Artificial Intelligence for the control and orchestration of mobile networks

Albert Banchs

Developing the Science of Networks

Network control and orchestration

- **Network management/orchestration** is becoming more and more complex
- **Adapting** to dynamics of a tangled environment ⇒ **Anticipatory actions**
  - **Distributed** infrastructure, **heterogeneity**, new **paradigms and use cases…**
- **But, currently**
  - Network management is made by **human**, thus, optimize **generic**, **non-flexible**, and **manually designed** objectives which will render the **promised** goals **impossible**
- **Solution?**
  - Towards **zero-touch** approaches → **Artificial Intelligence**

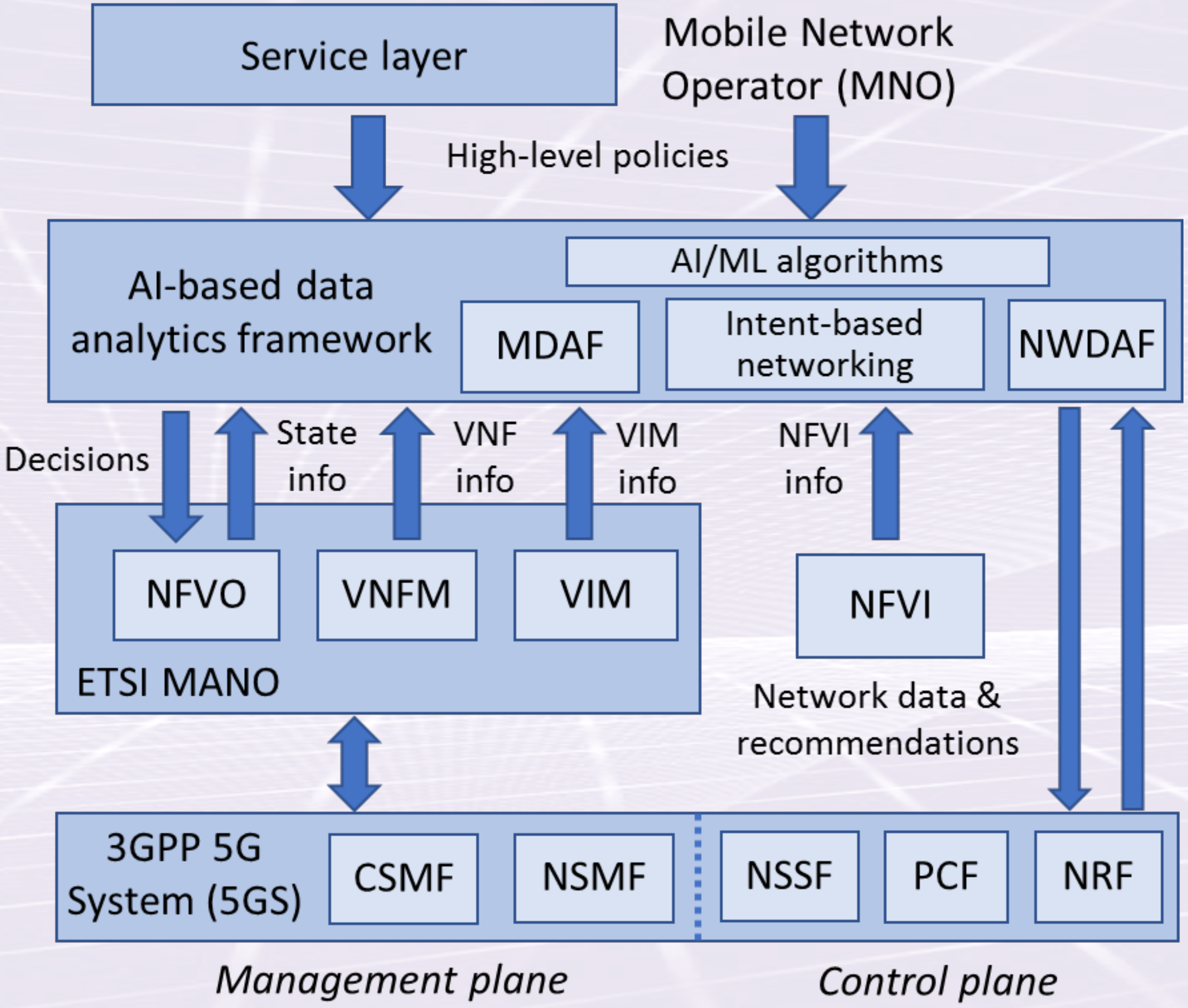
Outline

- Data Analytics and AI framework
- Analysis of benefits of Dynamic orchestration
- Realizing Dynamic orchestration with machine learning
- Combining m achine learning with Intent-based Networking

Data analytics and Artificial Intelligence for Orchestration

- Artificial Intelligence is a natural choice for driving orchestration decision
  - We need to make predictions, classifications and decisions based on data
- 3GPP has identified this and is pursuing efforts towards defining an AI-based Data Analytics
  - Autonomous and efficient control, management and orchestration
- Modules defined by 3GPP to this end
  - Network Data Analytics Function (NWDAF)
  - Management Data Analytics Function (MDAF)

AI-based data Analytics framework



Data analytics for the control plane

- In the control plane, analytics allow NFs to optimize their behavior at run-time, typically at a much faster speed than what network management and orchestration systems allow
- NWDAF analytics can be leveraged to improve
  - Slice-level load balancing
  - Service experience and Quality of Experience (QoE)
- Examples of data analytics usage
  - NSSF: Selecting the set of Network Slice instances serving a UE
  - PCF: Unified policy framework to govern network behavior, including the QoS parameters
  - NRF: Selection of a NF instance when a certain NF type is needed

Data analytics for the management plane

- Data used as input by the AI-based analytics framework
  - NFV Infrastructure (NFVI): knowledge on the computational resources' capabilities (such as the type of CPU and memory, accelerators, etc.) along with their availability (i.e., the status and utilization level)
  - MANO system: requirements of the network slices
- Decision taken
  - NFVO: NF placement and resource allocation decisions while ensuring that the resulting resource allocation satisfies the respective slice SLA
  - VNFM: Run-time up and down scaling of resources
  - CSMF (Communication Service Management Function) and NSMF (Network Slice Management Function (NSMF): Admission control of new slices

Artificial intelligence & data analytics

- AI is a computation paradigm that endows machines with intelligence
  - Aiming to teach them how to work, react, and learn like humans
  - Many techniques fall under this broad umbrella
- Machine learning enables the artificial processes to absorb knowledge from data and make decisions without being explicitly programmed
  - Data needs to be collected and made availably to AI algorithms
  - Machine learning is closely related to data analytics
- Machine learning has become very popular driven by:
  - Modern challenges are "high-dimensional" in nature
  - We have rich data sources and processing power that can be used to solve problems
  - Machine learning can be integrated into working software to support products demanded by industry
- In line with the rising popularity of machine learning, this tool is being widely used for many networking problems including 5G

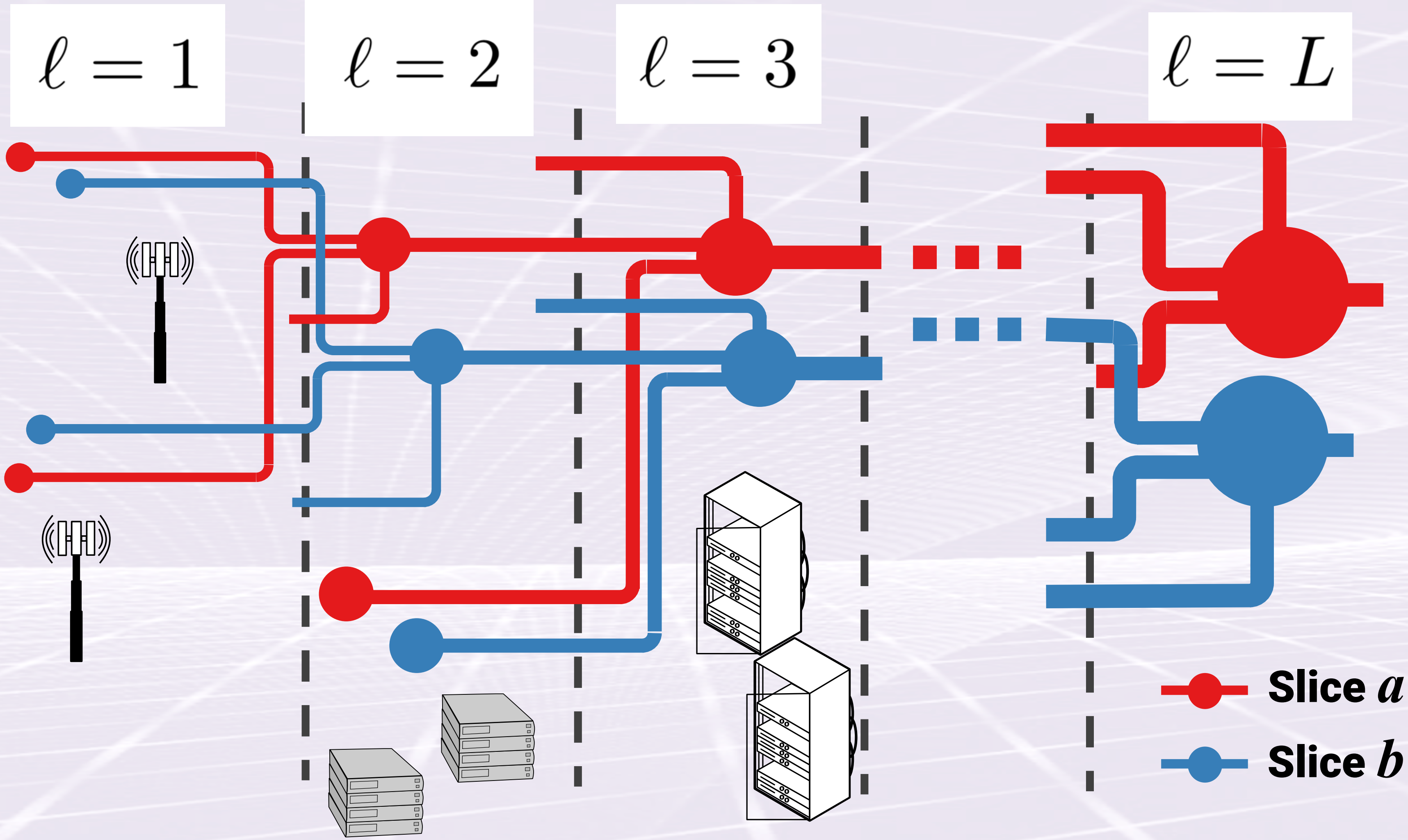
Outline

- Data Analytics and AI framework
- Analysis of benefits of Dynamic orchestration
- Realizing Dynamic orchestration with machine learning
- Combining m achine learning with Intent-based Networking

Empirical evaluation of network slicing efficiency

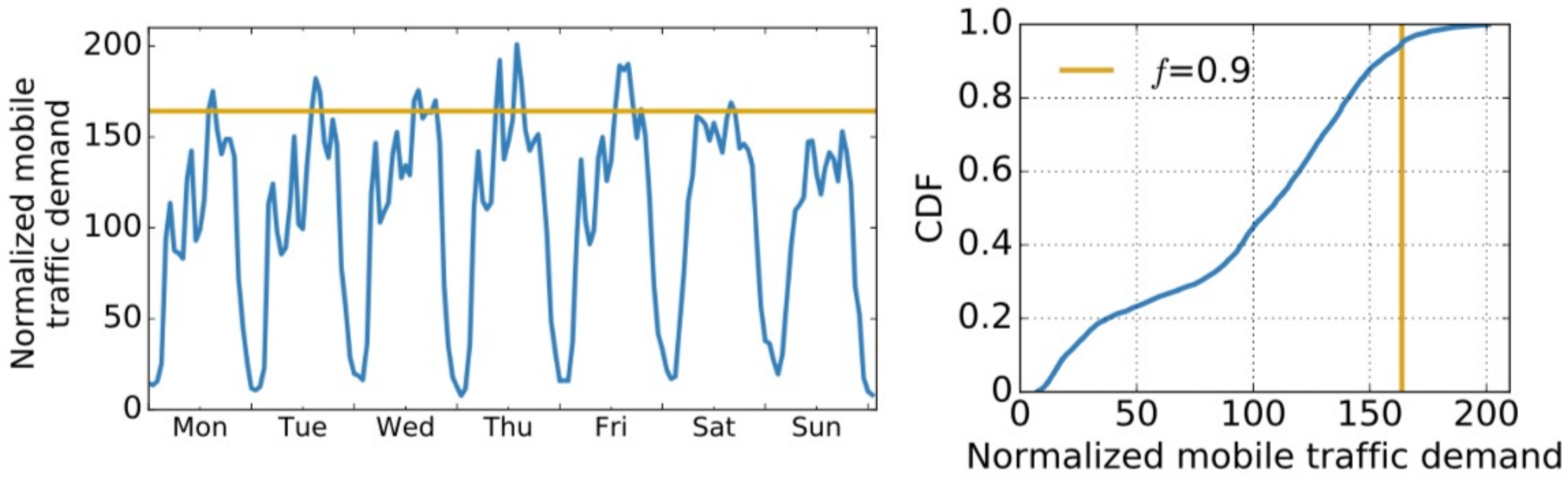
- Following a data driven approach we want to
  - Quantify the price paid in efficiency when suitable algorithms for dynamic resource allocation are not available, and the operator has to resort to physical network duplication
  - Evaluate the impact of sharing resources at different levels of the network, including the cloudified core, the virtualized radio access, or the individual antennas
  - Outline the benefit of dynamic resource allocation at different timescales under various slice specifications
- Methodology
  - Our approach can be used for generic kinds of resource allocation
  - Still, it is not an optimization, but rather an indication of how well slices will behave

Network level & Aggregation



Meeting slice requirements

f 90% w 1 hour



Efficiency evaluation

We evaluate the efficiency of a multi-slice scenarios by comparing

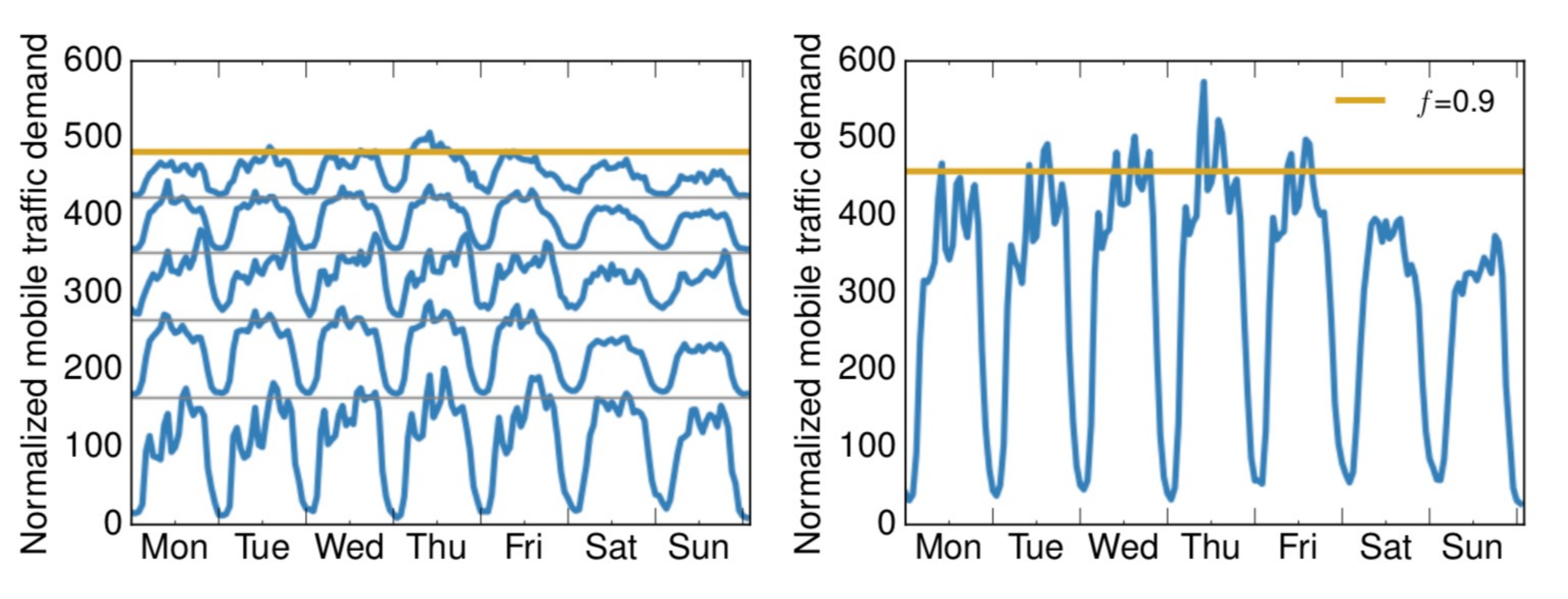
- A sliced scenario in which we need to statically provision each slice with the necessary resources to meet the slice requirements

$$\mathbb{R}^{\mathbb{Z}}_{\ell,\tau} = \sum_{s \in \mathcal{S}} \sum_{c \in \mathcal{C}_\ell} \sum_{n \in \mathcal{T}} \tau \cdot \hat{r}^{\mathbb{Z}}_{c,s}(n).$$

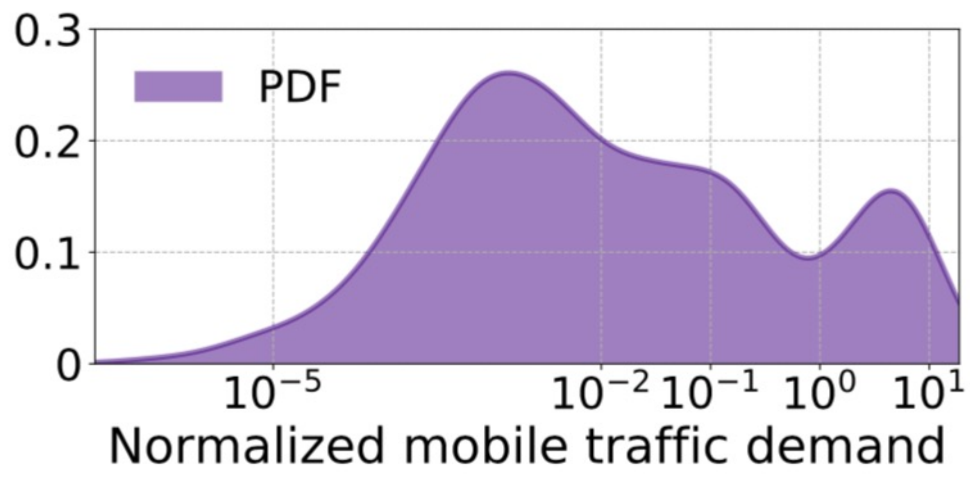
- A perfect slicing scenario, in which the exact amount of resources are shared instantaneously among all slides

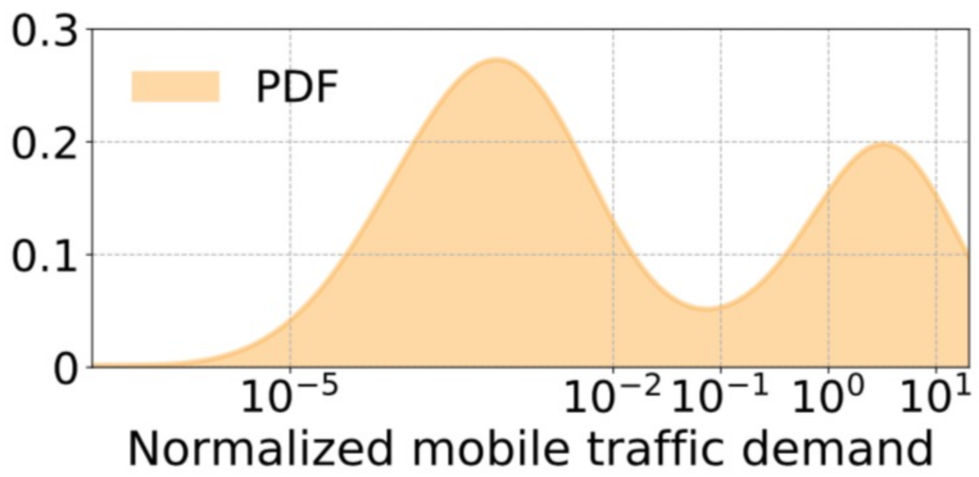
$$\mathbb{P}^{\mathbb{Z}}_{\ell,\tau} = \sum_{c \in \mathcal{C}_\ell} \sum_{n \in \mathcal{T}} \tau \cdot \hat{r}^{\mathbb{Z}}_{c}(n),$$

Efficiency example

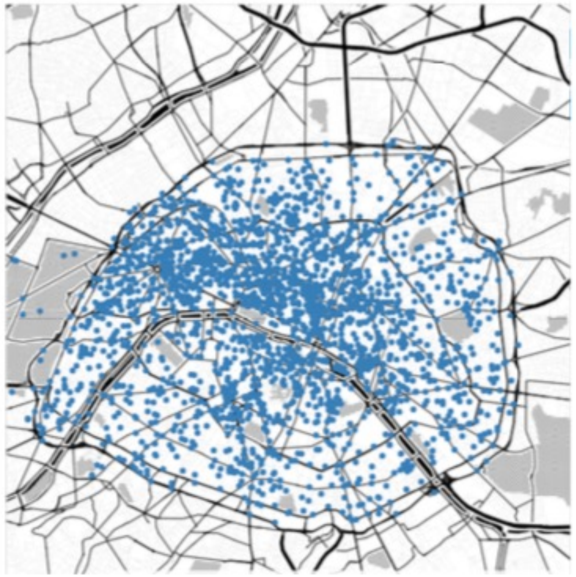


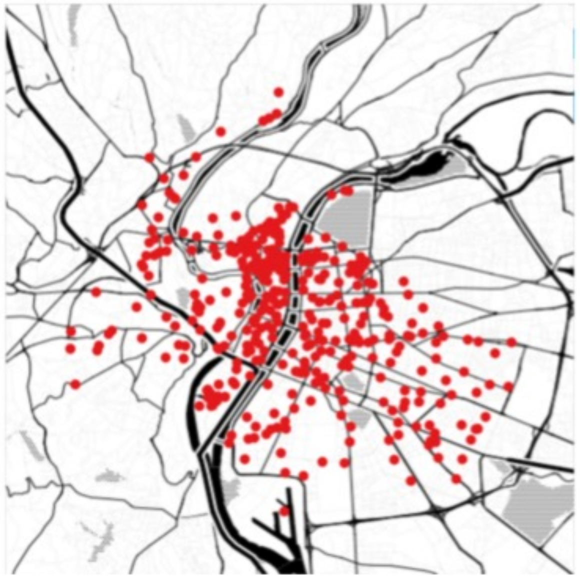
Empirical evaluation





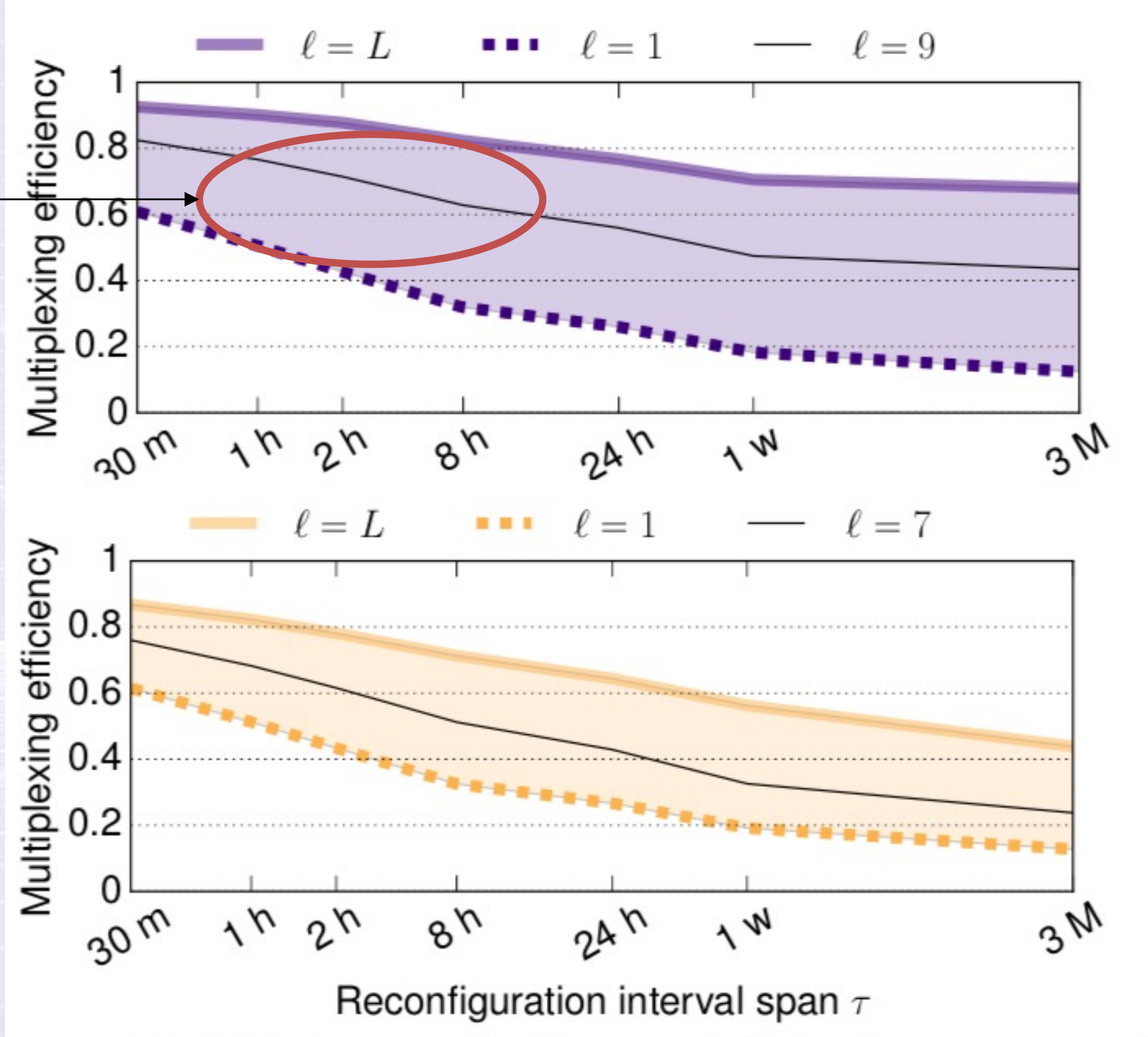
Two large cities
Three months of data
Granularity in space: sector
Granularity in time: 5 minutes
38 services in total





Results

Reconfiguration sweetspot is here



Either we allow such timescale, otherwise we don't have much gain over static

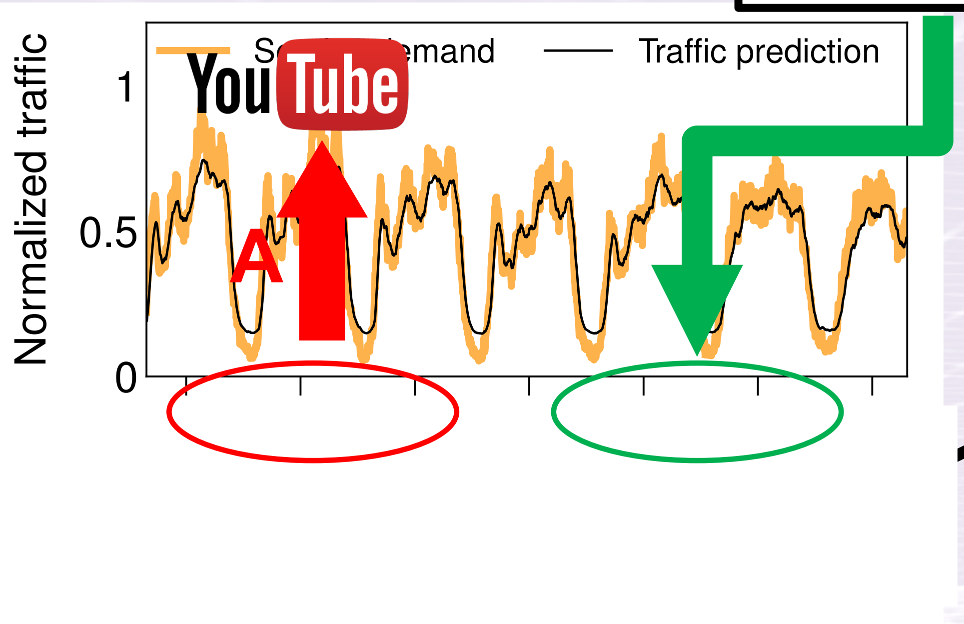
Outline

- Data Analytics and AI framework
- Analysis of benefits of Dynamic orchestration
- <u>Realizing Dynamic orchestration with machine learning</u>
- Combining m achine learning with Intent-based Networking

Capacity vs Demand forecasting

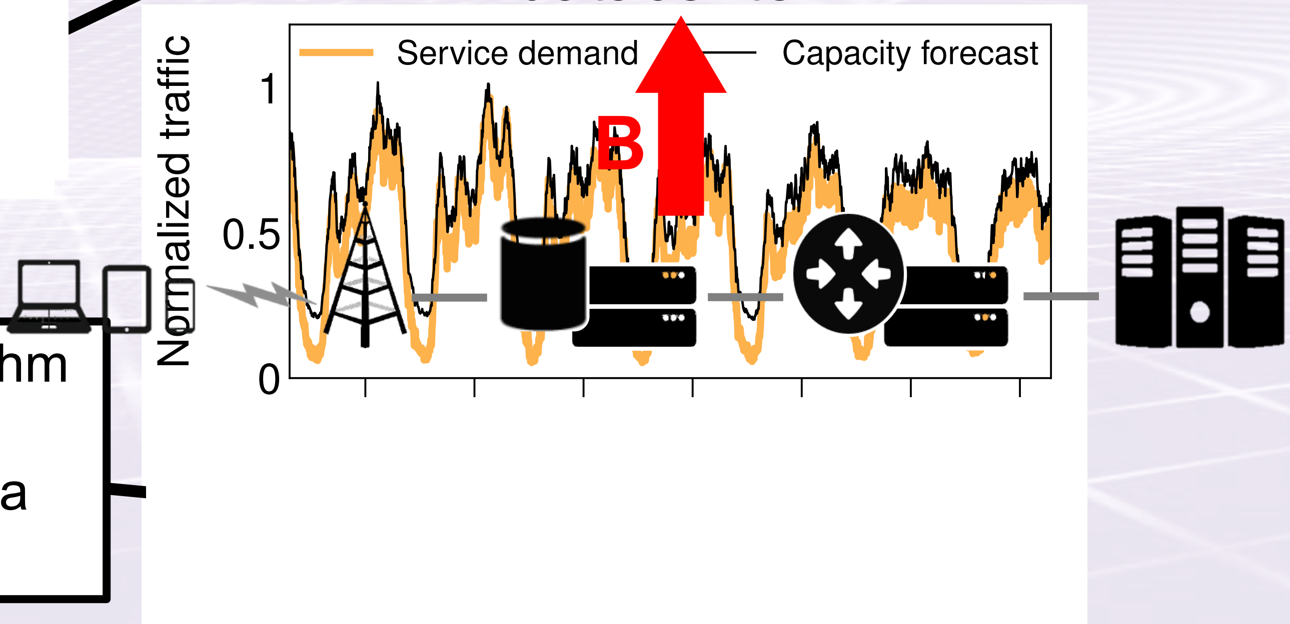
- Traditional approaches dea … orecasting

DeepCog



A traffic demand forecasting algorithm aims to minimize the error wrt to the original data, so underestimation is possible

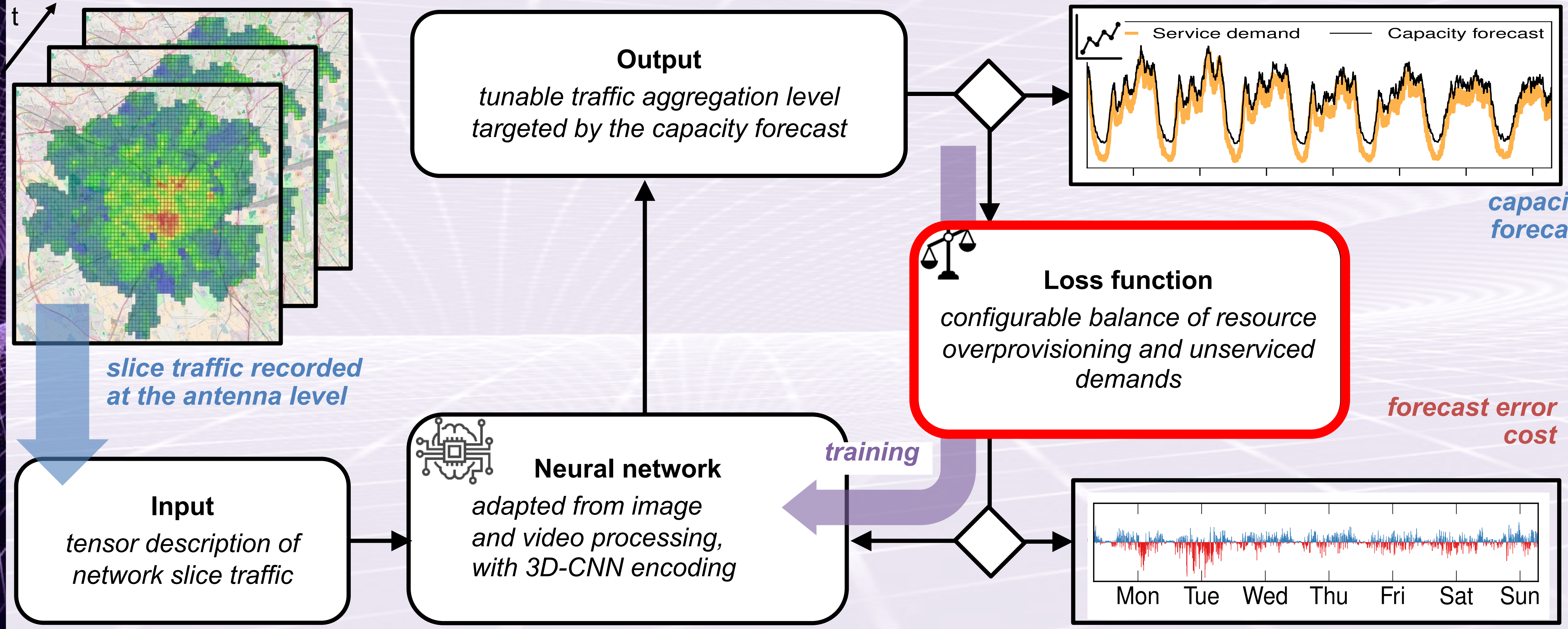
cloud RAN datacenter



A capacity forecasting algorithm minimizes the amount of resources needed to serve a given demand

DeepCog

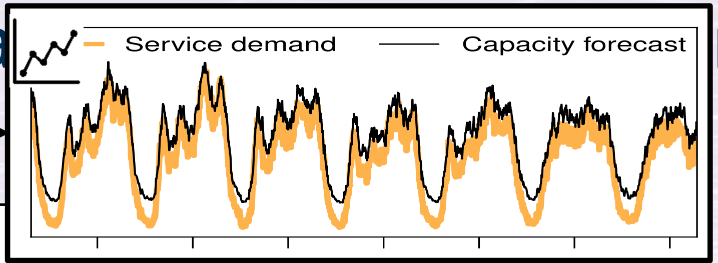
- DeepCog's design follows a deep learning approach



DeepCog

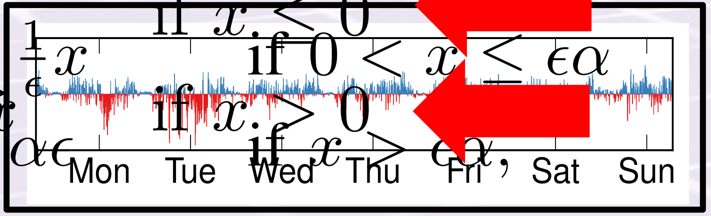
Loss function

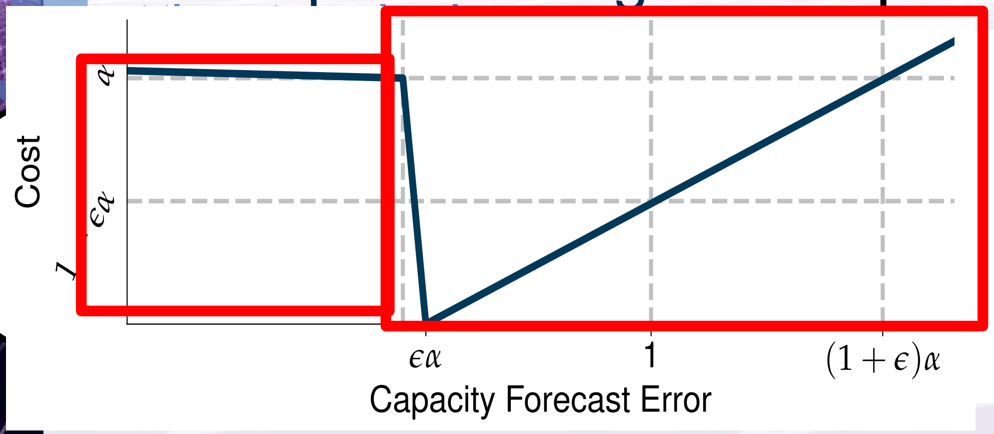
- DeepCog's design follows a deep learning approach
- Determines the penalty incurred when ma… r
- Tailored to capacity forecast problem
- It accounts for the costs resulting from
  - SLA violations
  - Overprovisioning



Unique system parameter

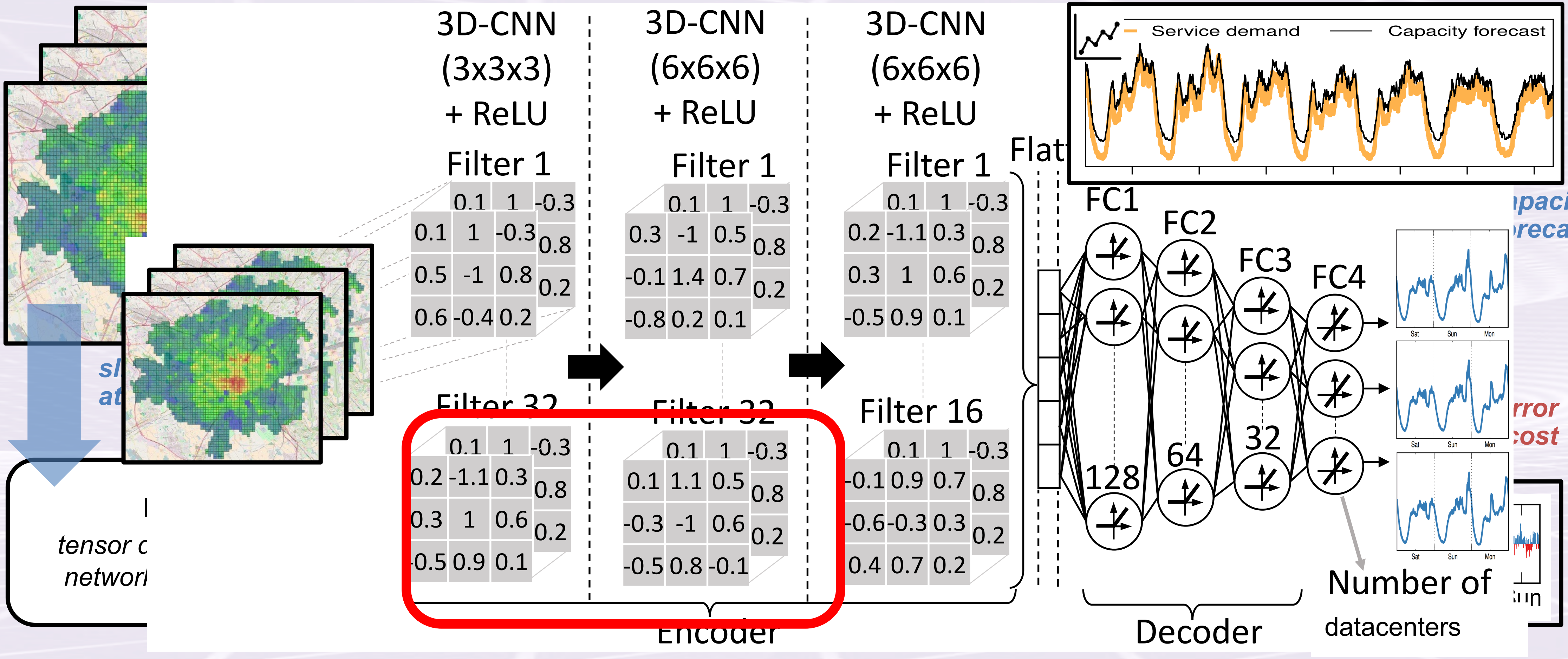
$$\ell(x) = \begin{cases} \alpha - \epsilon \cdot x & \text{if } x \le 0 \\ \alpha - \frac{1}{\epsilon} x & \text{if } 0 < x \le \epsilon\alpha \\ x - \alpha\epsilon & \text{if } x > \epsilon\alpha, \end{cases}$$





DeepCog

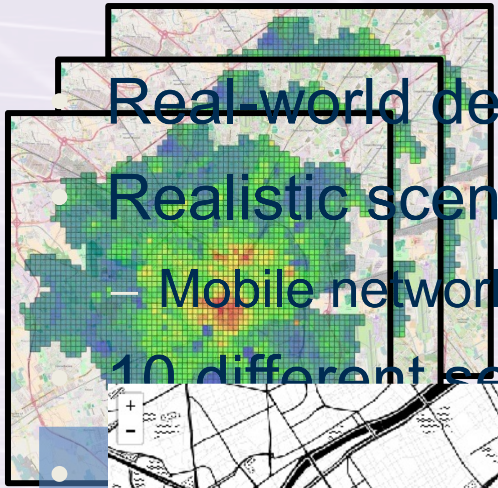
DeepCog’s settings

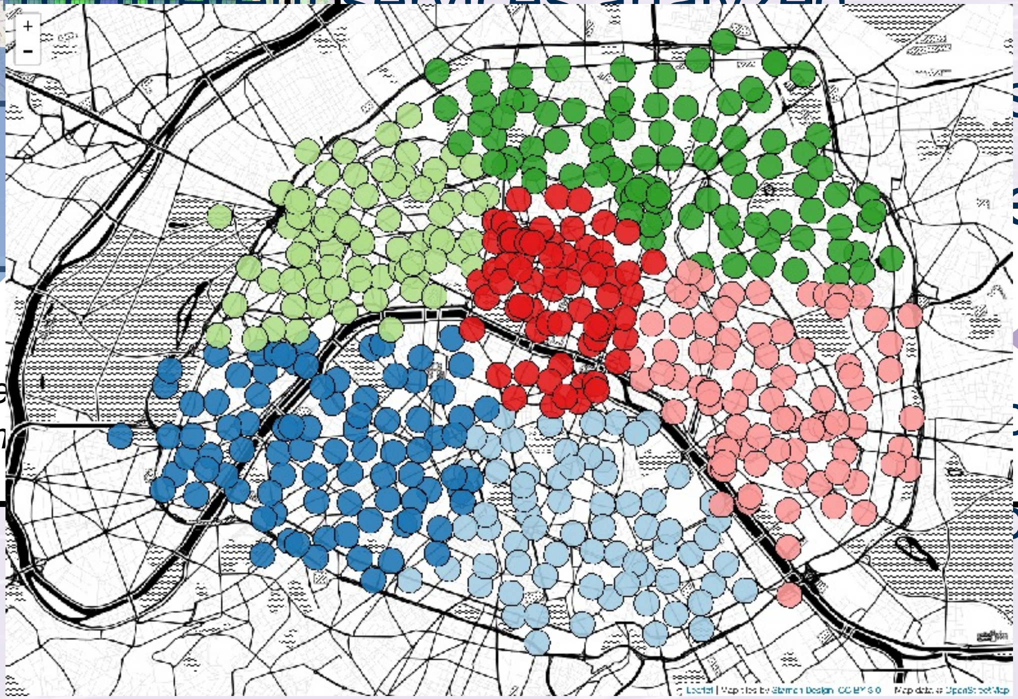


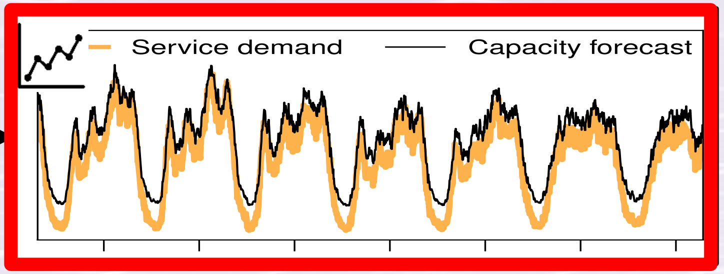
DeepCog

Reference Case Studies

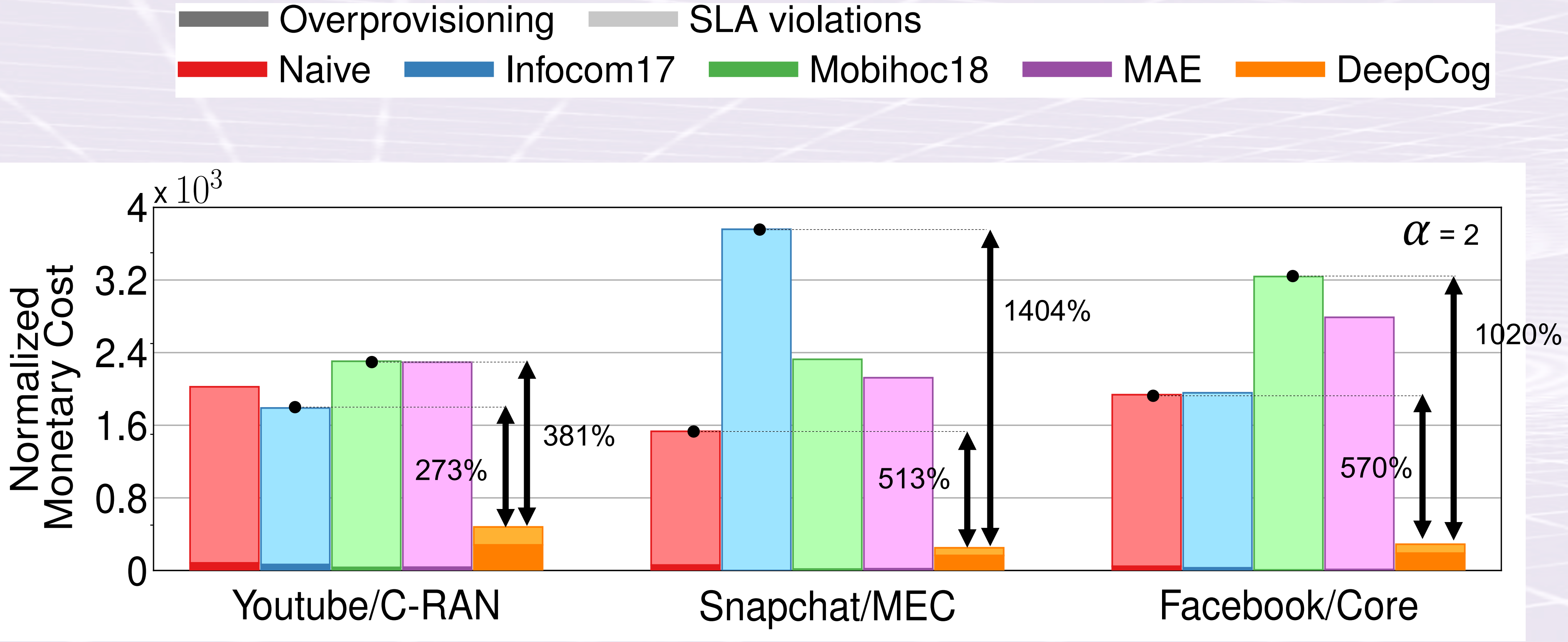
- Real-world demand generated by several millions
- Realistic scenario
  - Mobile network deployed in a large metropolitan region
- 10 different services analyzed
- ...s interval
- ...e studies
- ...odeBs)
- ...centers (70 ...des each)
- ...baseband processing or







Results



Overprovisioning
SLA violations
Naive
Infocom17
Mobihoc18
MAE
DeepCog
x 10^3
4
3.2
2.4
1.6
0.8
0
Normalized Monetary Cost
α = 2
273%
381%
513%
1404%
570%
1020%
Youtube/C-RAN
Snapchat/MEC
Facebook/Core

Outline

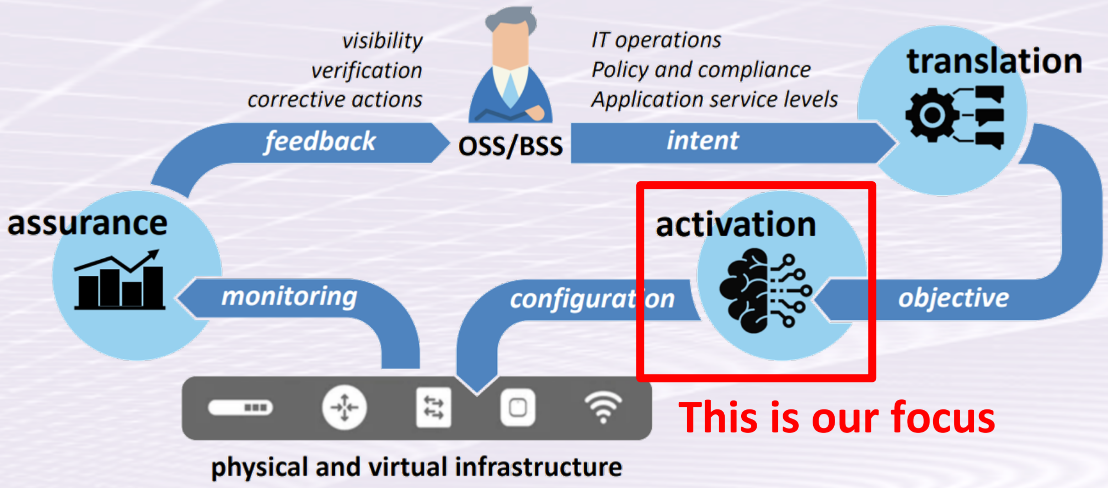
- **Data Analytics and AI framework**
- **Analysis of benefits of Dynamic orchestration**
- **Realizing Dynamic orchestration with machine learning**
- **Combining machine learning with Intent-based Networking**

Intent-Based Networking (IBN)

- Human controller dictates **high-level human-understandable** intents
  - They must be **automatically** interpreted and **implemented** by **network management entities**.

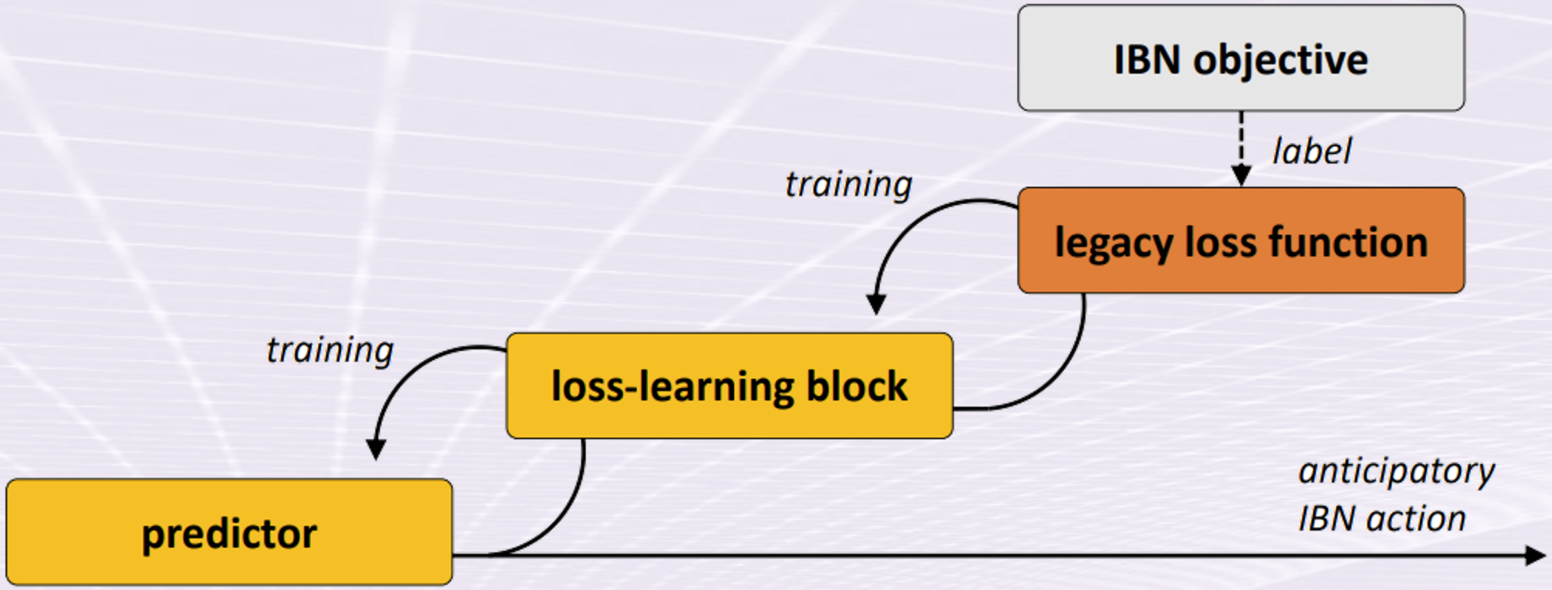
- *E.g., ensure high reliability to all Twitch traffic streaming from the Fusion Arena in Philadelphia in the next hour".*

- **Impossible** to define models to solve each possible exact task



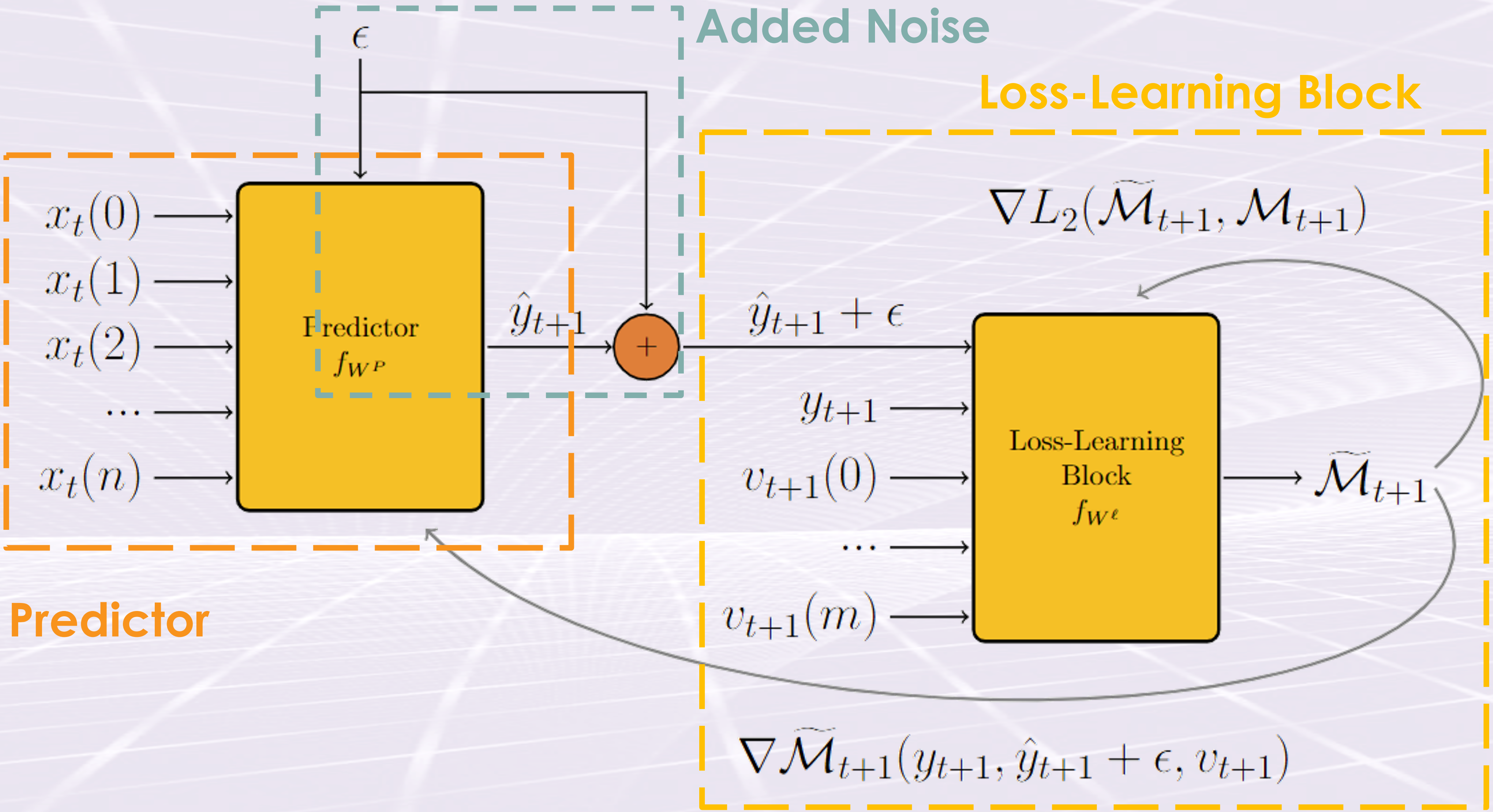
- In **anticipatory network management tasks** we can not automatically optimise **not known a-priori** metrics on-demand even with the most performing model
  - *E.g. end users QoE, depending on multiples KPIs*

The Loss Learning Predictor (LossLeaP) approach



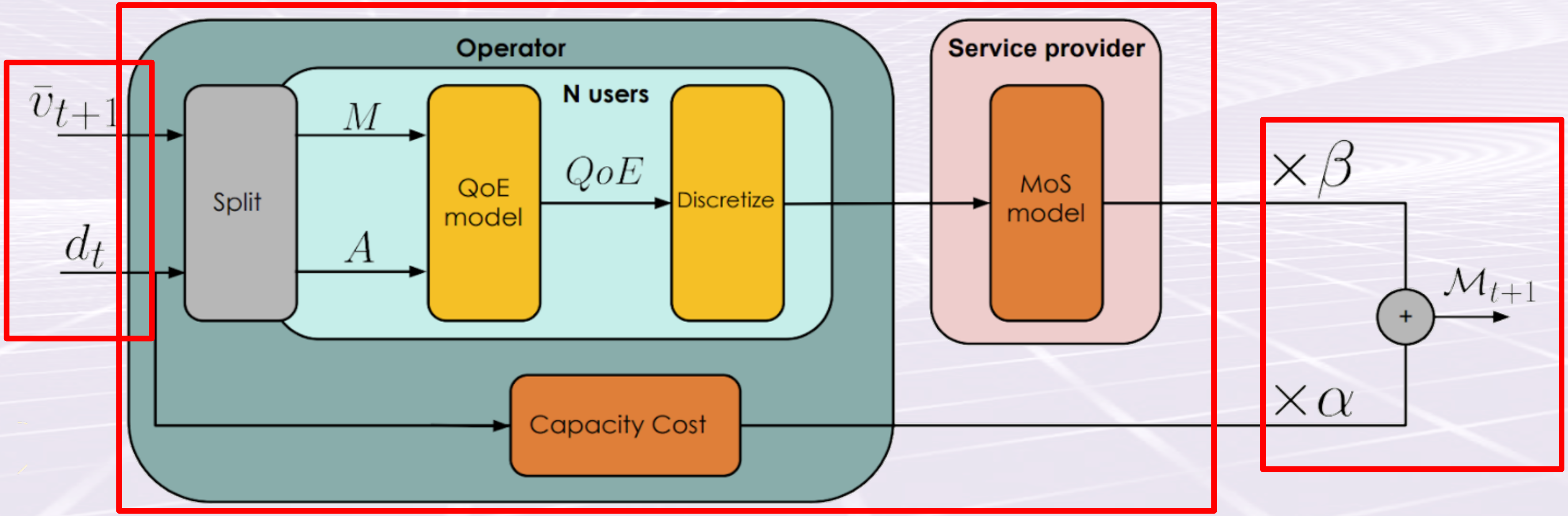
- Simple need of a Metric (**no** need **differentiability/continuity**)
- Adapt itself to any dataset **without any external tuning**
- Can shape complex **multi-dimensional** loss functions

Global architecture



Use case: maximize Incomes according to QoE

- **Full pipeline** as objective and not only an objective function
- Traffic splitted using a probability distribution among users
- **Empirical Model of QoE**
- Discretized into a **stepwise function**
- Cost if presence of **SLA violations** / Cost of **provided capacity**



Conclusions

- 3GPP has defined a framework to leverage data analytics and artificial intelligence to improve network performance
- Data-driven analyses show that performance can be very substantially improved by dynamically orchestrating network slices
- We have proposed a machine learning approach that realices the potential, focusing on capacity provisioning rather than simple prediction as existing approaches do
- In many cases loss functions are not known a priori
  - We can learn the loss function from the feedback received
  - This is a component for instant-based networking