Autonomous Network Slice Management for 5G Vertical Services

POC#9

PoC in a nutshell

- **Ambition**
  - Empower the creation of 5G Vertical Service with ENI principles
  - Apply AI/ML to Vertical Service Management and Network Slice Management functions

- **Technical goals**
  - Identify characteristics and profiles of 5G network slices in an automatic manner to meet the requirements of vertical services
    - **Exp. impact/outcome: Intent based interface**
  - Manage composition, sharing and actions for automated lifecycle of 5G network slices through AI/ML
    - **Exp. impact/outcome: ENI procedures and interfaces**

Members

Role	Organization	R&D track of origin for PoC		
		5G EVE	5GROWTH	5G-TOURS
Operator	TIM	✓	✓	✓
Manufacturer	WINGS ICT SOLUTIONS	✓		✓
Manufacturer	SAMSUNG			✓
Other	uc3m | Universidad Carlos III de Madrid	✓	✓	✓
Other	NEXTWORKS ENGINEERING FORWARD	✓	✓	

PoC Goals

DEFINED BY THE ENI FRAMEWORK

Detailed Goals

- **Use case #2-8: Automatic service and resource design framework for cloud services**
  - Extend the concepts of this use case from cloud services to 5G services, deployed across radio and transport, edge and cloud domains
  - Additional modeling items for descriptors:
    - network connectivity requirements in terms of virtual links capacity
    - QoS characteristics at the transport network level
    - service profiles expected at the radio access segment
  - Highly context-dependent:
    - action change depending on the network status

- **Use case #3-2: Intelligent network slice management**
  - Automation of the management of 5G network slices associated with multiple, concurrent Vertical Services
  - Meet service-level requirements, while optimizing the usage of the underlying 5G infrastructure, jointly considering access, core, edge, cloud resources
  - Design and implement algorithms that will be fed and assisted by the ENI system, based on short-term and long-term profiles

PoC Assessment

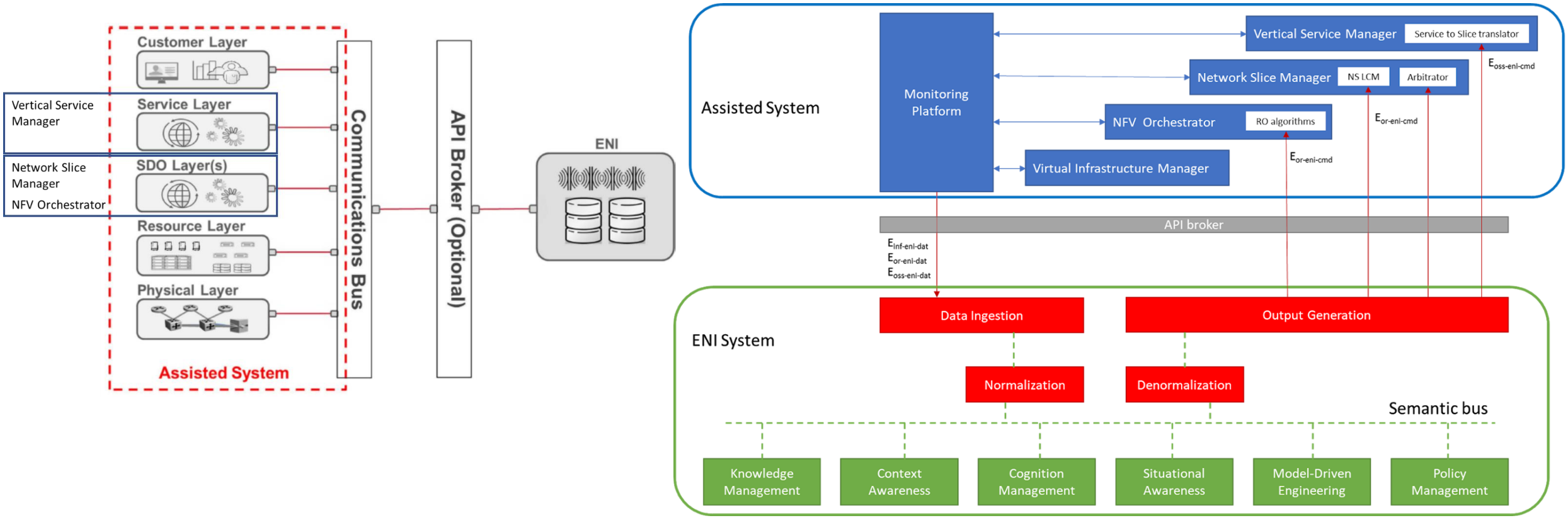
- Major functionalities of the ETSI ENI system to be validated:
  - **Ingestion and normalization** of multi-source, heterogeneous input data, related to service demands, service application performance, physical and virtual infrastructure utilization and NFV orchestration
  - **Processing of input data** to build a cross-domain knowledge about the trends of service demands, resource utilization, application and infrastructure performances and about how these elements are correlated
  - **Decision-making procedures**, generated through the Policy Management functional block
  - Assessment of the system through the **Performance Diagnostics** component which will be part of the Situational awareness module

PoC Technical Details

PoC Overview

- **Goal:** Design, develop and validate an ENI-assisted system for the intelligent management of network slices in support of vertical services operating over 5G network infrastructures
- **Software components**
  - A reference implementation of the ENI System, compliant with the architecture defined in ETSI GS ENI 005
    - ingestion and normalization of input data
    - knowledge management and processing
    - policy management
  - A multi-layer ENI-assisted system, implemented as an extended NFV MANO platform, for the management of vertical services and network slices in 5G network infrastructures.

PoC Architecture

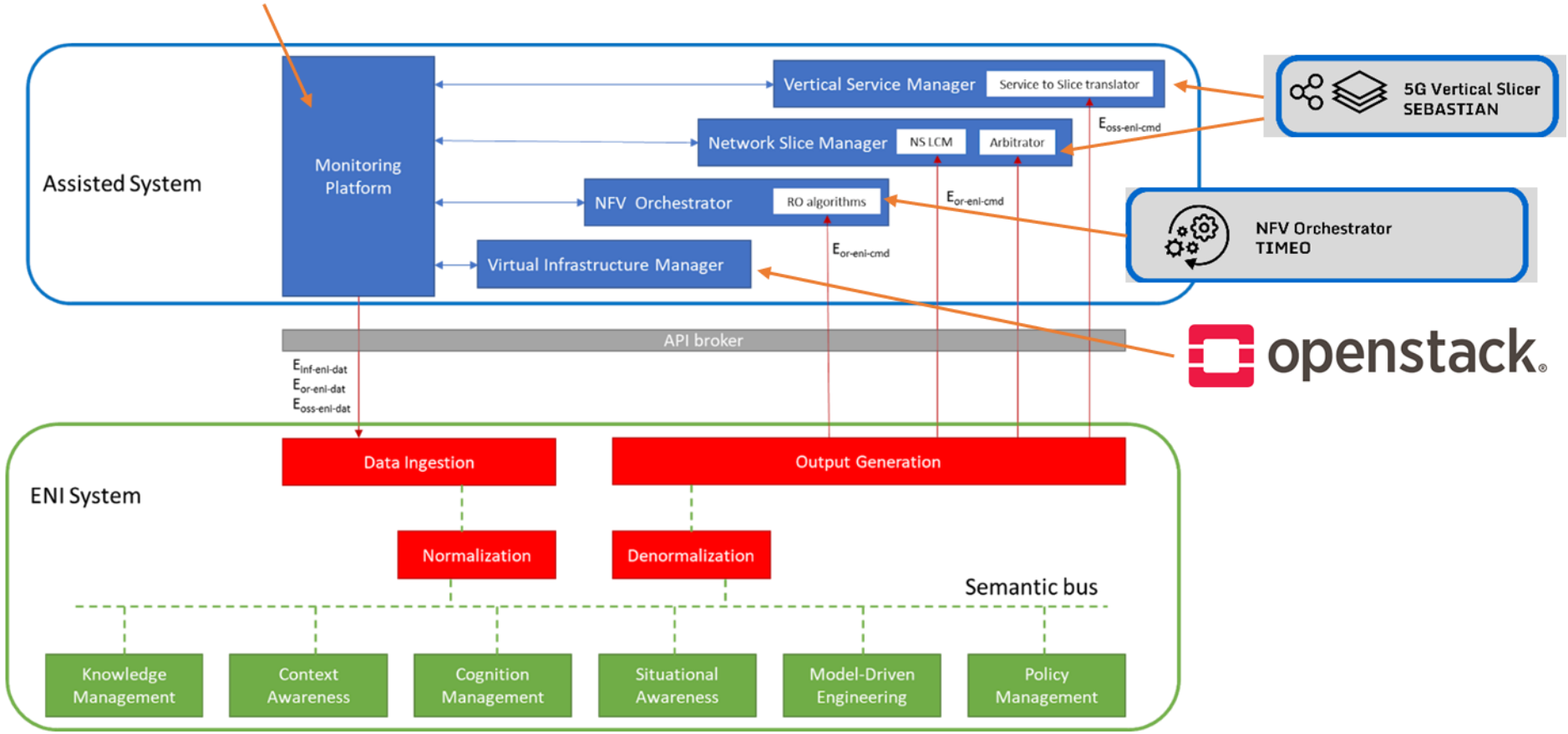


PoC Success Criteria

Goal to be verified	KPI	Stand-alone mode	ENI-assisted mode
Translation between intent-based Vertical Service definition and resource-based descriptor of the end-to-end 5G network slice	Service performance (see note 1)	Translation based on static rules preconfigured by the system administrator	Translation rules dynamically modified through policies injected by the ENI system, according to historical data about relationships between network slice characteristics and service performance
Enhanced strategies for sharing and composition of network slices.	Utilization of the 5G infrastructure. Total amount of resources used by the global set of network slices, for radio, transport, edge and core computing resources. Service performance (see note 1).	Static rules for network slice composition and sharing, applied at the provisioning time only and based on the current resource utilization and the currently active network slices.	Slice composition and sharing rules are dynamically modified through policies injected by the ENI system, according to short-term and long-term predictions for future service demands.
Automation of scaling and migration procedures for self-re-optimization of the global set of network slices.	Utilization of the 5G infrastructure. Total amount of resources used by the global set of network slices, for radio, transport, edge and core computing resources. Service performance (see note 1).	Feature not supported. Network slices are scaled manually or automatically, based on the real-time performance of single services following a threshold-based mechanism.	Suggested commands for network slice re-optimization are triggered from the ENI system, according to cross-layer and cross-domain monitoring data feeding a decisions process related to the entire set of network slices.
Note 1: Service performance will be measured through application-based KPIs, to be defined for each of the services adopted in the PoC.			

PoC Status

PoC software prototype: implementation



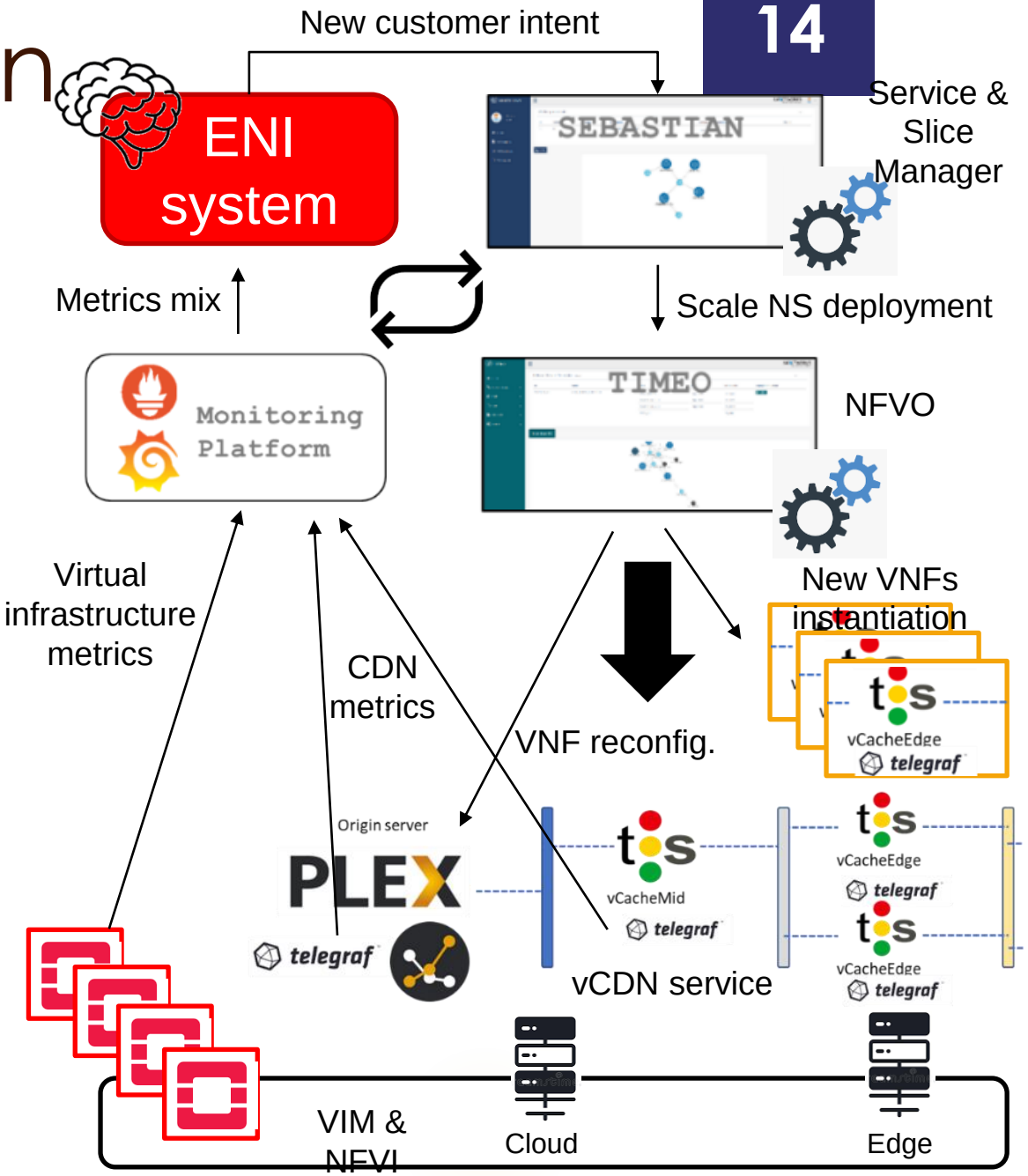
Refs:
Vertical Slicer: https://github.com/nextworks-it/slicer
TIMEO (NFVO): https://github.com/nextworks-it/timeo

Extensions implemented in the ENI-enabled NFV MANO stack

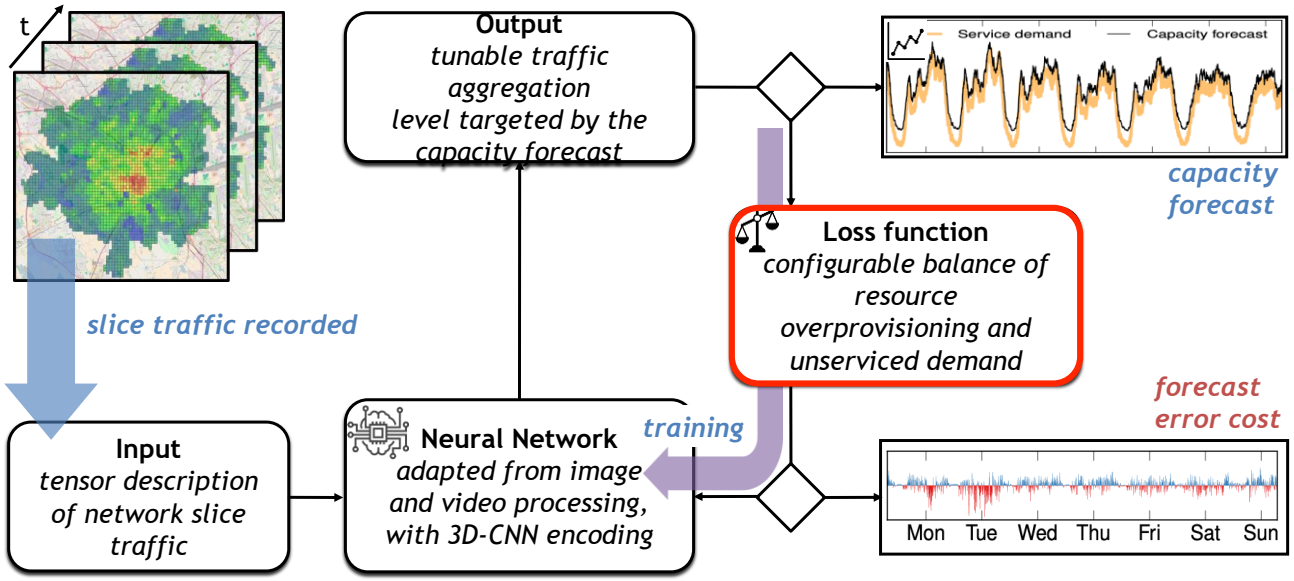
- **Programmable and cross-layer monitoring** to feed the ENI system with multi-source data: consumption of virtual resources, application statistics and performance, service demands...
  - New **probes dynamically installed and configured** in virtual application functions or as additional elements of the network service
  - Dedicated **exporters to normalize the data** sent to the monitoring platform
  - **Programmable APIs for real-time retrieval and processing of monitoring data** for training and decisions at ENI system
- Network slice management and NFV orchestration tools with extended **programmable interfaces to receive feedbacks, commands, suggestions and management policies from the ENI system**
  - Key enabler for a **multi-layer closed-loop automation**: joint decisions across vertical services, network slices, network services and virtual resources
  - Open interfaces to **externalize the logic for LCM decision making,** e.g. intent translation, slice sharing, service arbitration, resource placement
  - Enhanced internal logic to **coordinate multiple and concurrent inputs** from the ENI system, to guarantee the overall **system stability** while actuating ENI decisions
    - **conflict resolution** strategies, verification of **consistency in cross-layer commands** actuation, …

Experimental validation

- System validated with eMBB services
  - **Content Delivery Networks** with hierarchical deployment of virtual caches in edge/core network
- New ENI-enabled functionalities
  - Dynamic and automated **detection of customer's intent evolution** at service runtime
    - Based on a trained model mixing **application and virtual infrastructure metrics**
  - **Automated trigger of scaling** actions based on real-time customer's intention
  - **ENI-driven translation of customer's intent into service requirements** at the application and deployment level
    - Automated adjustment of virtual edge caches dimension, number and placement



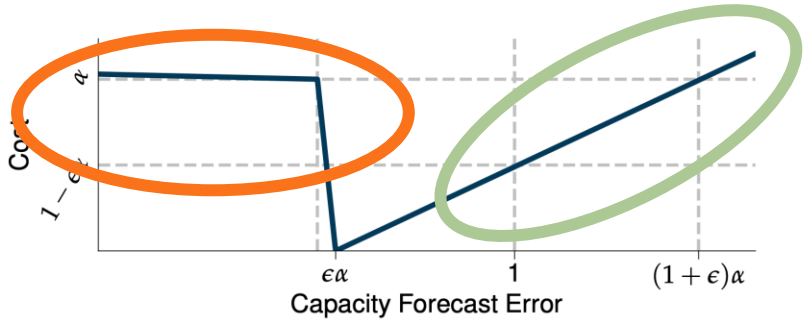
Cognitive vertical service management using deep learning



Configurable with user parameters

Fixed penalty for SLA

Proportional cost for overprovisioning

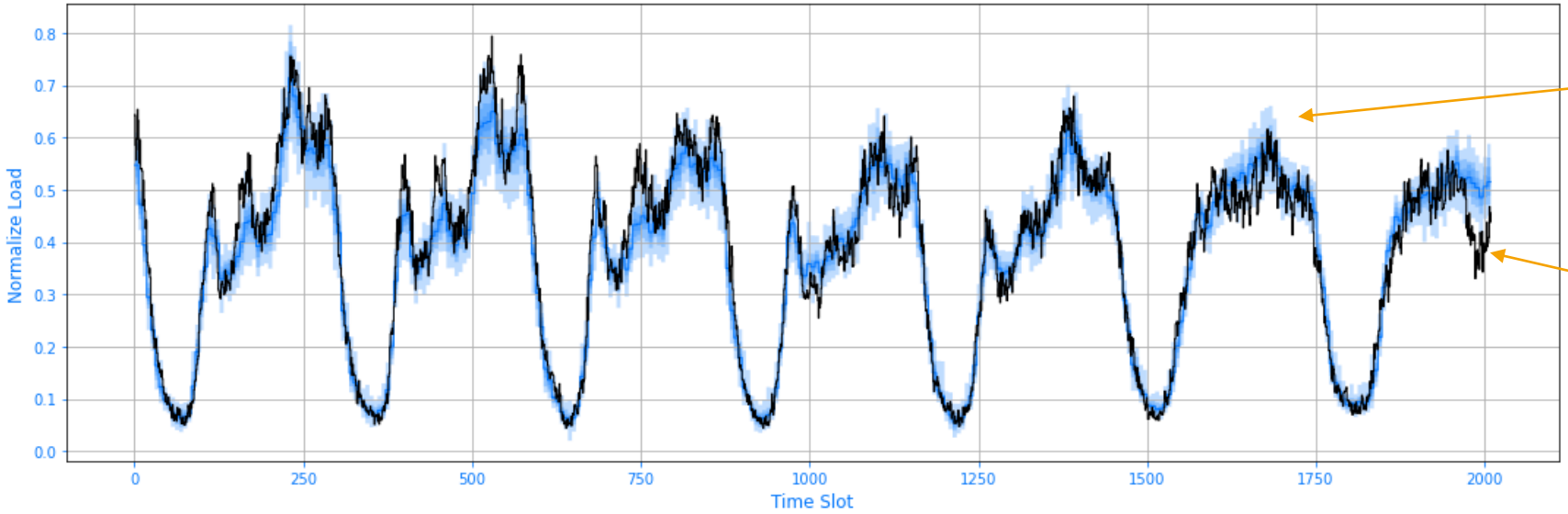


Code available at: https://github.com/wnlUc3m/deepcog

Modelling uncertainty

- Scaling decision cannot be taken too fast
  - there are certain timing that cannot be fulfilled too fast
  - Booting times of VM may take minutes
- Two timescales approach:
  - Long term forecast, that covers N prediction slots, which takes the decision about scaling or not
  - Short term forecast, to reconsider the previously taken decision, which include uncertainty
    - The rationale is: to reallocate resources only if needed
    - Modelling Bayesian network with the usage of dropout layer and Montecarlo simulators

Load forecasting for VNF scaling

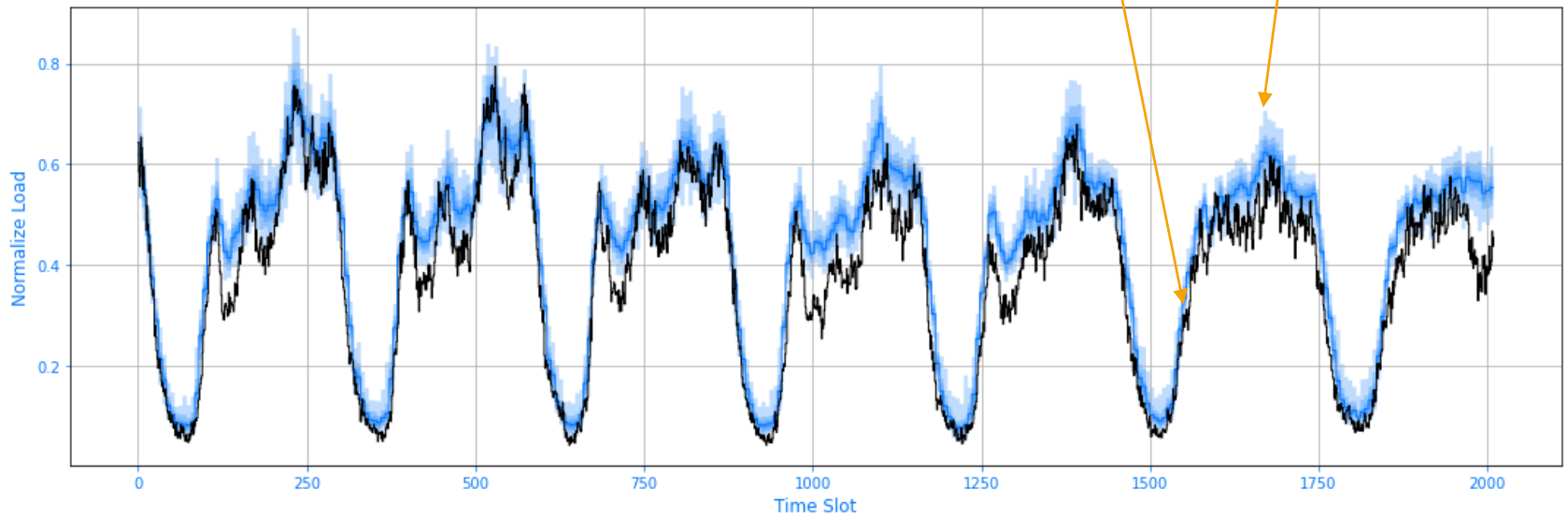


α = 1

Ground Truth

α = 5

Higher values of β entail more resource expenditure (scaling up are performed more frequently), but they are not incurring into resource shortages.

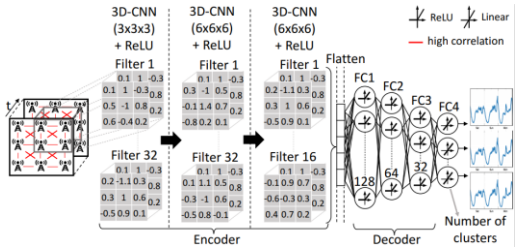


Overview of ETSI ENI enhancements in the NFV MANO stack

Service scaling commands for new intents

ENI System

Application and infrastructure metrics



5G Service Development Kit (SDK)

5G Vertical Slicer SEBASTIAN

ENI-driven intent translation and scaling automation

5G Apps & Services Catalogue

Plug & Play Control Plane

NFV Orchestrator TIMEO

ENI-driven placement algorithms & monit. Config.

Monitoring & Performance

Probes for appl. metrics

WIM

VIM

Virtual Network Functions

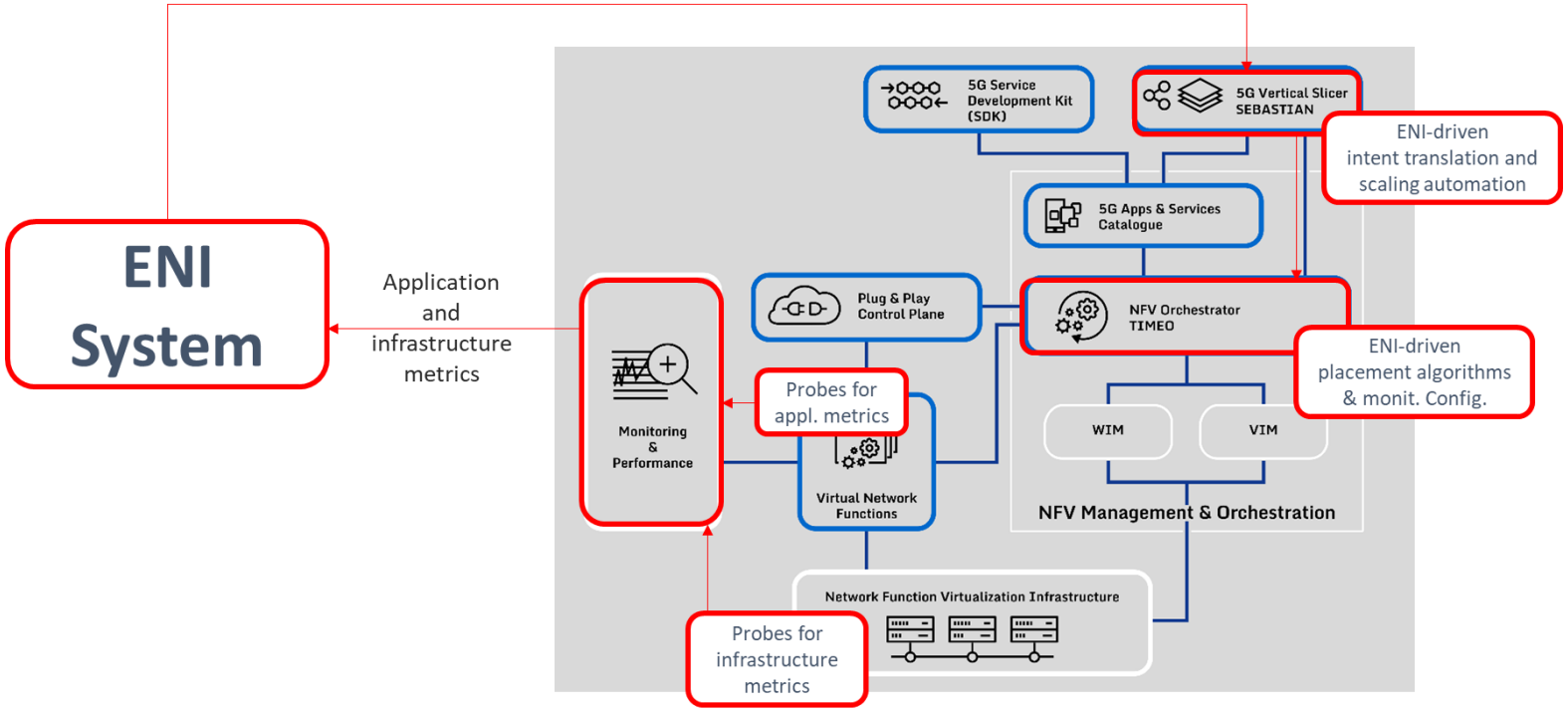
NFV Management & Orchestration

Network Function Virtualization Infrastructure

Probes for infrastructure metrics

Integration of ENI System with OSM

- Work in progress: **integration of ETSI OSM in ETSI ENI PoC system**
  - Vertical Slicer with a **new driver** to interact with OSM
    - Info model translation for VNFDs and NSDs
    - LCM commands translation, including mapping between **different NS scaling approaches**
  - **Adaptation of OSM monitoring mechanisms**
    - Unify different metrics as input for ENI system
    - Dynamic monit. configuration
  - External **placement algorithms**



PoC Success Criteria Validation

Goal to be verified	Stand-alone mode	ENI-assisted mode	Validation
Translation between intent-based Vertical Service definition and resource-based descriptor of the end-to-end 5G network slice	Translation based on static rules preconfigured by the system administrator	Translation rules dynamically modified through policies injected by the ENI system, according to historical data about relationships between network slice characteristics and service performance	Intent of the Vertical captured through the configuration parameter of the deep learning function. Interface exposed to the vertical
Enhanced strategies for sharing and composition of network slices.	Static rules for network slice composition and sharing, applied at the provisioning time only and based on the current resource utilization and the currently active network slices.	Slice composition and sharing rules are dynamically modified through policies injected by the ENI system, according to short-term and long-term predictions for future service demands.	No-multi slice setup was tested, but the specific functionality is implemented for one slice
Automation of scaling and migration procedures for self-re-optimization of the global set of network slices.	Feature not supported. Network slices are scaled manually or automatically, based on the real-time performance of single services following a threshold-based mechanism.	Suggested commands for network slice re-optimization are triggered from the ENI system, according to cross-layer and cross-domain monitoring data feeding a decisions process related to the entire set of network slices.	Fully validated with scaling techniques implemented on the VNFs
Note 1: Service performance will be measured through application-based KPIs, to be defined for each of the services adopted in the PoC.			

Conclusion

- We successfully implemented PoC#9 using an eMBB network slice as reference
- Due to the COVID-19 pandemic it was not possible to test it using a real use case, as initially planned
- We validated the criteria:
  - Translation of vertical intent into orchestration primitives, through the setup of the parameter β
  - Automatic scaling of the VNF instances, according to the CPU load
  - Full integration of the solution into the OSM
- We also found some possible points to be further analysed

Additional architectural considerations

- Multi-timescale approach
  - It is fundamental when dealing with real environment with non-immediate actions
- Exposure of machine learning parameters to verticals
  - Connect business intelligence with network operation
  - Respecting security and privacy of the exposed information

Questions?