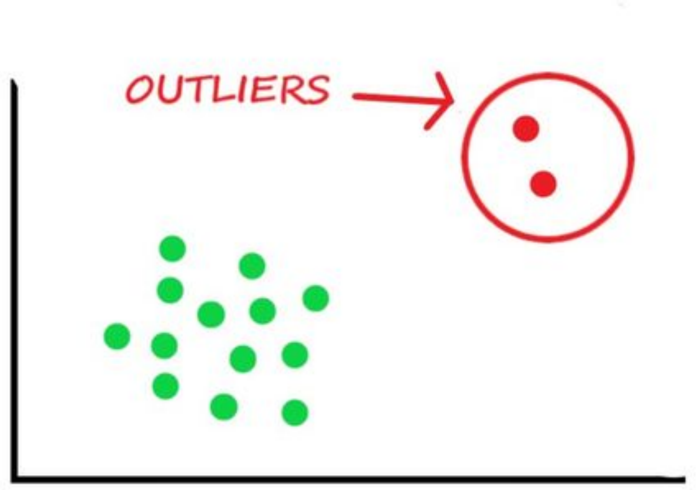
R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D84-D99, Oct. 2018.

UNICAMP

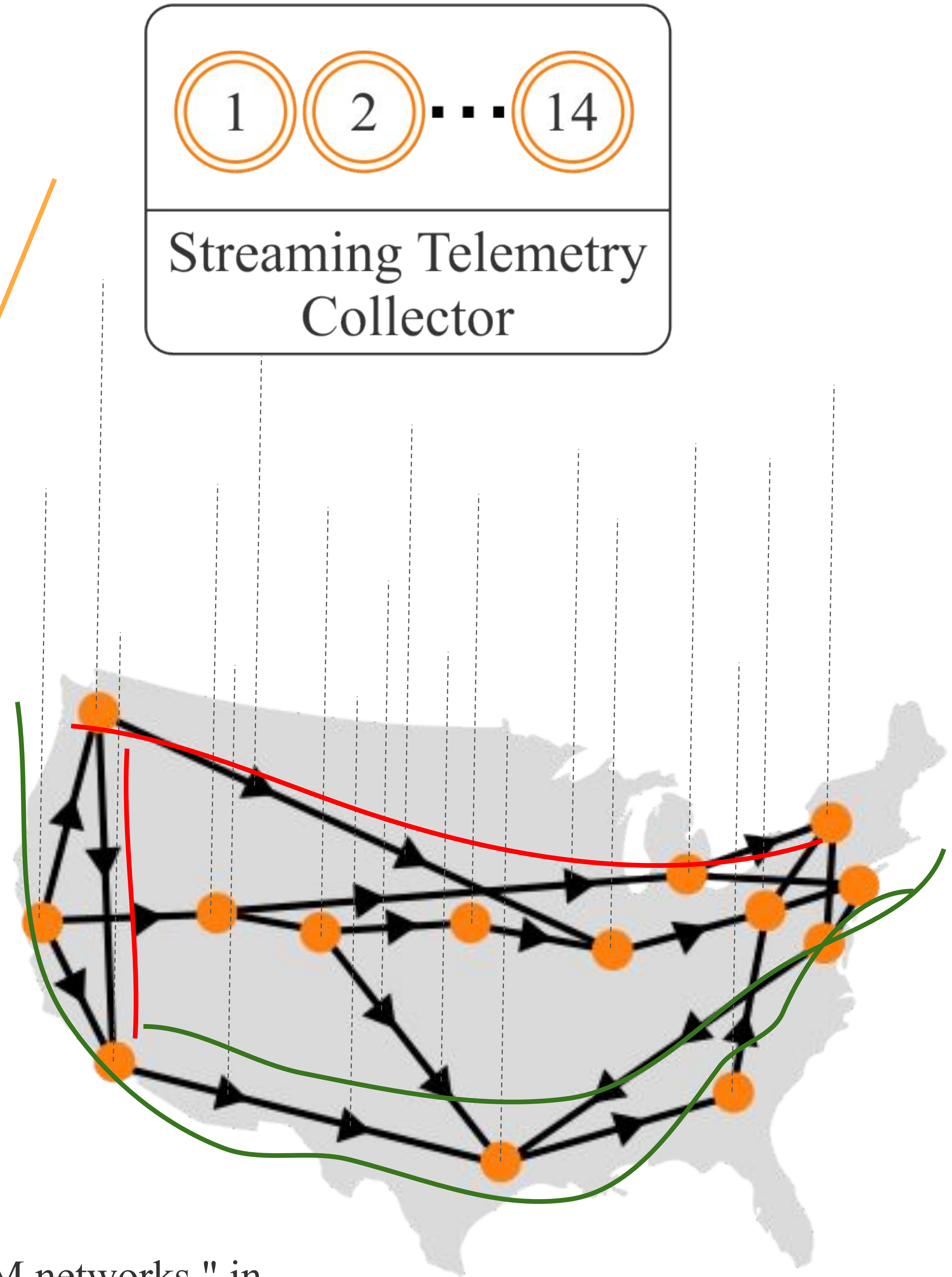
Data Analytics Applications

Health Check

- Identification of anomalous lightpaths



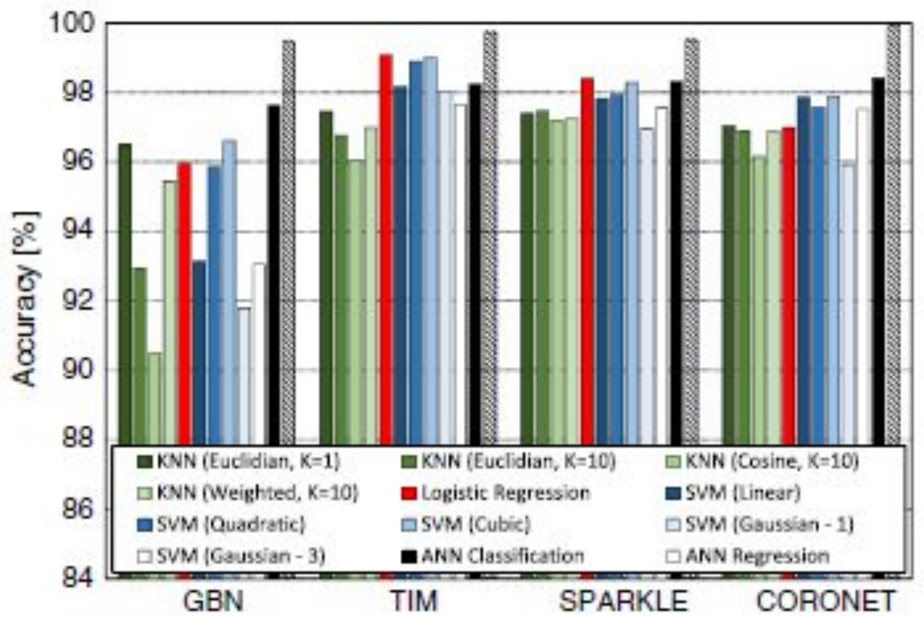
Network Operation



Network Planning

Number of hops
Number of spans
Total length
Average link length
Maximum link length
Average span attenuation
Average dispersion
Modulation format

Data Analytics



Feasible/Unfeasible
Margin [dB]

R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D84-D99, Oct. 2018.

UNICAMP

Data Analytics Applications

Failure prediction

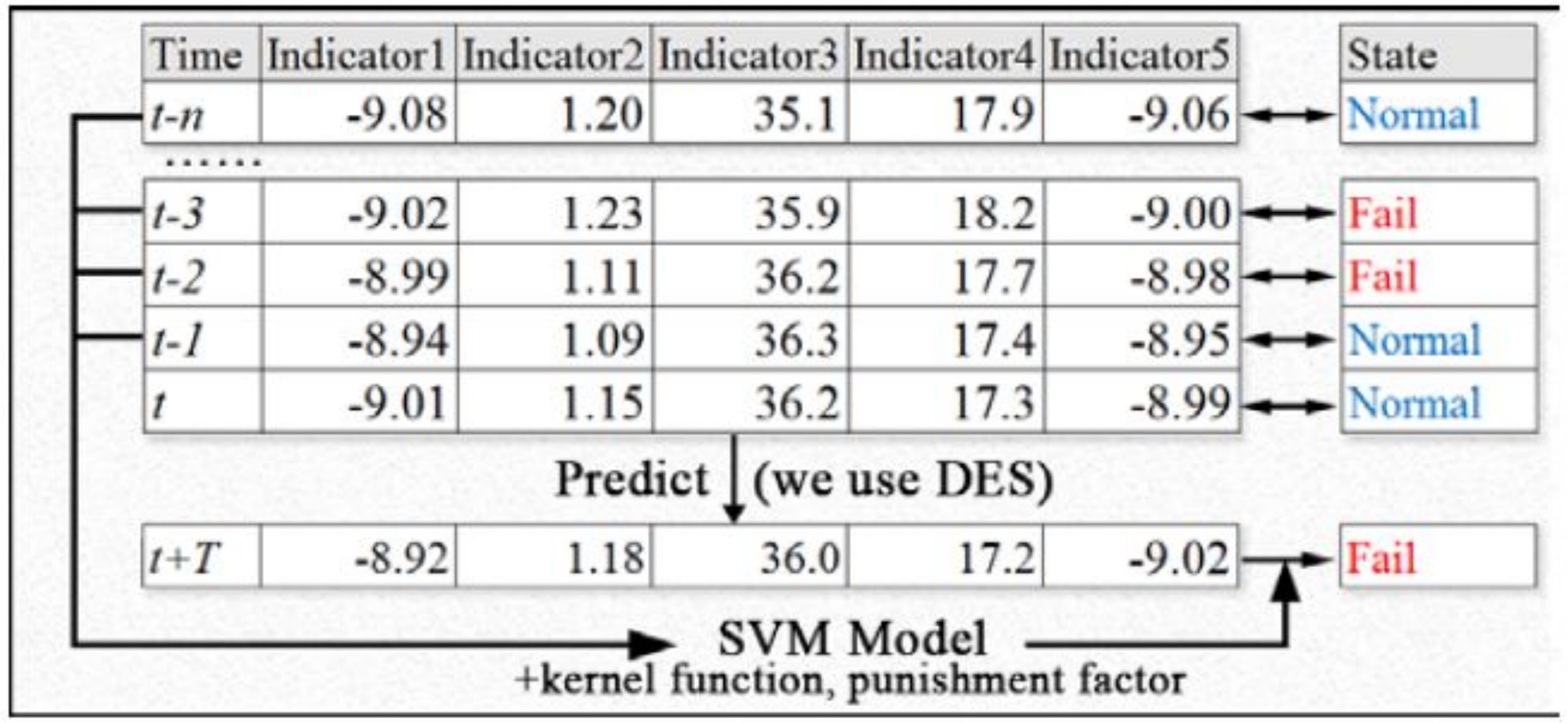
Indicator Name	Units
Input Optical Power (IOP)	dBm
Laser Bias Current (LBC)	mA
Laser Temperature Offset (LTO)	°C
Output Optical Power (OOP)	dBm
Environmental Temperature (ET)	°C
Unusable Time	s

Research Article
Vol. 25, No. 16 | 7 Aug 2017 | OPTICS EXPRESS 18553
Optics EXPRESS

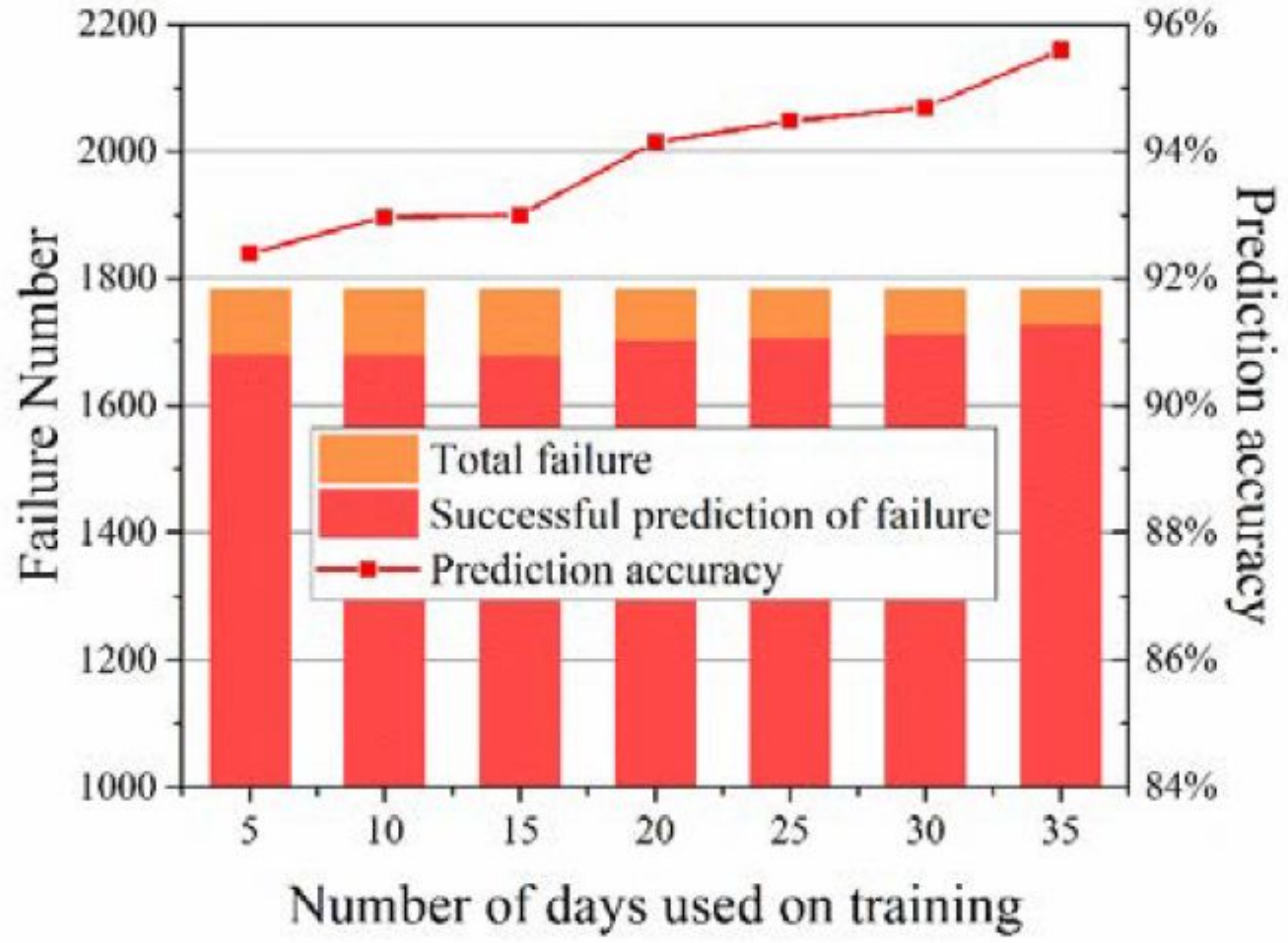
Failure prediction using machine learning and time series in optical network

ZHILONG WANG,[1] MIN ZHANG,[1,*] DANSHI WANG,[1] CHUANG SONG,[1] MIN LIU,[1] JIN LI,[1] LIQI LOU,[1] AND ZHUO LIU[2]

[1]*State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*China Mobile Communications Corporation, Beijing 100033, China*
[*]*mzhang@bupt.edu.cn*



- Monitors boards of China Telecom to predict failures in the next day
- Uses SVM+Time series processing
- Identifies the parameters (laser bias, laser temperature, environmental temperature) most related to the failures
- Achieves high prediction levels



UNICAMP

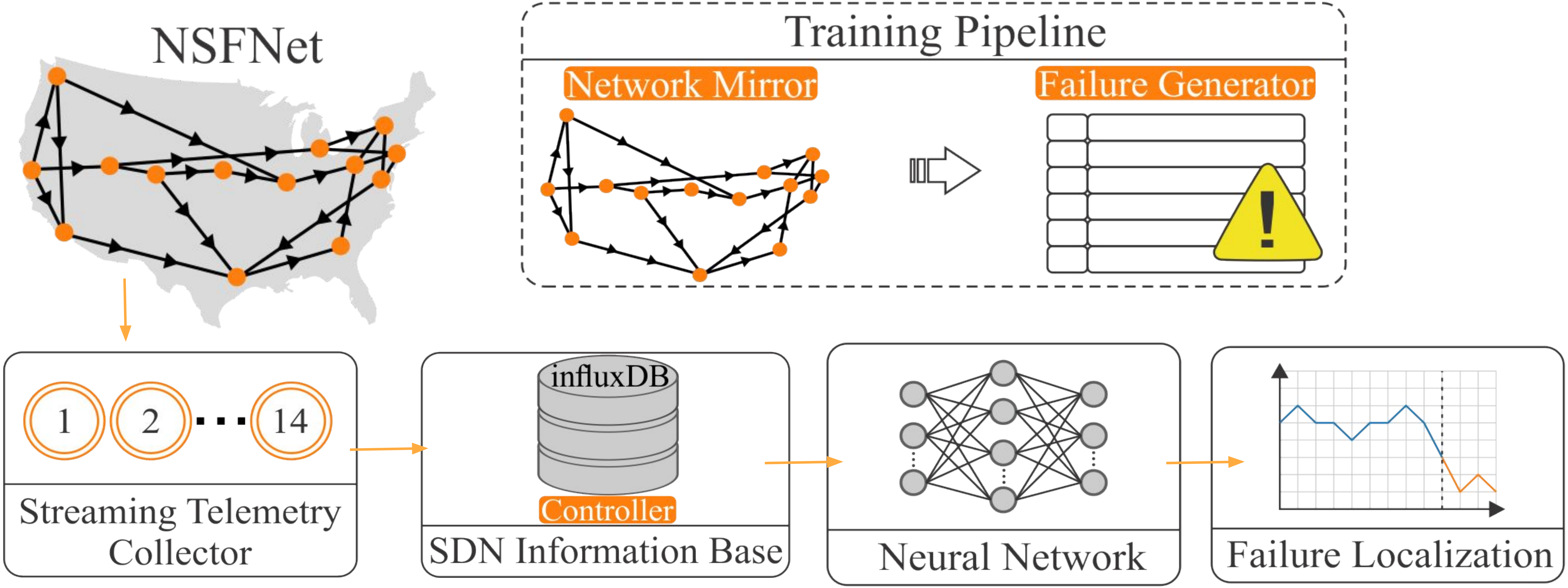
Soft Failure Localization

Failure localization

Soft Failure Localization Using Machine Learning with SDN-based Network-wide Telemetry

Kayol S. Mayer[1,+], Jonathan A. Soares[1], Rossano P. Pinto[2], Christian E. Rothenberg[2], Dalton S. Arantes[1], and Darli A. A. Mello[1]

Accepted at ECOC 2020!



Soft Failure Localization

Case study

2136 Components:

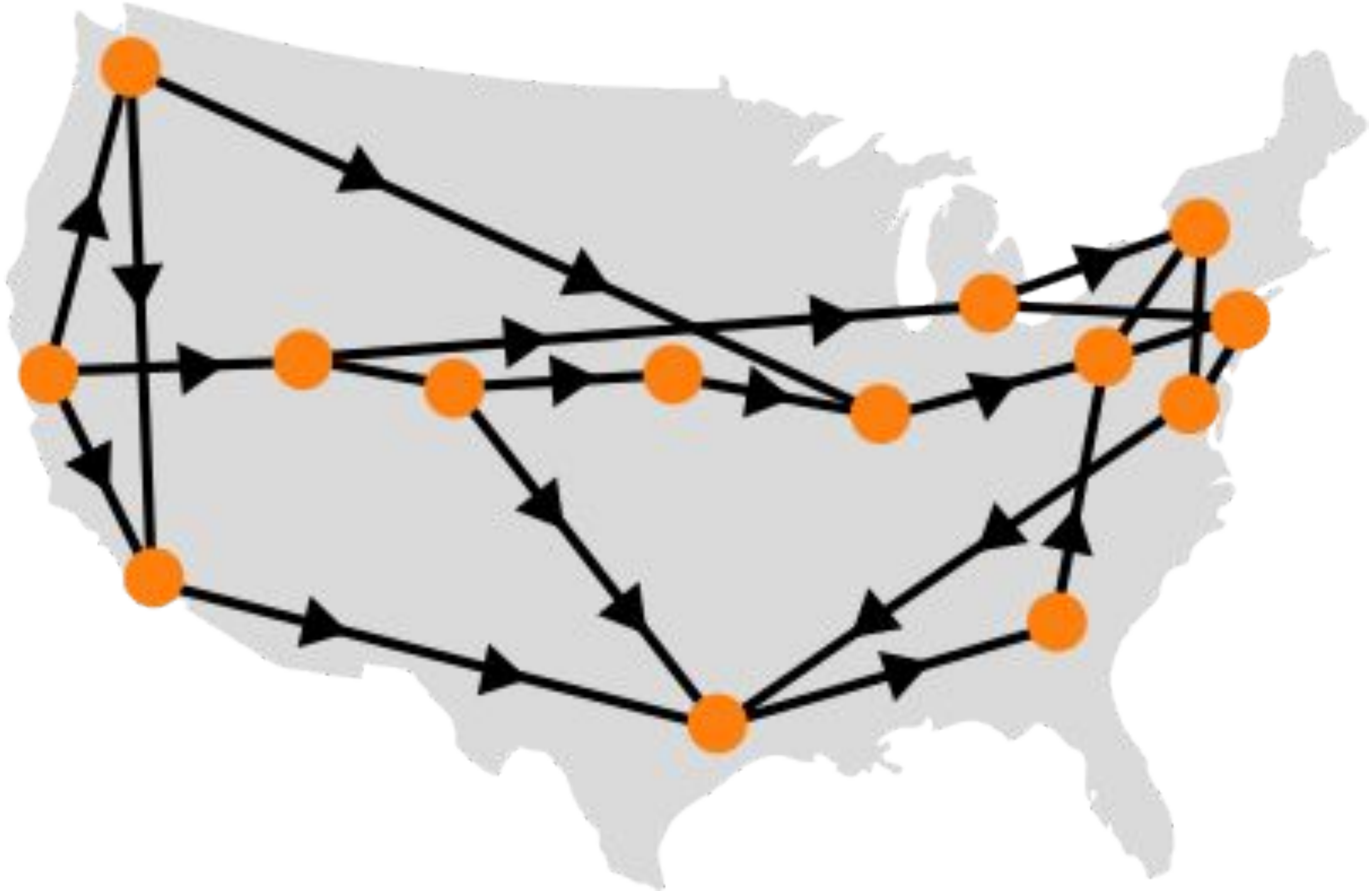
- 580 Unidirectional fiber spam;
- 624 Amplifiers;
- 42 OCMs;
- 72 WSSs;
- 44 Splitters;
- 772 Transceivers;

Requested services:

- 1000 demands;
- 386 accepted;

Monitored Parameters

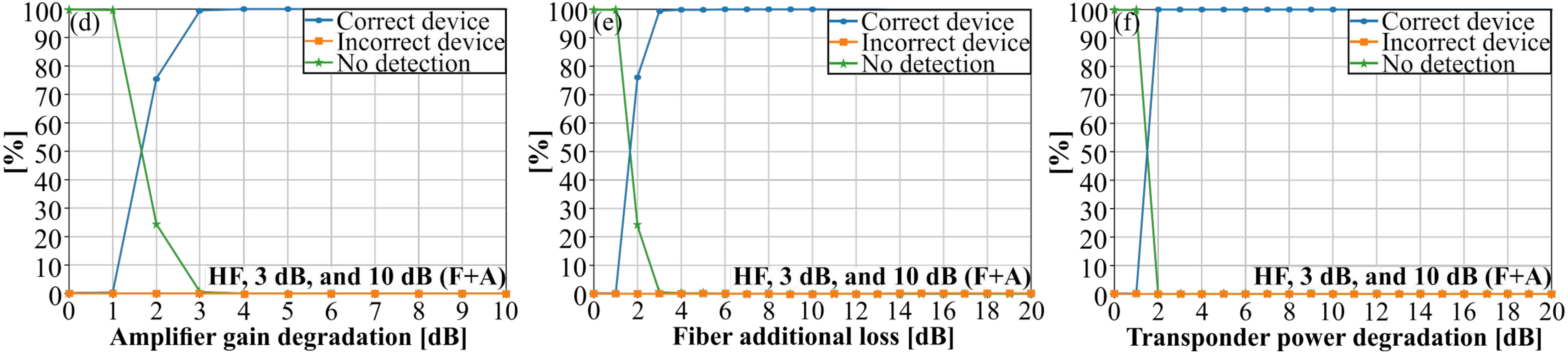
Network element	# Monitored parameters per card	# of Cards	# Monitored parameters
Transceivers	3 (Pin, Pout, OSNR)	772	2316
Amplifiers	2 (Pin, Pout)	624	1248
			3564



Links

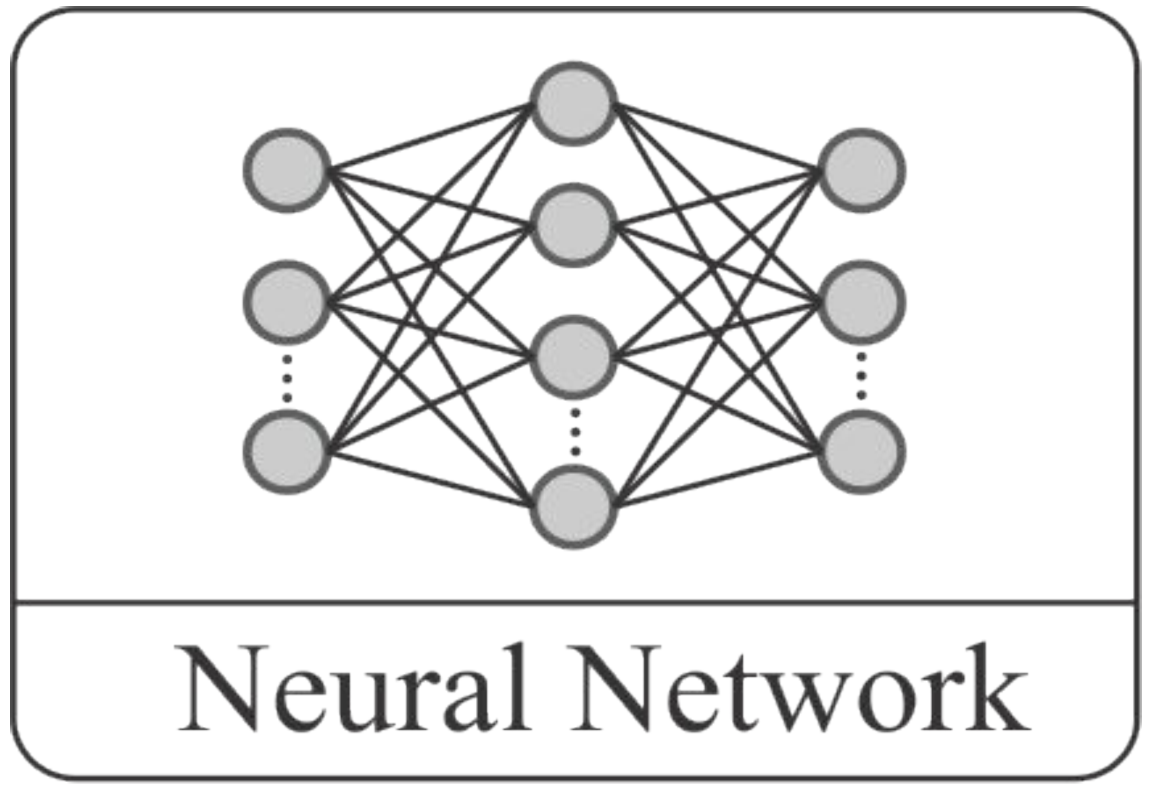
Source	Destination	Distance (km)
Seattle (WA)	Palo Alto (CA)	1100
Seattle (WA)	San Diego (CA)	1600
Seattle (WA)	Champaing (IL)	2800
Palo Alto (CA)	Salt Lake City (UT)	1000
Palo Alto (CA)	San Diego (CA)	600
San Diego (CA)	Houston (TX)	2000
Salt Lake City (UT)	Boulder (CO)	600
Salt Lake City (UT)	Ann Arbour (MI)	2400
Boulder (CO)	Lincoln (NE)	800
Boulder (CO)	Houston (TX)	1100
Lincoln (NE)	Champaing (IL)	700
Champaing (IL)	Pittsburg (PA)	700
Houston (TX)	Atlanta (GA)	1200
Houston (TX)	College Pk (MD)	2000
Atlanta (GA)	Pittsburg (PA)	900
Pittsburg (PA)	Ithaca (NY)	500
Pittsburg (PA)	Princeton (NJ)	500
College Pk (MD)	Ithaca (NY)	500
College Pk (MD)	Princeton (NJ)	300
Ann Arbour (MI)	Ithaca (NY)	800
Ann Arbour (MI)	Princeton (NJ)	1000

Failure Localization Results



UNICAMP

Failure Localization Results



Dataset generation: 80 min
ANN training time: 6 min
ANN Failure localization: <1s