AI and ML in Software Defined Networks

Software-Defined Mobile Networks Course

Department of Electrical Engineering
Sharif University of Technology

Winter 1401

AI and ML not new concepts

- They have been around for many decades
  - Roots of ML in what we already knew for many years in EE as Adaptive Filters, Neural Networks and Estimation Theory
- Why are they so trendy now?!
  - First: Computational power
  - Second: Accessibility of data and ability of real time systems to be more controlled by software

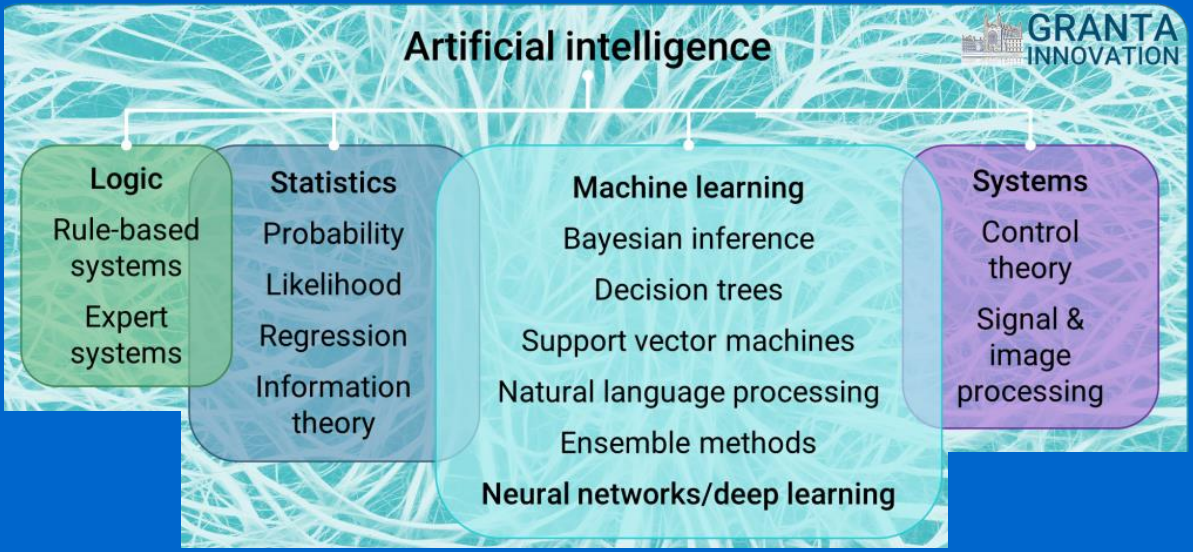
What is AI?

- What is the first key step to become *intelligent*?

Feedback!

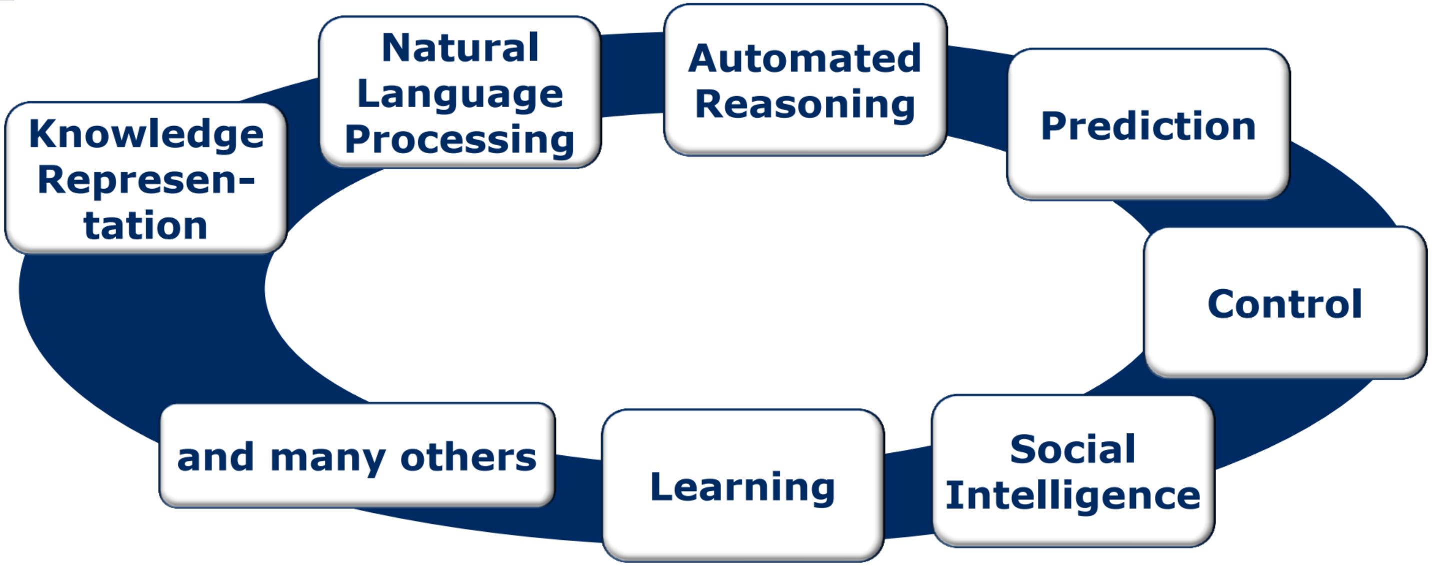
(A lot more than 10^20 feedback loops in our body!)

- But that is not enough. We also need to *adopt to, learn* new situations, and *act* accordingly!
- Systems that perform *reasoning*, *planning* and object management using *knowledge* and *perceived information*.



AI (Machine Intelligence)

- **There is a huge variety of approaches to AI . . .**



- **. . . with just a few big success stories**
  - (Deep) Machine Learning – biggest AI boom today
  - Natural Language processing – Siri, Alexa, Cortina, etc.
  - Profiling and Prediction – web ad placement

Plenty of AI Methods

Limited number of Functionalities

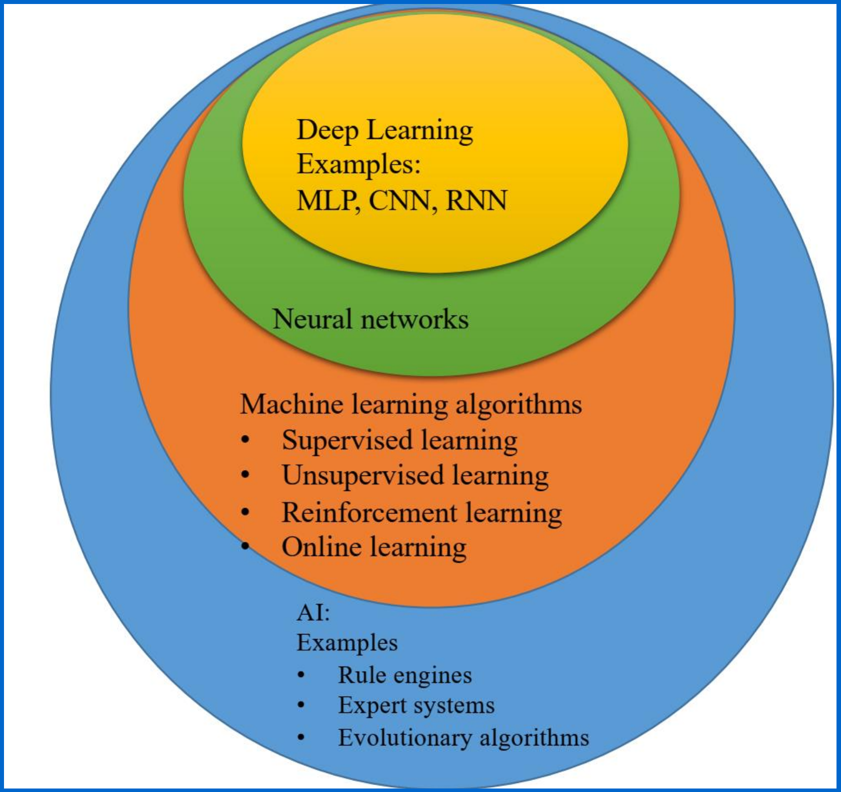
Data description	**Deriving statistical characteristics of data**
Data segmentation	**Grouping of data into homogeneous clusters**
Data association	**Discovering interesting relations between variables**
Data classification	**Finding the function linking target categorical variables with input variables**
Data regression	**Finding the function linking target numerical variables with input variables**
Data forecasting	**Predicting the value of a target variable for the future**
Variation detection	**Determining possible drifts in data characteristics**
Anomaly detection	**Identifying items which do not conform to an expected pattern**
Sequential optimi-zation of parameters	**Controlling an interactive system or environment**

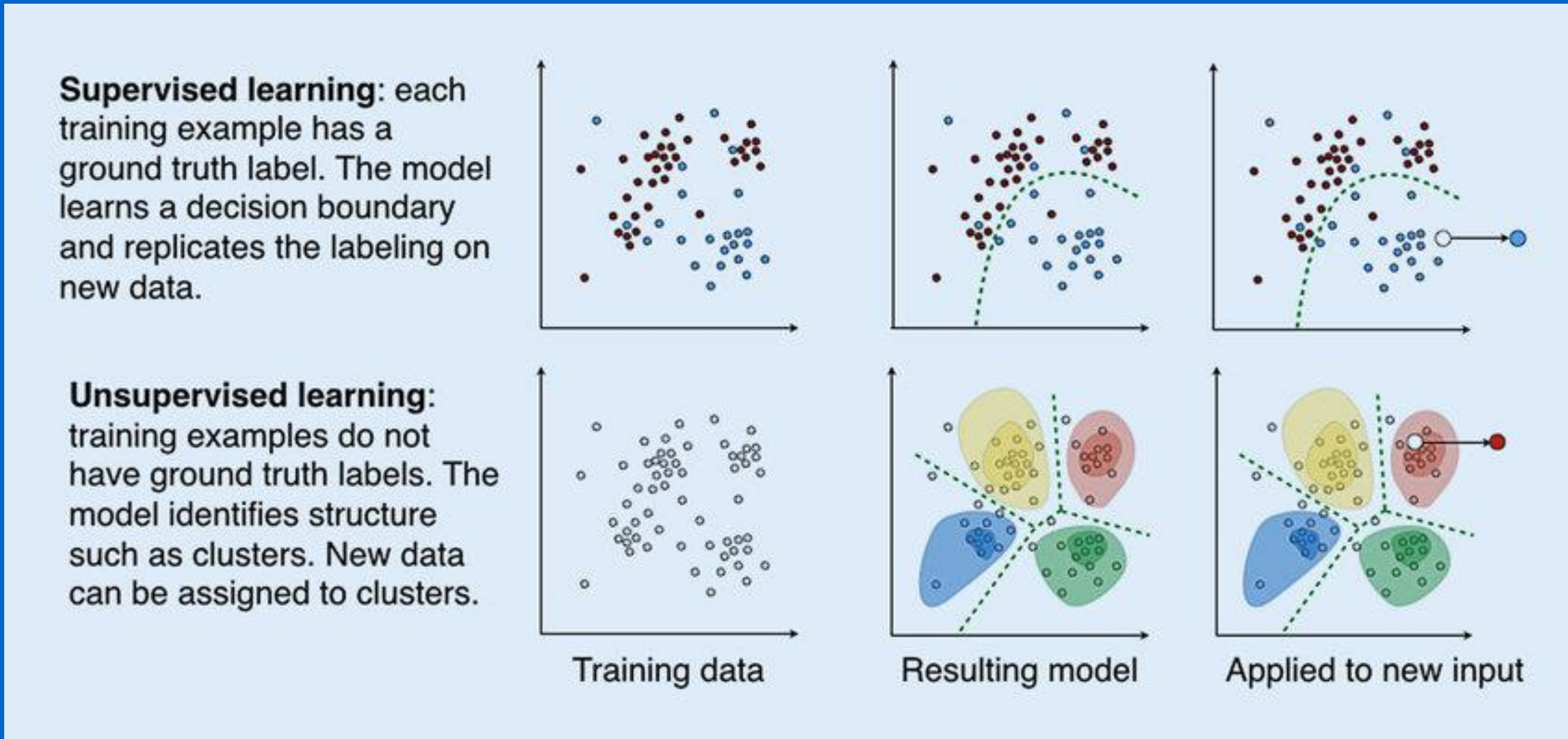
Steps to take in real-world use cases

Research	Proof of Concept (PoC)	Trial	Operational
The use case is investigated by theoretical and simulation studies	The use case has been explored with real operational data on a limited scope	The use case has been explored with real operational data on a limited scope	The use case is already implemented in the field
Datasets do not necessarily come from the field	OR was run on a real software / network infrastructure	AND was run on a real software / network infrastructure	Still improvements of AI method and its implementation can be sought

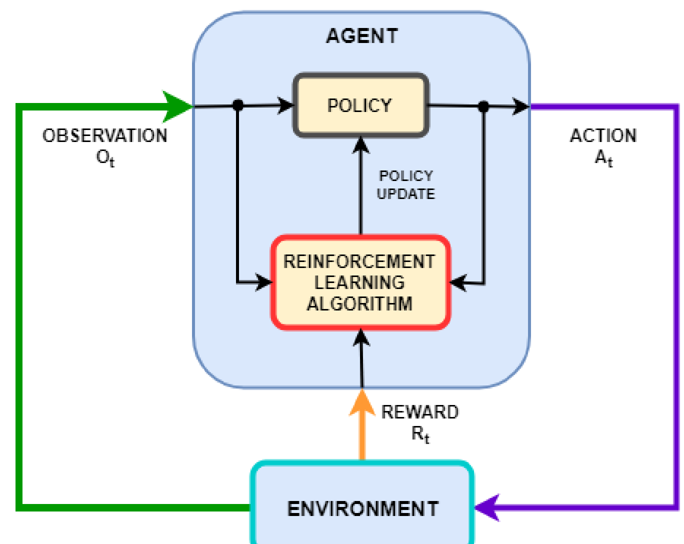
Building an AI Application

Learning Approach



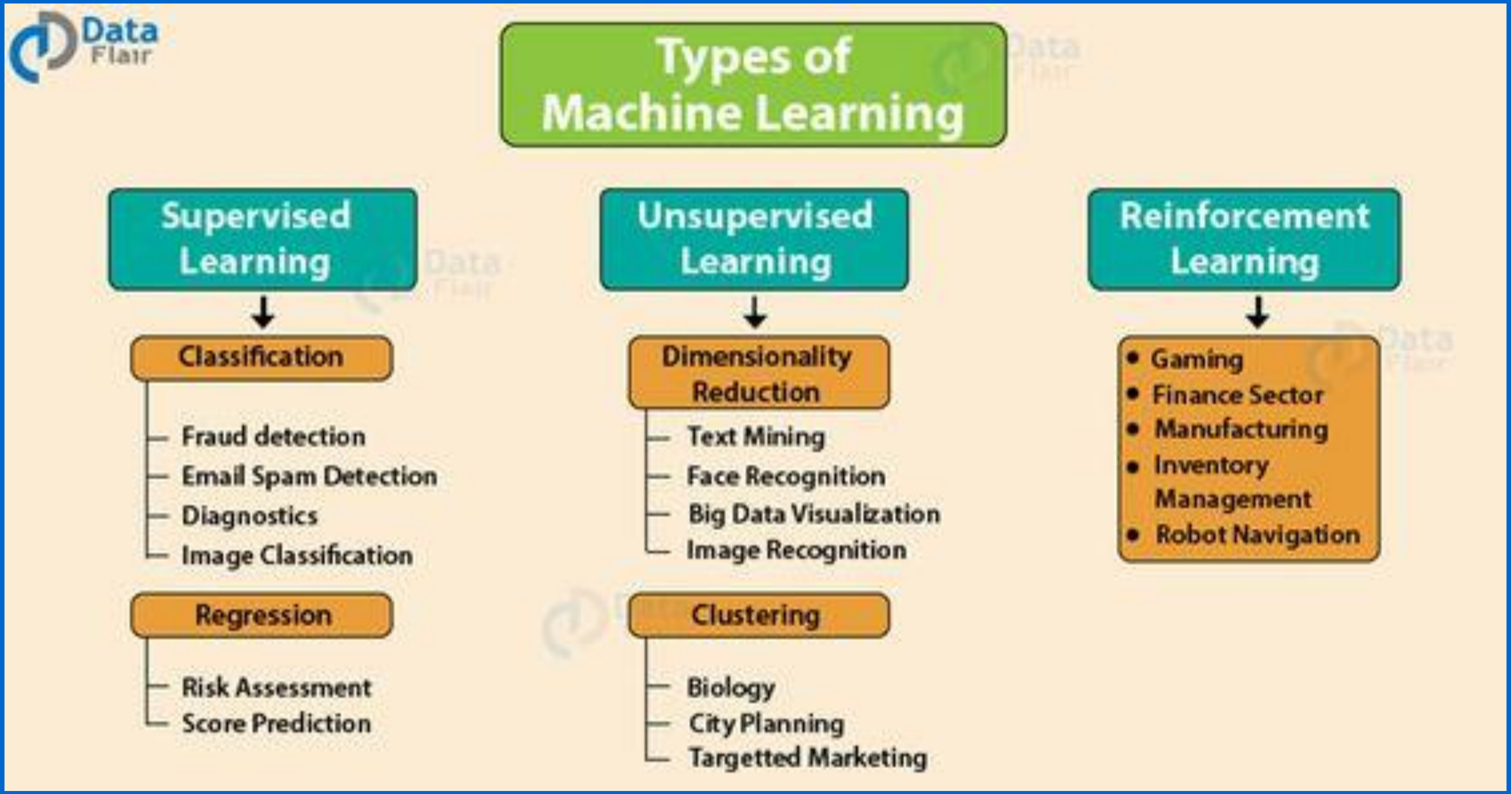


- Agent performs action A to change state of the system
- No indication of correct action is given, just a feedback value called reward after one or more time steps
- Learning system tries to optimize for expected cumulative future reward
- Well-known as MDP (Markov Decision Process) in EE



Reinforcement Learning

Application of ML Variations



Laundry List of ML Algorithms

- Linear and Logistic Regression
- Support Vector Machines
- Random Forest
- Naïve Bayes Classification
- K-means
- Ensemble Methods
- Principal Component Analysis
- Singular Value Decomposition
- Independent Component Analysis
- …

Role of AI in Telecommunications

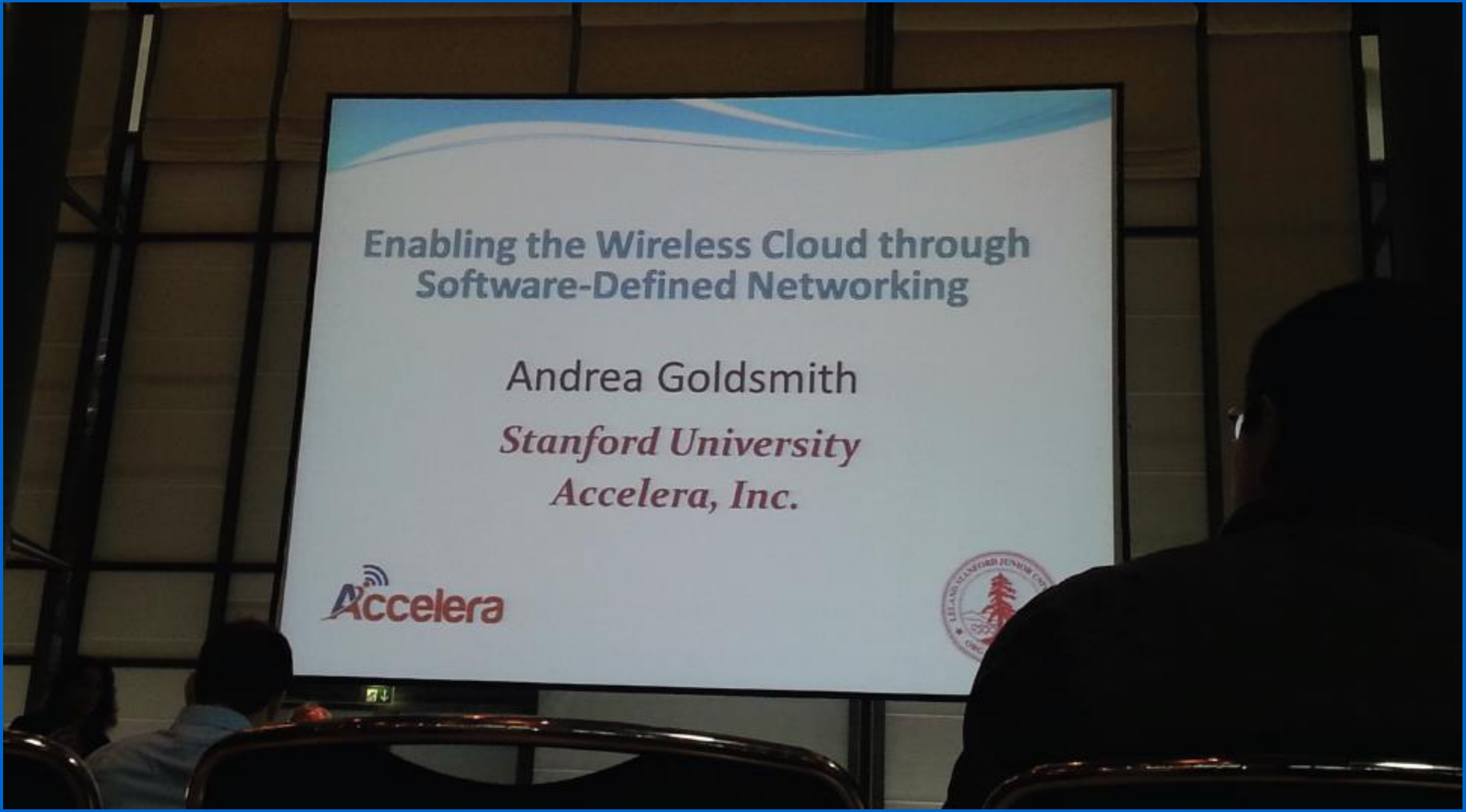
- For network operators, there are two large classes of AI applications:
  1. those that run over the network to improve operator revenue through big data analytics (customer behavior,…)
  2. those that run within the network to <u>optimize some aspect of network operation</u> or <u>management</u>.

- For first type of application:
  - **Goal:** New Network Services and Monetizable Data Insights
  - Estimate customer preferences and usage, to provide improved user experience
  - Perform service personalization by tracking device behavior such as mobility patterns or the types of services the customer uses
  - Use expert systems to improve customer contact, support, sales activities and overall customer satisfaction.

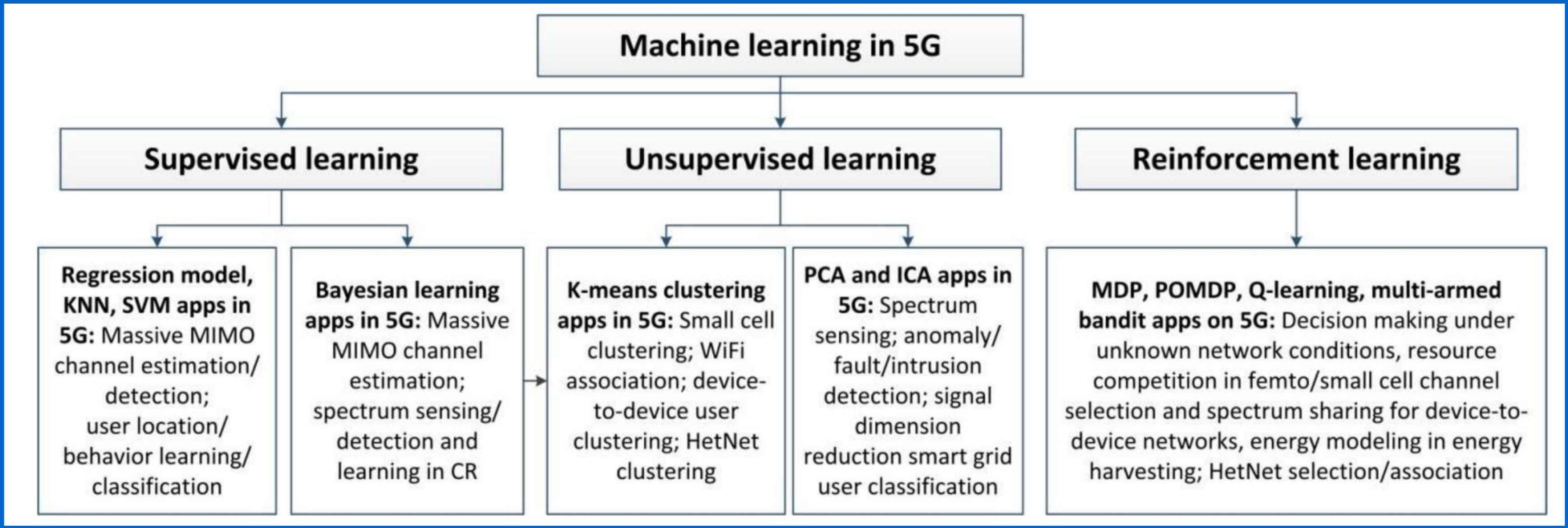
AI Inside Telecom Network

- For second type of application:
  - Network AI could be used to replace or enhance network planning, service deployment and management functions, typically operating with a long time constant
  - It also could be used in near-real time to dynamically optimize network performance based on rapidly changing traffic patterns
- Two levels of AI:
  - Automation at RAN level: SON
  - Network Management and Control

RAN and SON


Enabling the Wireless Cloud through Software-Defined Networking
Andrea Goldsmith
Stanford University
Accelera, Inc.
Accelera

AI in RAN



Machine learning in 5G
Supervised learning
Unsupervised learning
Reinforcement learning
Regression model, KNN, SVM apps in 5G: Massive MIMO channel estimation/ detection; user location/ behavior learning/ classification
Bayesian learning apps in 5G: Massive MIMO channel estimation; spectrum sensing/ detection and learning in CR
K-means clustering apps in 5G: Small cell clustering; WiFi association; device-to-device user clustering; HetNet clustering
PCA and ICA apps in 5G: Spectrum sensing; anomaly/ fault/intrusion detection; signal dimension reduction smart grid user classification
MDP, POMDP, Q-learning, multi-armed bandit apps on 5G: Decision making under unknown network conditions, resource competition in femto/small cell channel selection and spectrum sharing for device-to-device networks, energy modeling in energy harvesting; HetNet selection/association

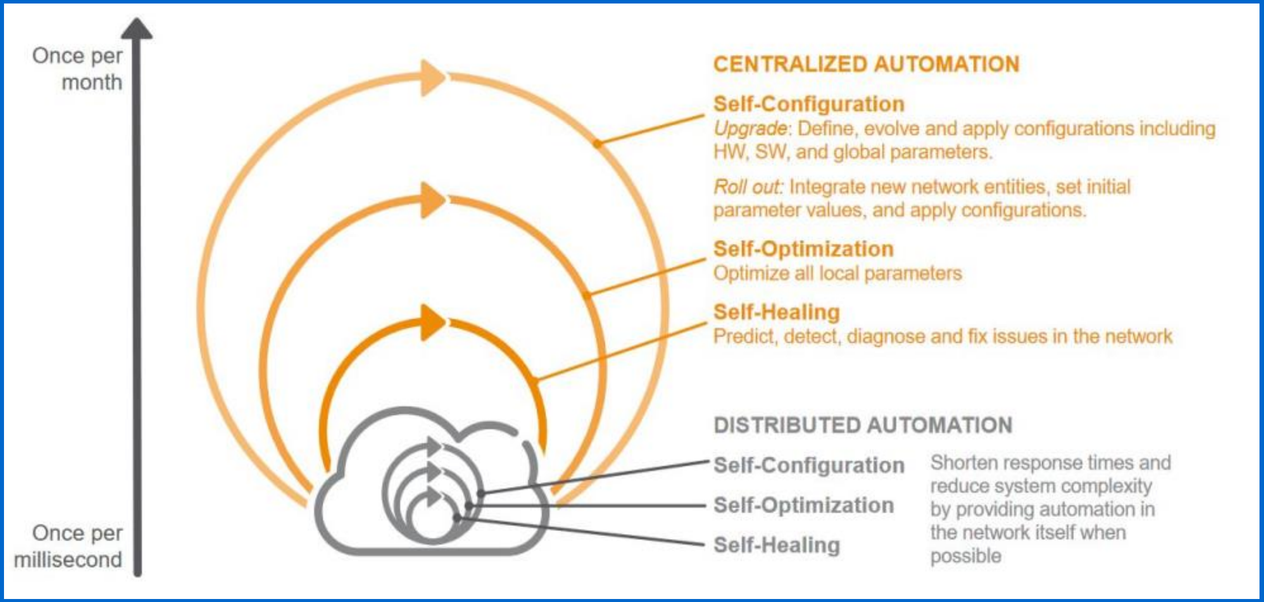
RAN from a Network Perspective

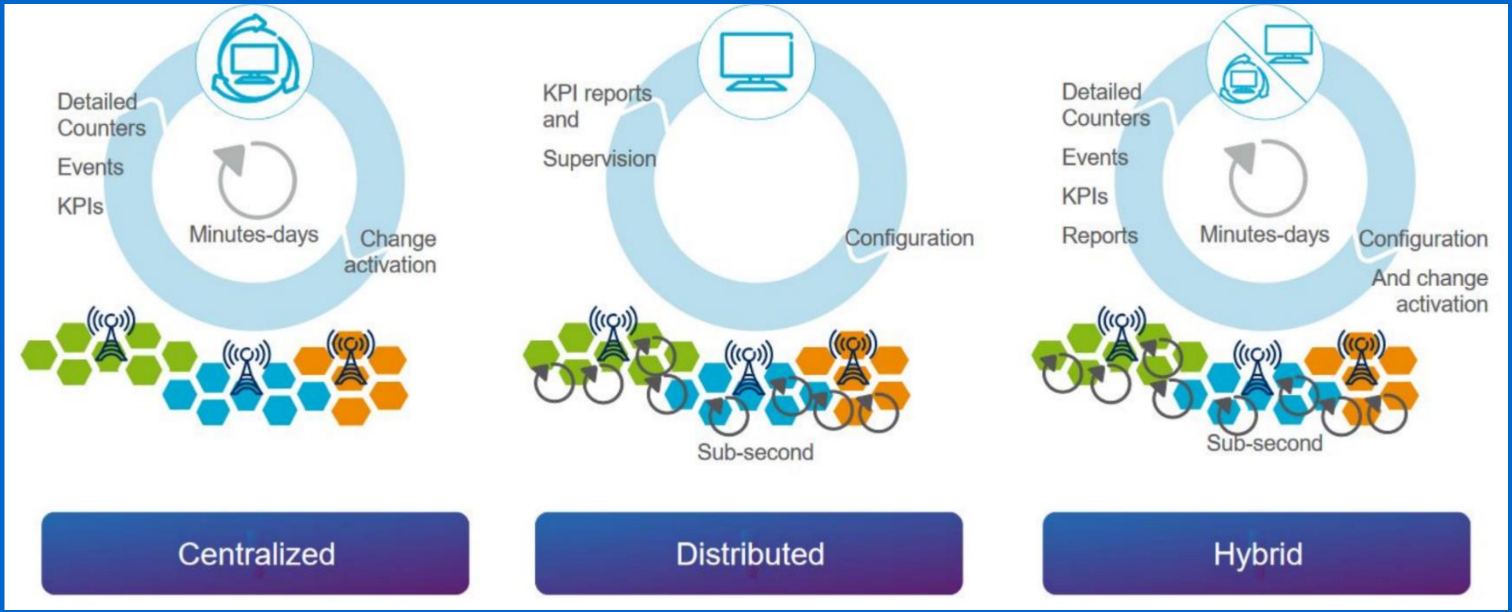
- Where to bring up RAN Virtual Network Functions, typically driven by constraints on latency and jitter
- Overcome inter-cell interference through coordination of transmissions
- How these nodes are associated and the routing between them is configured
- How to control schedulers and when handovers are performed
- Some degree of configuration via APIs, orchestrators or similar entities.

RAN **state information** and **performance metrics**

- Cell physical aspects such as topology, spectrum used (number of carriers in each band), power levels and cell neighbor relationships including small cell overlays within macro cells
- Number of active UEs in a cell, along with some measure or indication of UE specific traffic characteristics (e.g., voice, data, SMS)
- UE signal, interference and noise ratios in the cell and some measure of UE mobility
- Measurements of recent UE and/or cell performance including call/session drop rates, handover performance, overall packet capacity and performance
- Network transport and routing infrastructure metrics indicating packet loss, latency and throughput, information indicating what applications are running on what devices in which subscriber groups and network slices and at what locations
- …

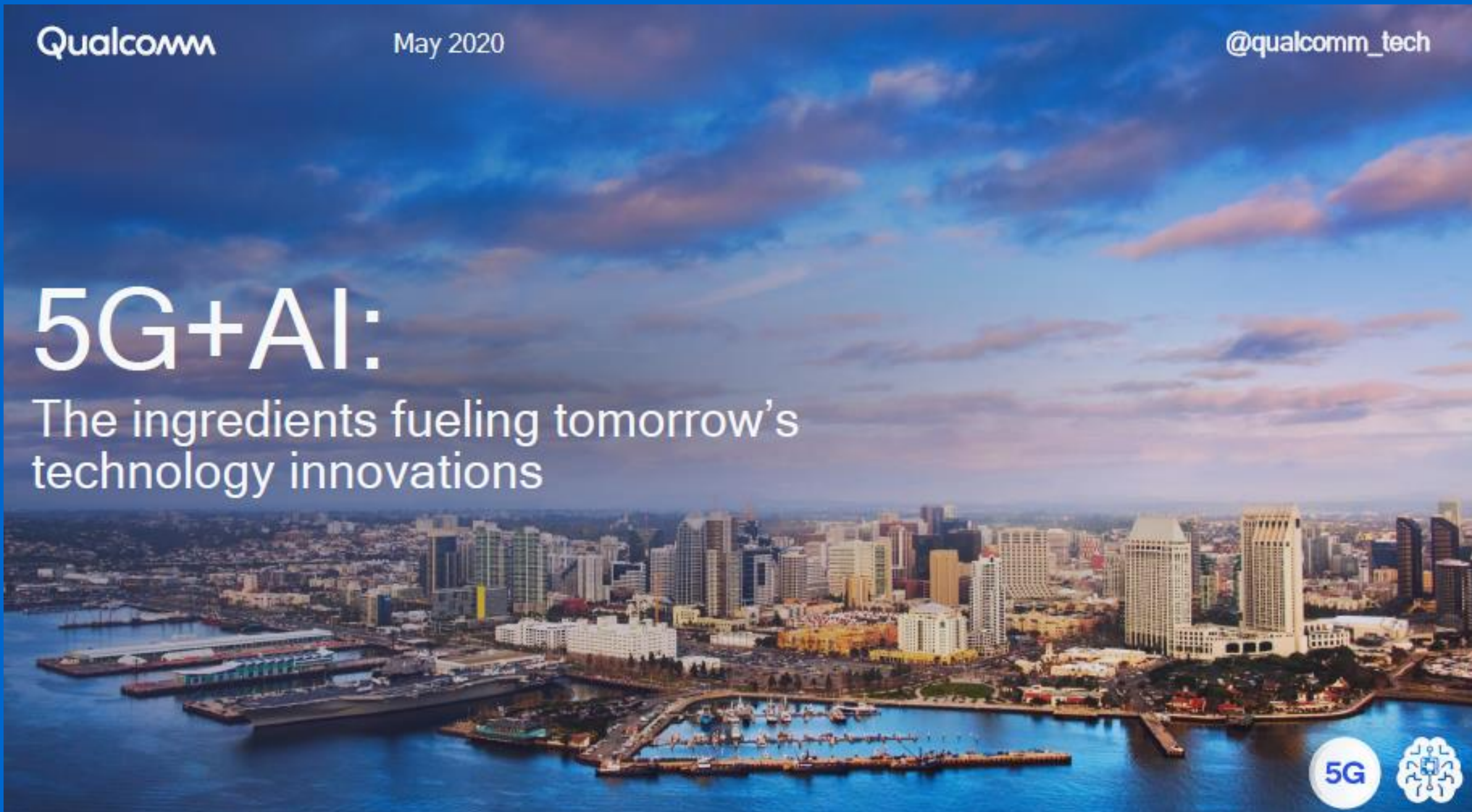
SON Automation Time Scales





Traditional vs. Advanced SON

Capability	Basic SON Solution	Advanced SON Solution
Infrastructure Optimization	*Vendor Proprietary*	*Vendor Agnostic Managed Service*
Automated modules	*Vendor specific*	*All vendors*
Product suite pre-populated with optimization rules	*User templates for Policy Rules*	*Pre-populated with optimization rules based on prior populated data*
Tools for Productivity (UI, Wizard etc.)	*GUI and Expert System Interface*	*Complementary dashboards for SON actions and reports. Self-Configuration Wizard for fully automated initial site configuration.*
Extensible software solution	*Upgradeable with new releases only*	*Rapidly add capabilities to handle new issues e.g. unique regional interference problems*
New capabilities rapidly integrated	*Upgrades on standard software release schedule*	*Software releases upgraded automatically to reflect new requirements*
Directional antenna statistics	*General Statistics for Network Traffic*	*Detect misalignment and crossed cable problems without drive testing*
Solutions adjust rapidly and automatically	*In Minutes or Hours*	*Adjust rapidly to unexpected disruptive events including when base station:* *(a) is isolated e.g. during an earthquake* *(b) lost in terrorist attack or other disaster* *(c) must reassign spectrum to first responders or public safety.* *Goal: Auto Recover in 10 -30 minutes*
Dynamic C-SON parameters	*Manual trigger for cell coverage recalculation*	*Automatically vary macro cell and small cell coverage and capacity to handle unexpected traffic or add coverage if competitors are out of service.*
Modules to automate HetNets and Spectrum Refarming	*Basic Multi-Frequency planning tools*	*Modules to determine best cells for refarming from 2G/3G to 4G and automate 2G/3G refarming for 4G/LTE deployment*
Optimized capacity planning across macrocells and small cells for multiple frequencies	*Area specific Macro cell optimization*	*Optimized network wide capacity planning across 2G, 3G and 4G - both macro cells and small cells*
'5G ready' for dynamic Hybrid HetNets	*2G, 3G and 4G separately*	*Dynamic HetNets where 5G micro sites or small cells only provide partial coverage*
5G Service ready 'Future proof'	*Varies*	*Upgradeable for New Radio, 'NB-IoT' 'Network Slicing' etc.*
Automated Scaling	*Scale in and out based on Policy or Triggers*	*Dynamic Scale in and out of small cells or virtual network functions based on Policy, Traffic or Events*
Machine Learning (ML) and AI	*Embedded Rules with Open Loop feedback based on New Data and Bayesian updaters*	*Predictive Smart Algorithms based on Neural Networks calculated with historical data on multiple production installations*


Qualcomm
May 2020
@qualcomm_tech
5G+AI:
The ingredients fueling tomorrow's technology innovations
5G

Applying AI to solve difficult wireless challenges

Deep wireless domain knowledge is required to optimally use AI capabilities

Wireless challenges

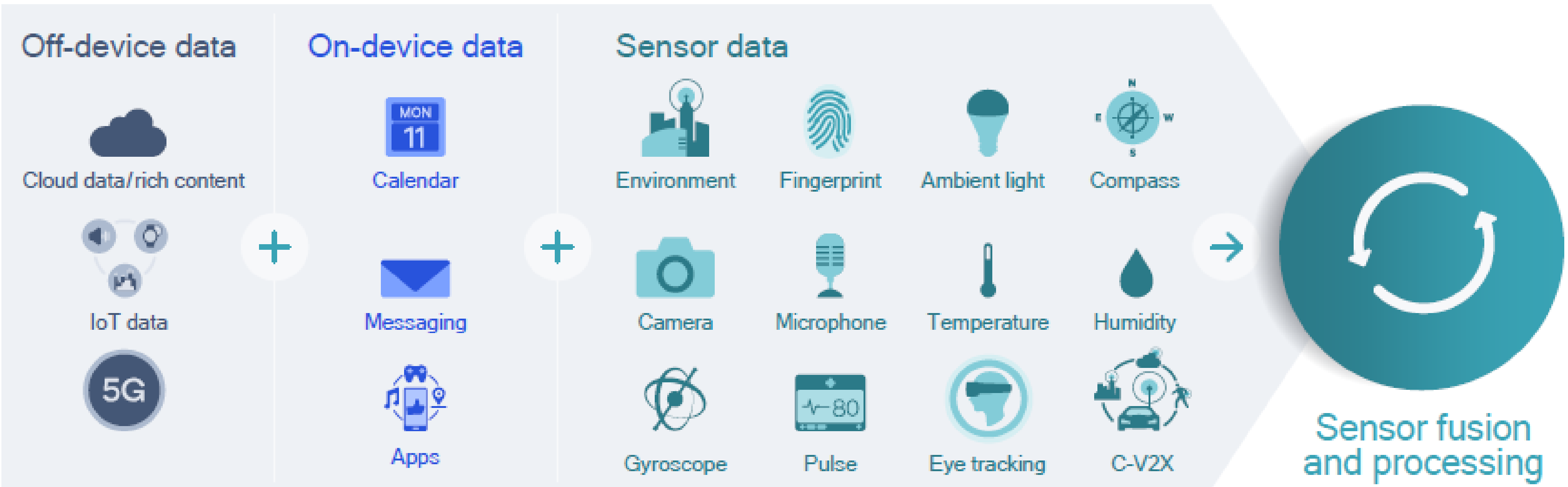
- Hard-to-model problems
- Computational infeasibility of optimal solution
- Efficient modem parameter optimization
- Dealing with non-linearity

AI-enhanced wireless communications

AI strengths

- Determining appropriate representations for hard-to-model problems
- Finding near-ideal and computationally realizable solutions
- Modeling non-linear functions

Devices generate and possess massive amounts of data



On-device AI processing of sensors and personal information conserves bandwidth while providing contextual intelligence, personalization, and privacy

On-device AI enhances 5G device experience

Better beam management

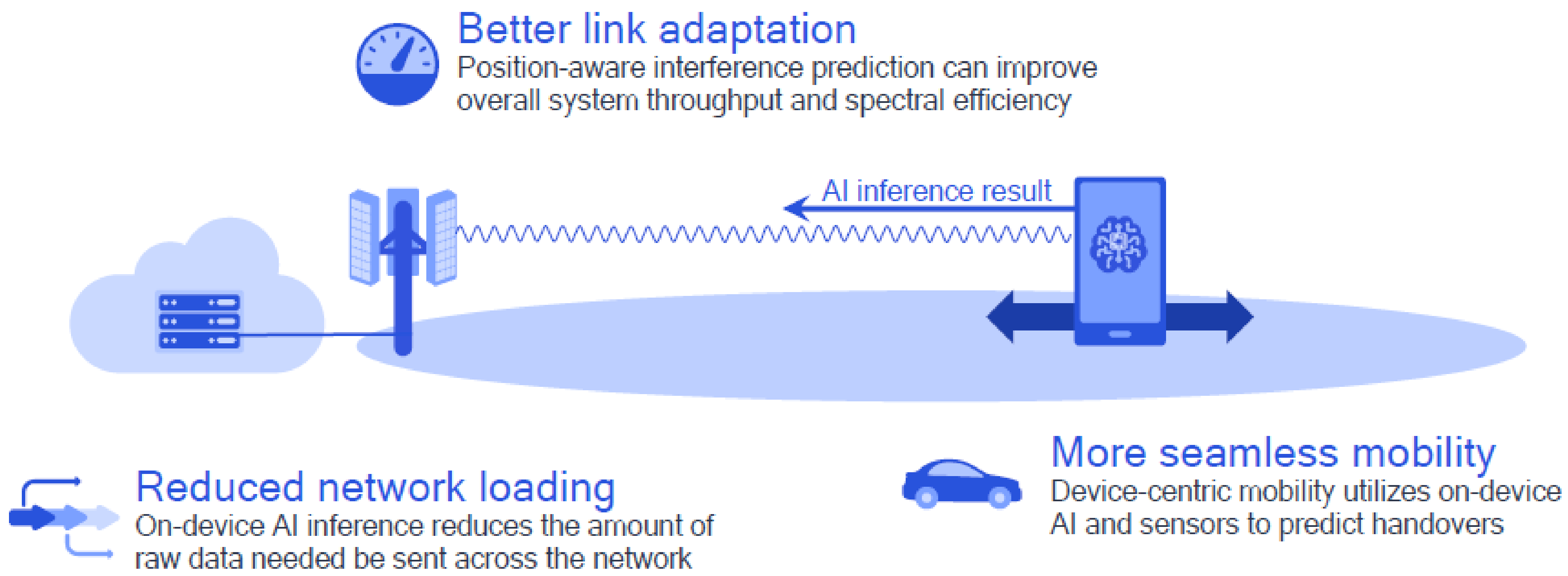
Incorporate location, velocity, other aspects of environmental and application awareness to improve robustness and throughput

More power saving

Optimize performance/power consumption tradeoffs by taking advantage of better contextual awareness on device

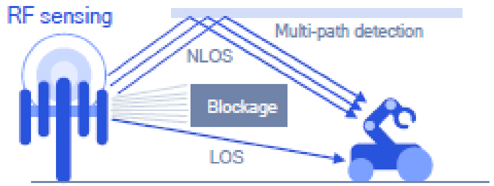
Deep reinforcement learning example

On-device AI improves 5G system performance



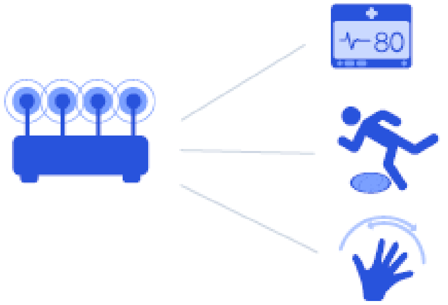
Applying AI for contextual awareness & environmental sensing

More accurate device positioning

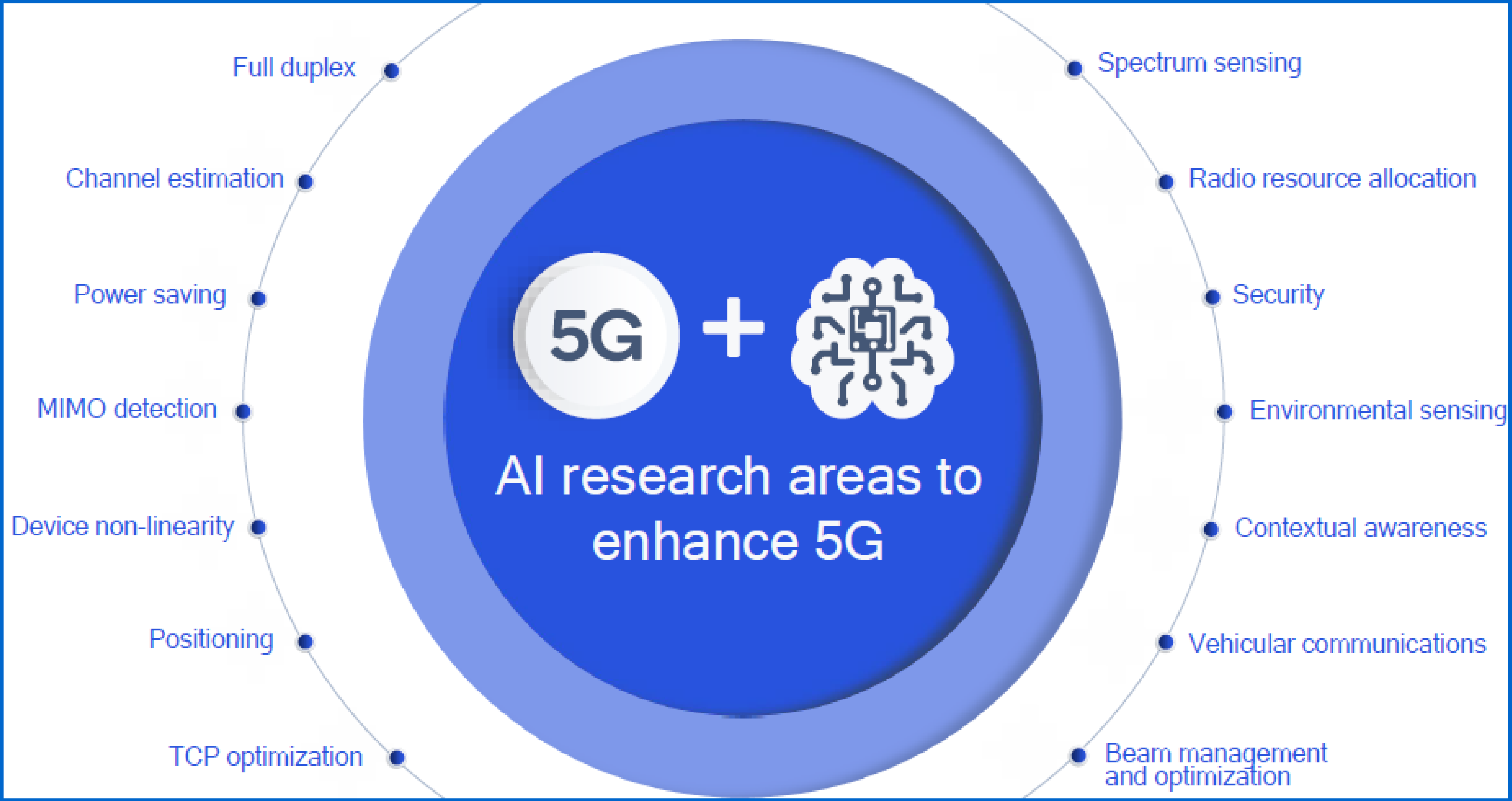


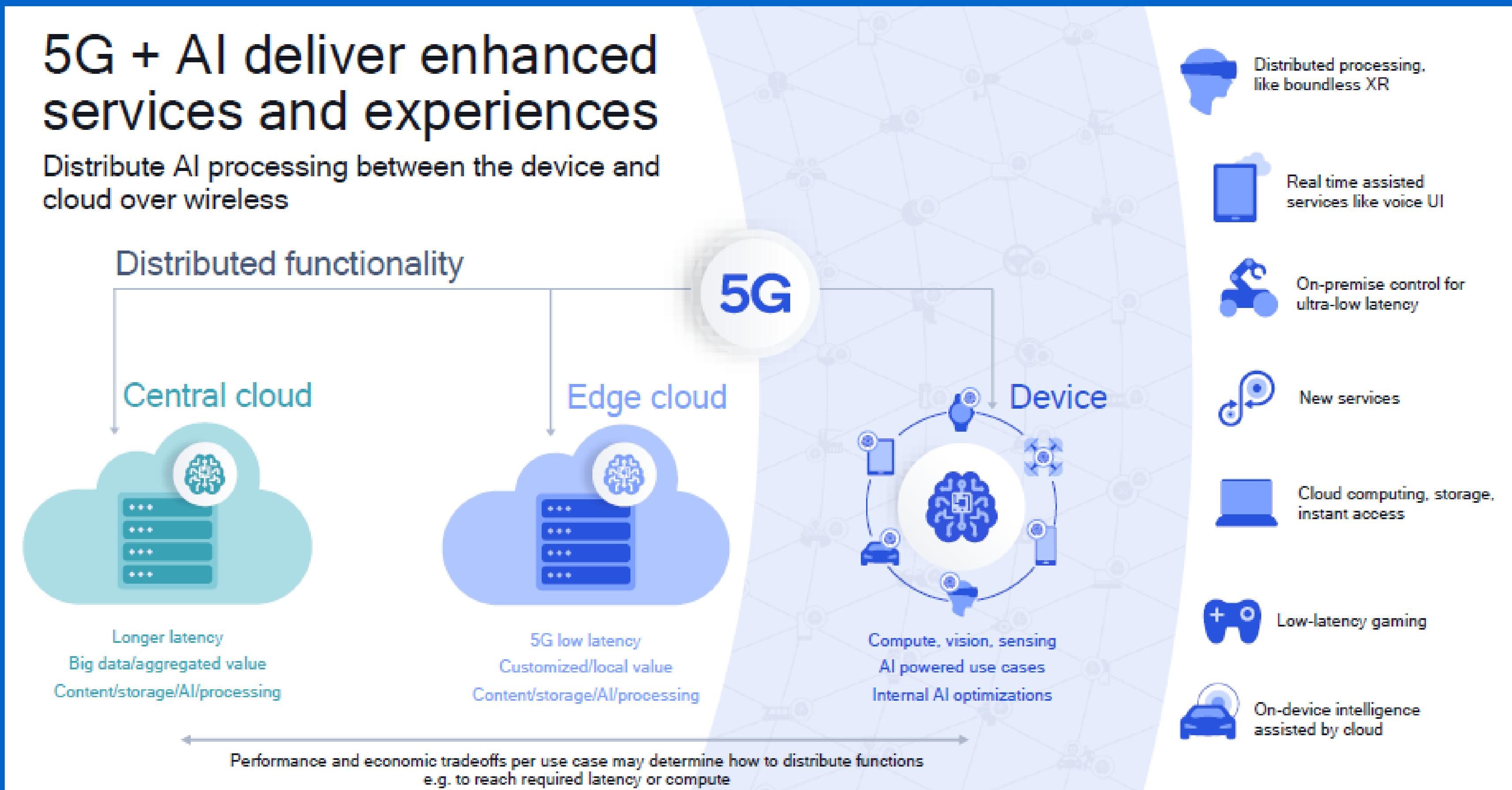
Learning device position over time without prior knowledge with RF sensing – complementing existing positioning methodologies[1]

Motion and gesture detection

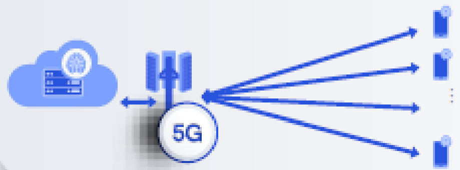


Sensing changes in environment to infer location and type of motion for a wide range of use cases (e.g., vital sign tracking, fall detection)


Full duplex
Channel estimation
Power saving
MIMO detection
Device non-linearity
Positioning
TCP optimization
5G +
AI research areas to enhance 5G
Spectrum sensing
Radio resource allocation
Security
Environmental sensing
Contextual awareness
Vehicular communications
Beam management and optimization

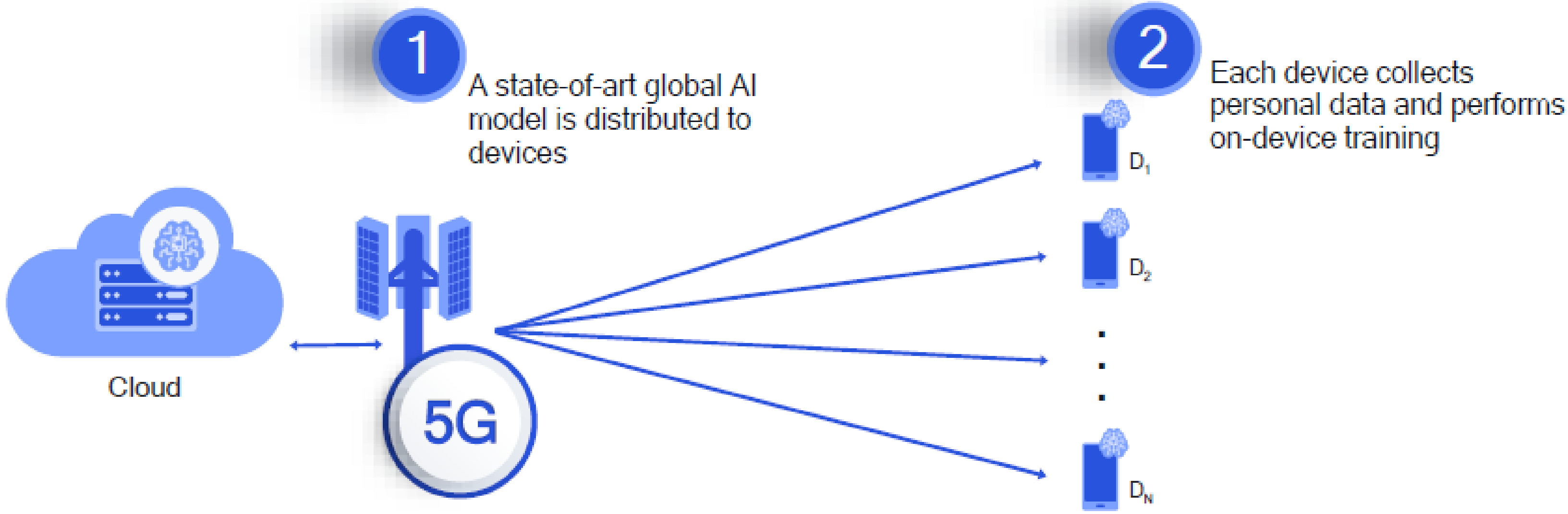

5G + AI deliver enhanced services and experiences
Distribute AI processing between the device and cloud over wireless
Distributed functionality
5G
Central cloud
Edge cloud
Device
Longer latency
Big data/aggregated value
Content/storage/AI/processing
5G low latency
Customized/local value
Content/storage/AI/processing
Compute, vision, sensing
AI powered use cases
Internal AI optimizations
Performance and economic tradeoffs per use case may determine how to distribute functions
e.g. to reach required latency or compute
Distributed processing, like boundless XR
Real time assisted services like voice UI
On-premise control for ultra-low latency
New services
Cloud computing, storage, instant access
Low-latency gaming
On-device intelligence assisted by cloud

Research directions for distributed learning over wireless



- **Optimizing the communication network**
  Compressing information sent on the uplink and downlink
- **Tackling statistical heterogeneity in data**
  Smartly combining model updates from a broad distribution
- **Privacy**
  Model parameter perturbations with privacy guarantees
- **Personalization**
  Meta learning with optimized global model starting point
- **On-device training efficiency**
  Light-weight models and training
- **Advanced topologies for distributed learning**
  Peer-to-peer, multi-cloud, and hierarchical privacy

Distributed learning over 5G is the way to scale intelligence



Scale

Processing is spread over many devices

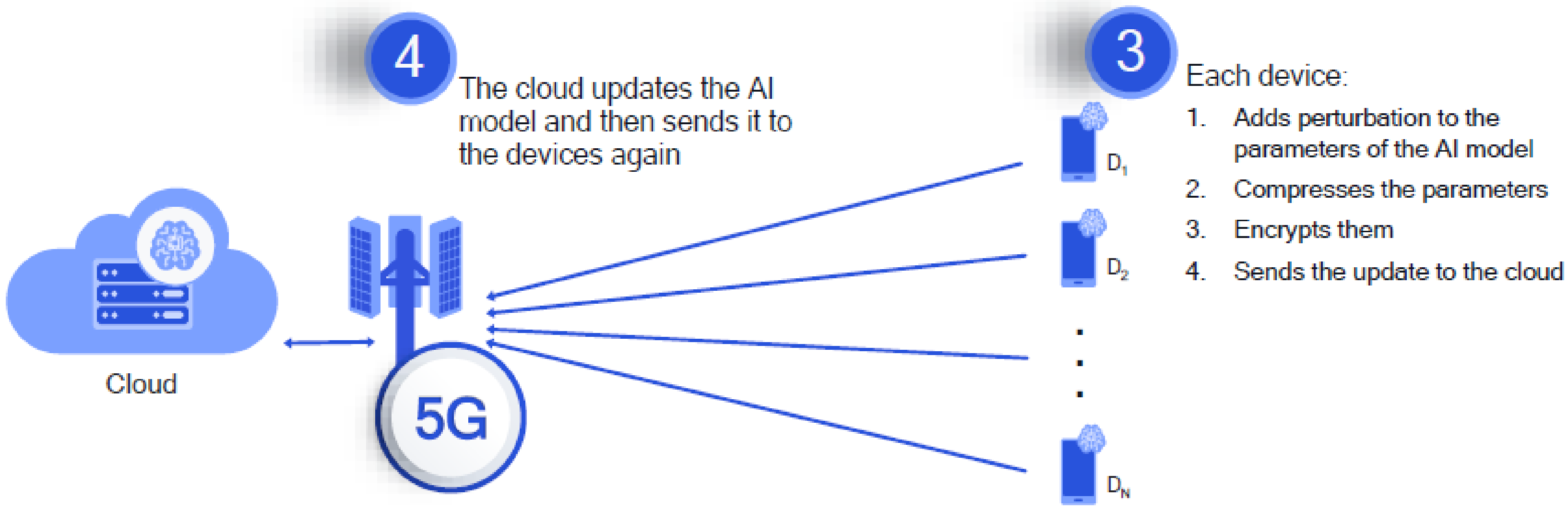
Personalization

Model customized based on your personal data

Privacy

Raw data stays on the device

Distributed learning over 5G is the way to scale intelligence



Scale
Network bandwidth is conserved

Personalization

Privacy
Only noisy and encrypted weights sent to the cloud

EXFO

Key inputs for ML-driven automated diagnostics

Passive monitoring	Active monitoring	Fiber monitoring	RAN	Topology	Telemetry
CDRs / XDRs	**QoS / QoE**	**Optical QoS**	**RAN-UE analytics**	**Dependencies**	**Infrastructure**
Customer and device-level QoE, location, services, impact and transactions.	Real-time pulse of network and service performance from core-to-edge.	Insight into how optical transmission performance impacts QoS and QoE.	Geo-analytics relate radio and RAN performance issues to QoE.	Visualize and isolate fault location using infrastructure dependencies.	Leverage infrastructure telemetry for full-stack assurance

Diverse data sources are essential to pinpoint and resolve customer-impacting events

Key AI Use Cases for 5G

Predictive Maintenance
Prevent performance issues & complex faults

Network Slicing
Dynamic Network Slicing

Network Security
DDoS Attacks and Bot prevention

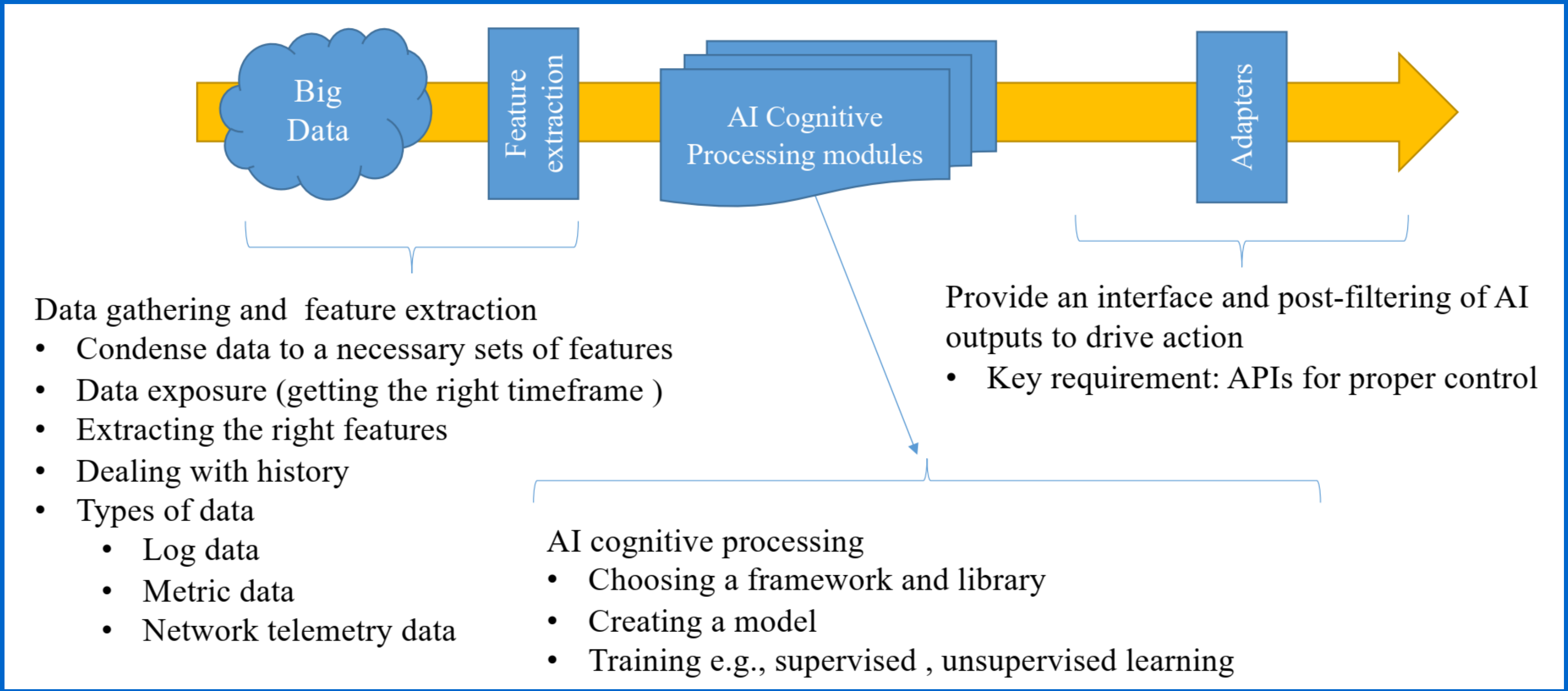
Intelligent Orchestration
Network Services scaling & Healing

NOC Automation
RCA & Guided Diagnostics

Intelligent Service Rollout
Network Planning

servicenow

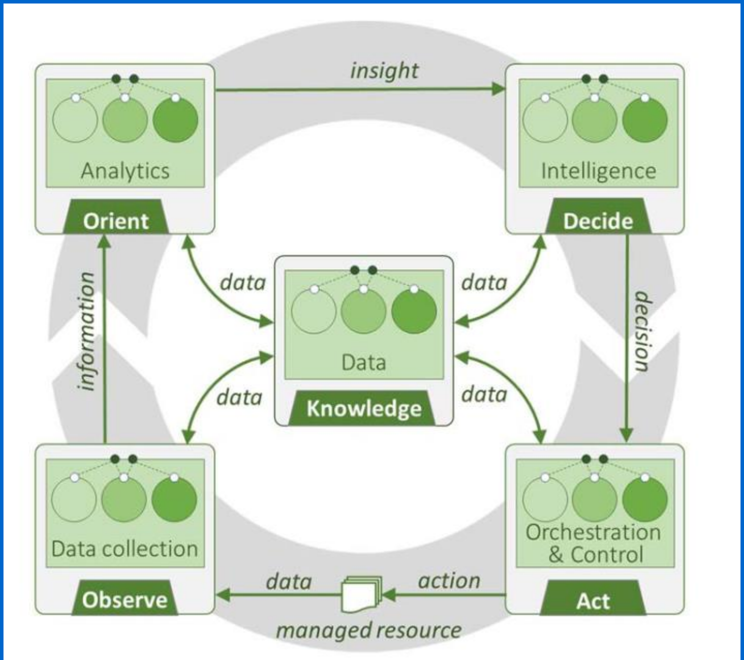
Network Based AI

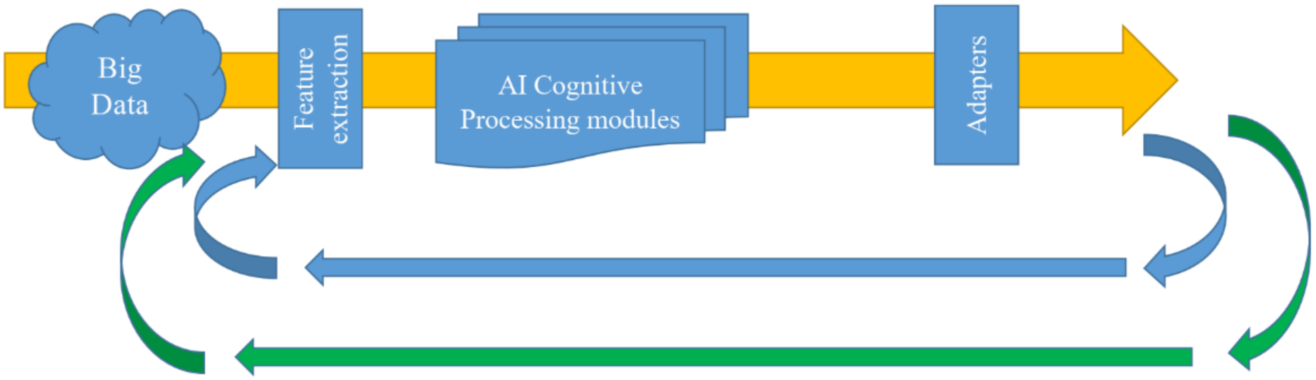


- **Appropriate APIs** may be required to apply the necessary controls
- Additionally, **orchestrators** and **policies** may be used to implement the end result

Control Loops in AI

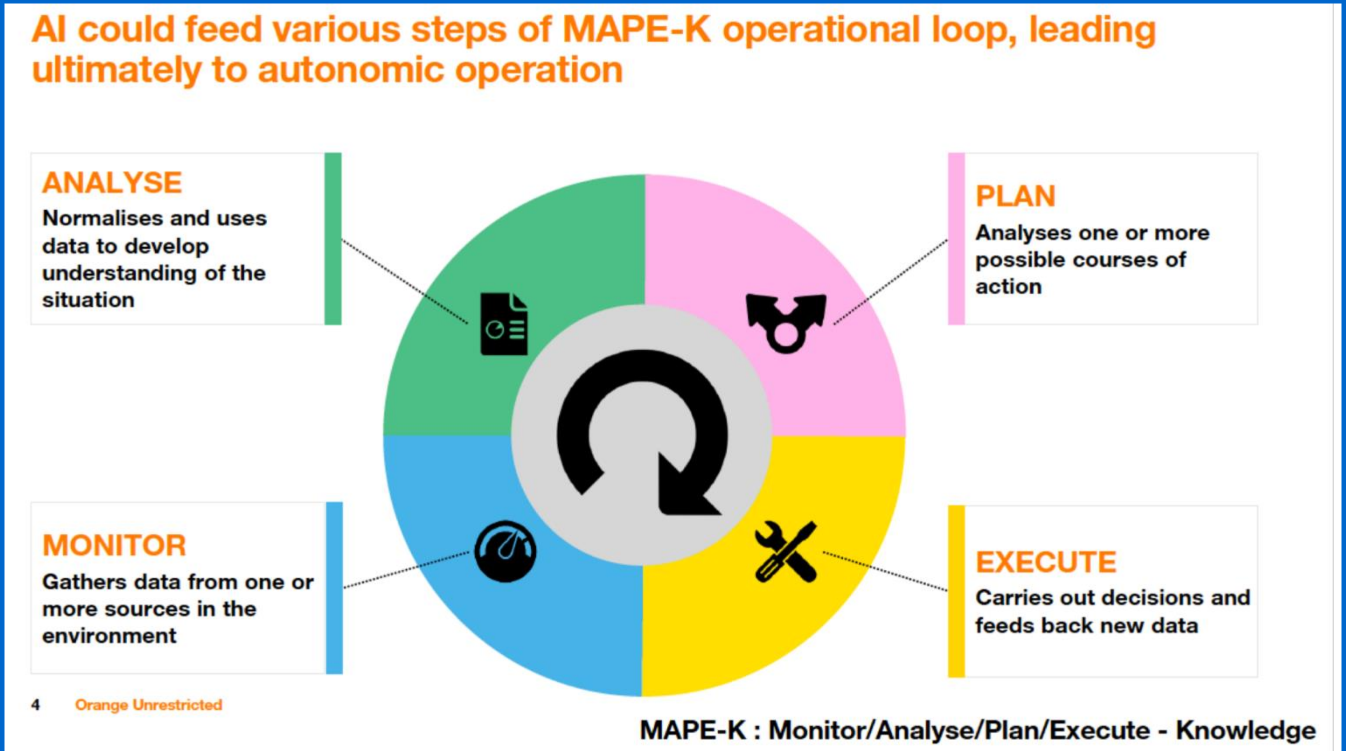
- Inner/shorter loops can operate at faster speeds, while longer outer loops need to operate at longer time constants to ensure the network has stabilized before making new changes.
- AI functions can sometimes be deployed using reinforcement learning techniques to dynamically optimize system performance.



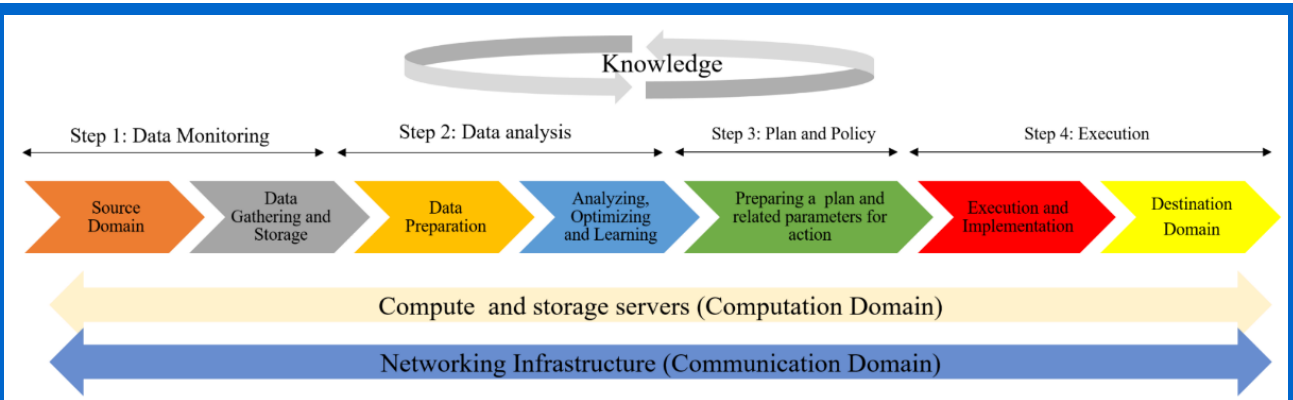


- AI network applications often deployed in closed-loop systems
- Multiple closed feedback loops may co-exist, and even be nested within each other
- Loops can interfere with each other, causing instability
- To avoid instability, special actions and policies may need to be considered.

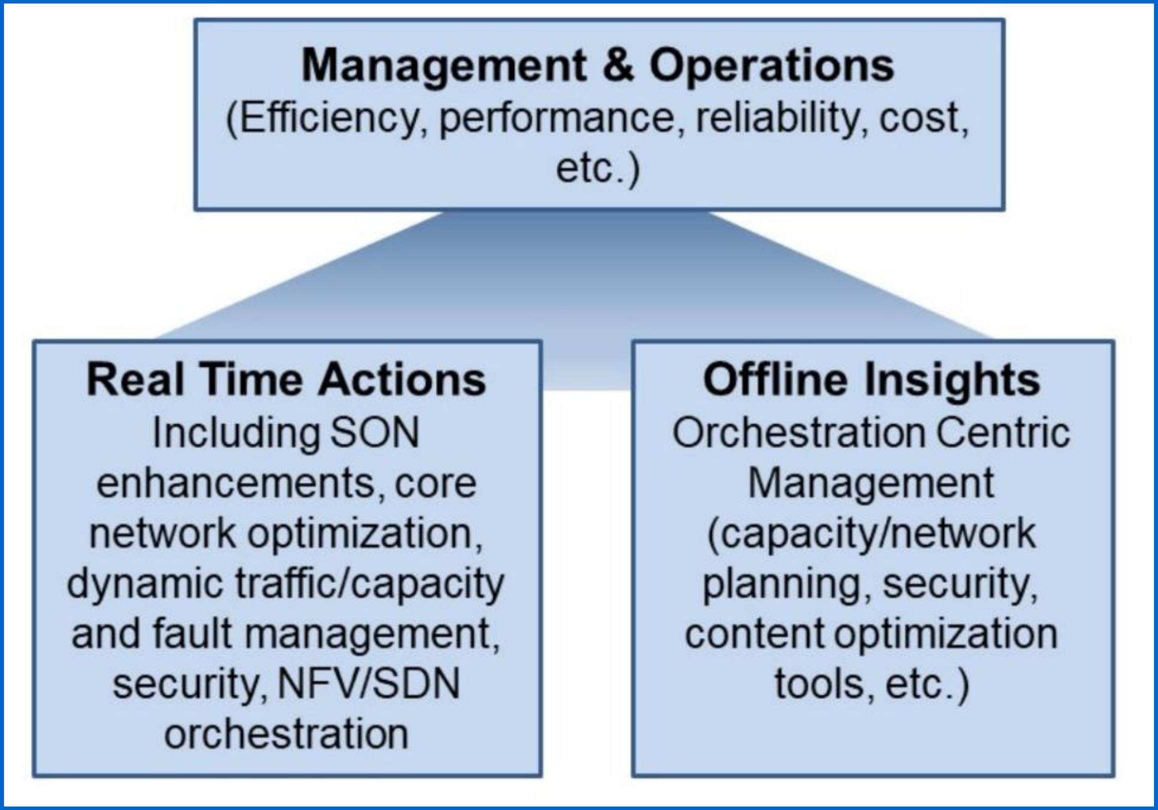
MAPE-K : Monitor/Analyse/Plan/Execute – Knowledge for Autonomic Operation



Source: Towards an AI unified platform using Acumos, OW2con'18, June 7-8, 2018, Paris



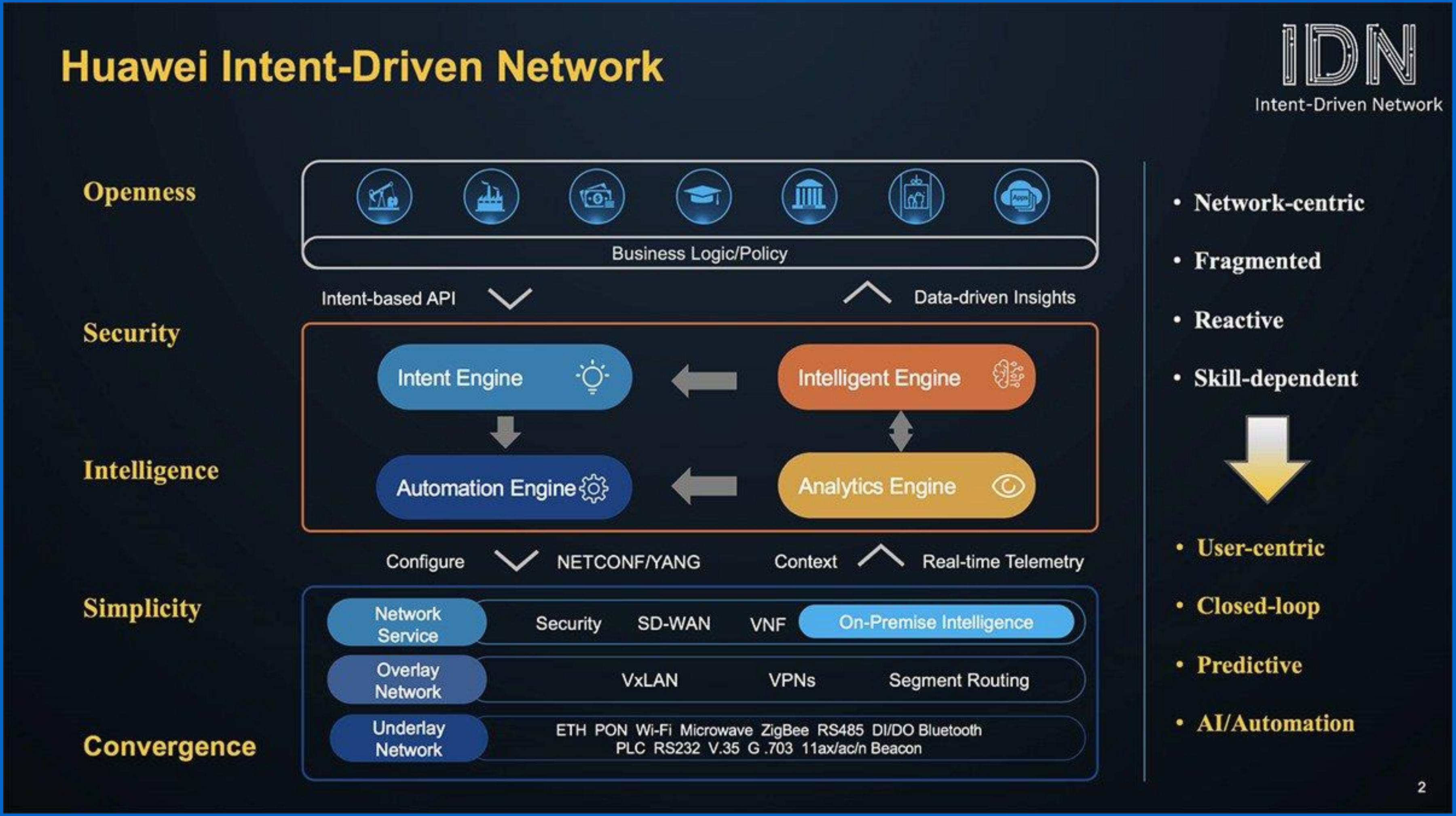
Types of Network Oriented AI



Intent-based and Zero Touch Networking (ZSM)

- Human administrators define the network's desired outcome in broad but descriptive terms
- Intent-based networking systems <u>monitor, detect and react</u> in real time to changing network conditions while automatically orchestrating new customer service deployments and configuration changes
- Actual network management and operations are done using automated systems that implement the **desired intent** of the expressed **policies**
- With intent-based networking, it is often useful to think of AI as <u>Automated Intent</u> rather than artificial intelligence

Intent-based Networking (MWC 2019, Barcelona)

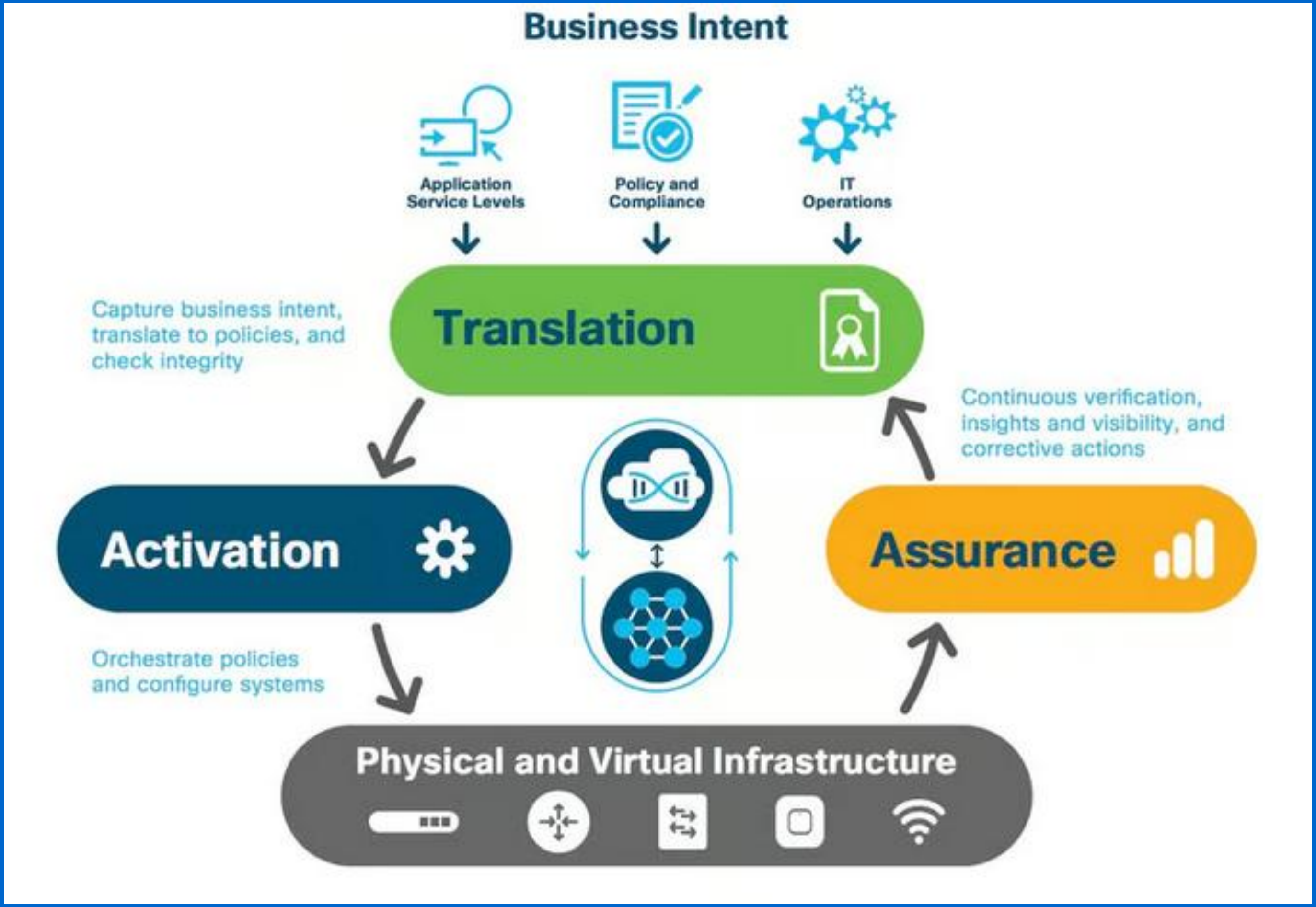


Huawei IDN: New SW blocks

- **CampusInsight 2.0.** This is Huawei's network insight engine and can be thought of as the "brains" of the solution. It collects network and application data from network devices and, in real time, turns it into streaming telemetry. Huawei then applies machine learning and artificial intelligence (AI) to it to continually monitor the network to predict problems.

- **SD-Campus 2.0.** After CampusInsight 2.0 does its magic, SD-Campus 2.0 implements the changes via a full lifecycle network automation solution for the campus network. Huawei uses a combination of software-defined network (SDN) technologies and cloud technologies to fully automate network configuration, policy management, and virtual network provisioning.

Cisco Intent-based Networking



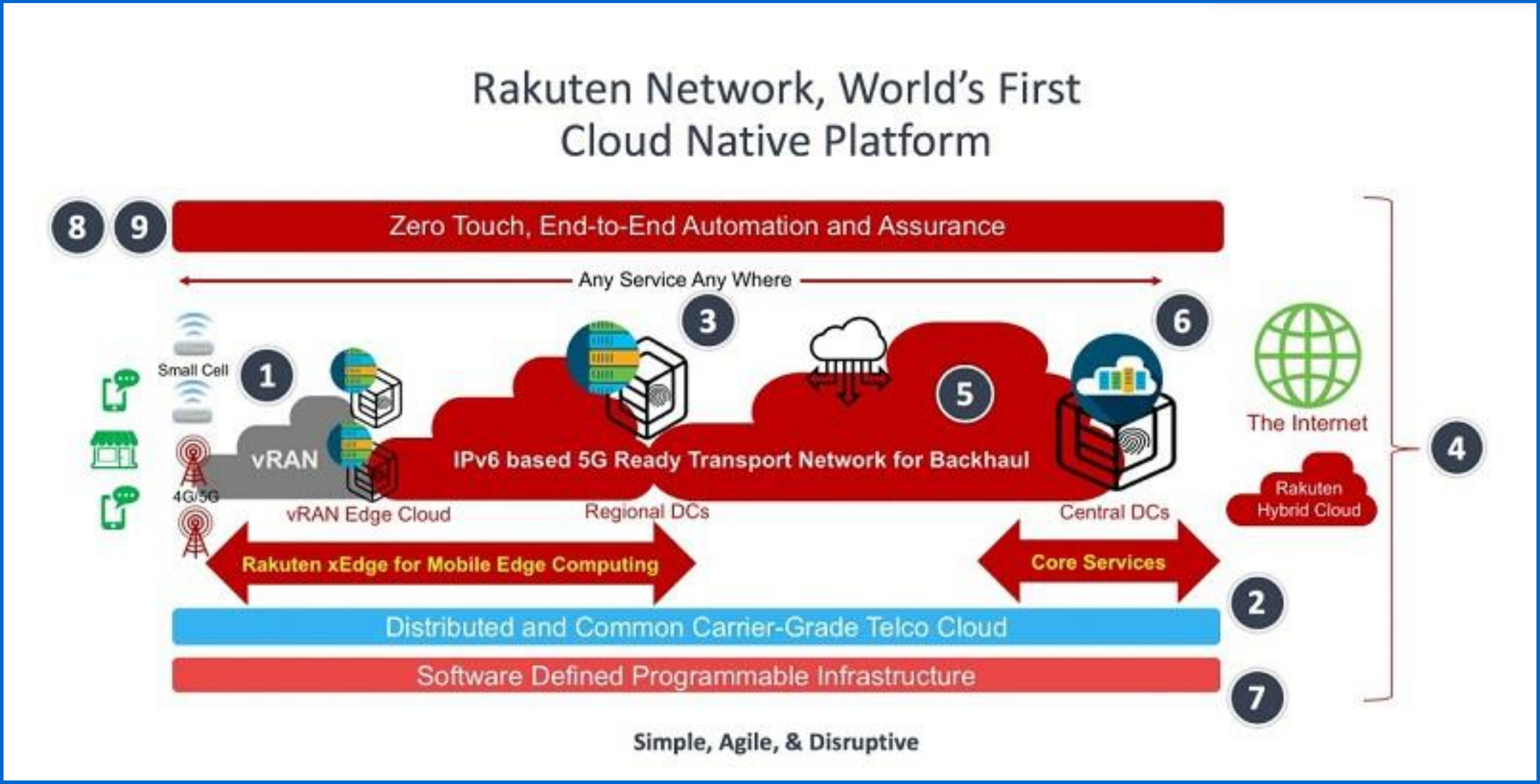
Rakuten Mobile

Partnerships for Success

Tareq Amin, EVP CTO
June 4, 2020

**World's First
Fully Virtualized
New Generation Network**

World’s first Cloud Native Operation Started in Japan 2018



Rakuten Elastic Cloud Network Architecture delivers

VIRTUALIZED
Virtualized and decoupled RAN architecture

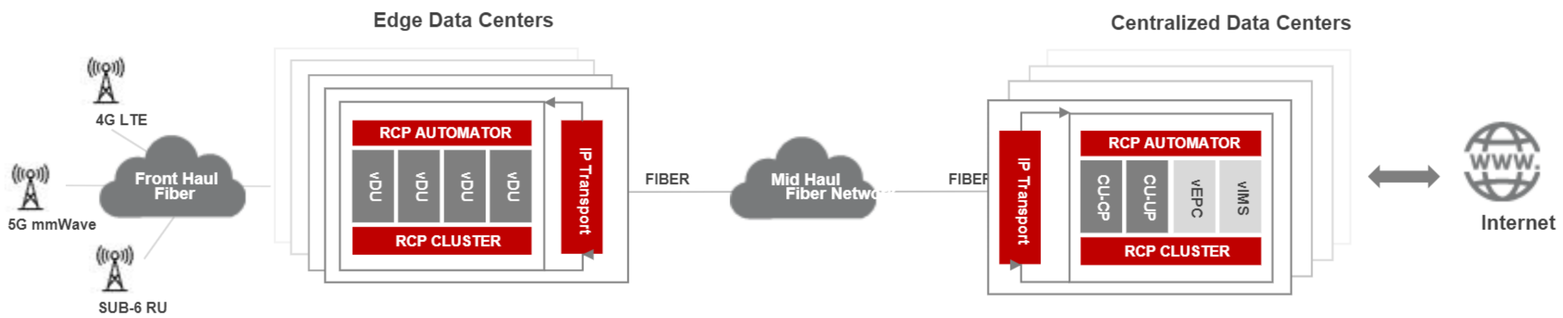
Sub-6 DU On FPGA Based Servers

AUTOMATED
Pervasive automation and zero ops control

Auto Scalable Highly Resilient Design

COTS/CUPS ARCHITECTURE
Standardized, open HW and SW frameworks

Separate Control/User Plane App For Different Traffic Type



330 Far Edge Sites	**58** Edge Sites*	**3** Central DCs	**8 Tbps** IPV6 Backbone Infrastructure

* Edge Sites as Regional Data Centers are currently being deployed.

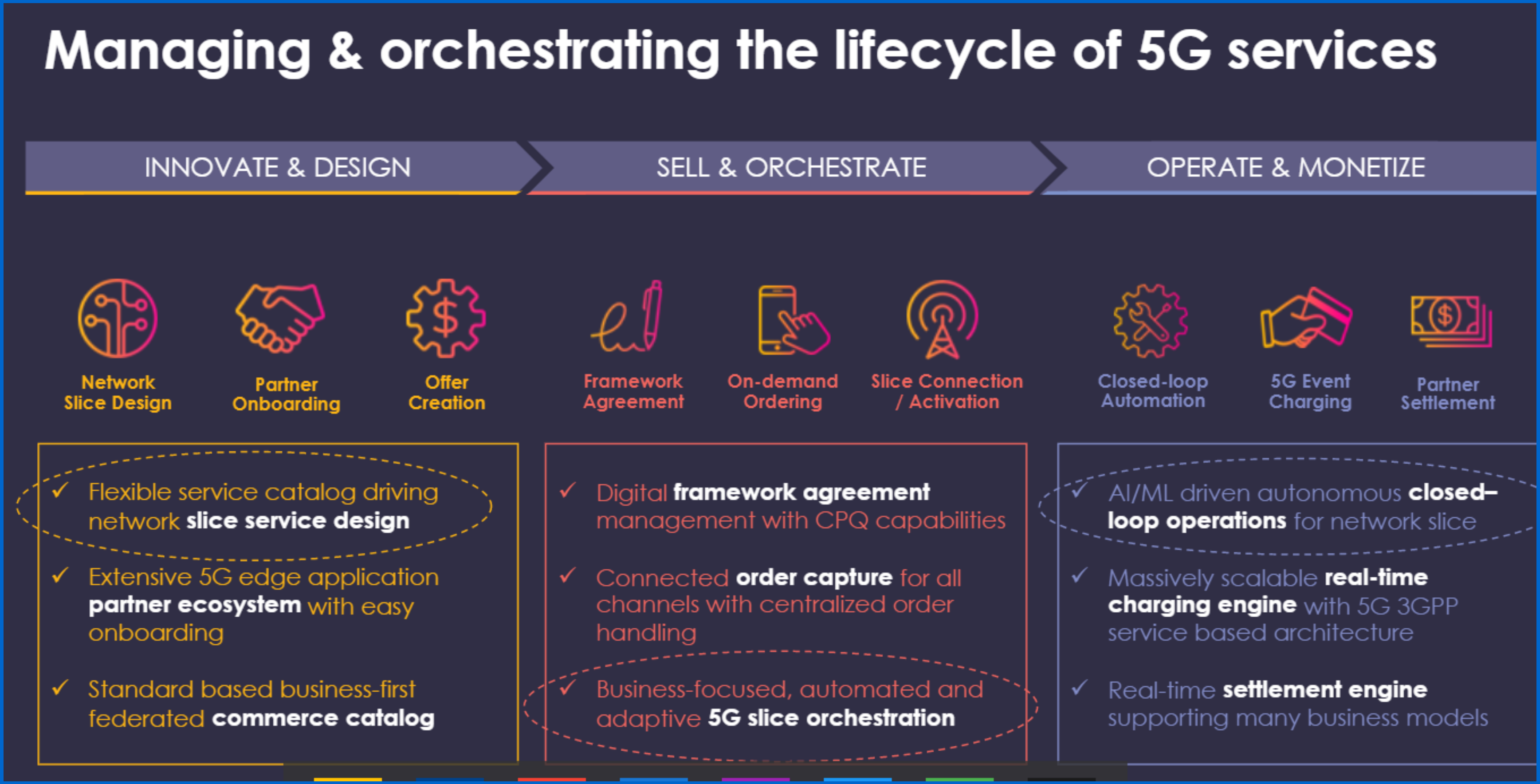
World’s First OpenRAN Compliant RAN Platform

5G Sub-6 Antenna
Massive MIMO 32x32 Antenna Element, 71 dBm EiRP

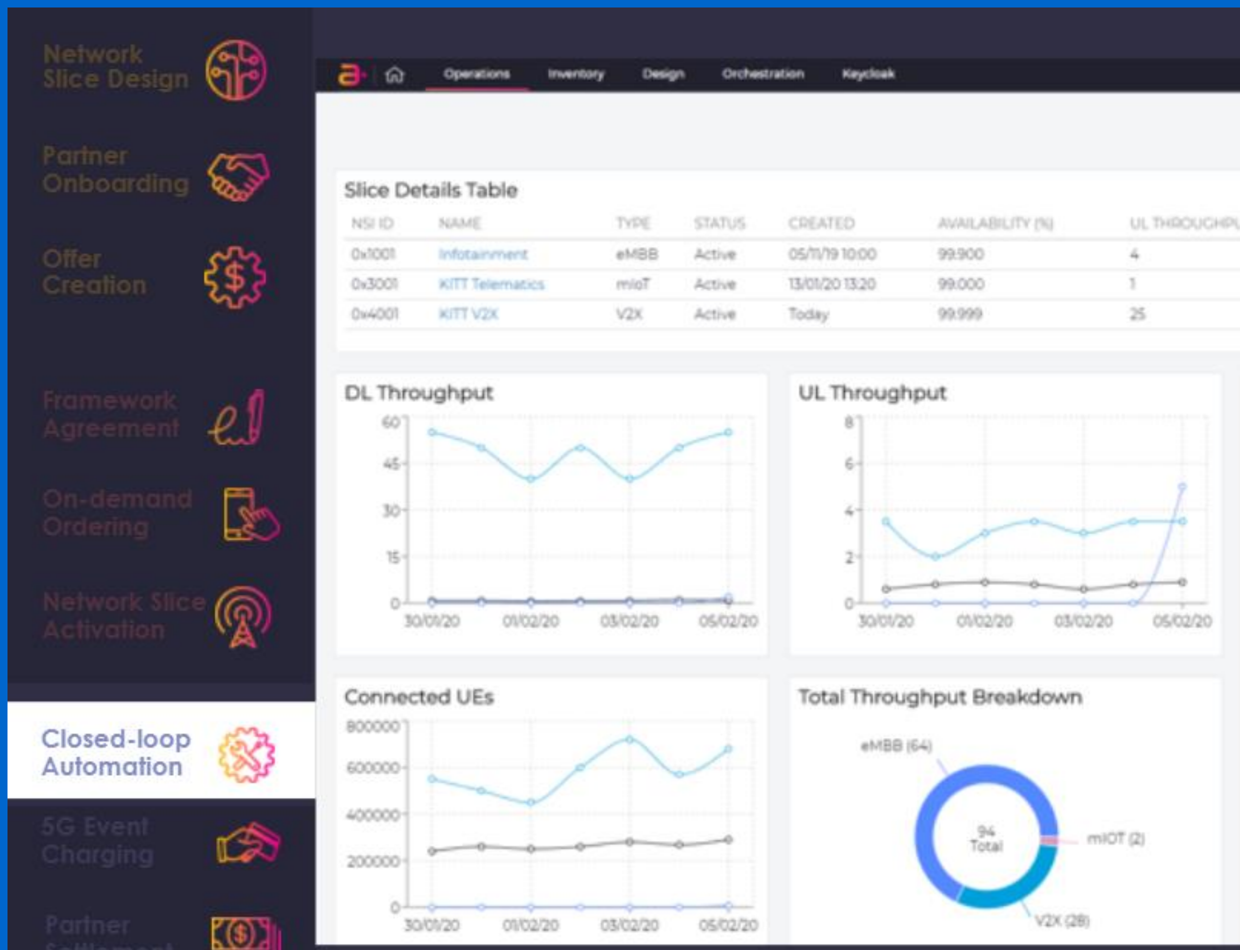
NEC Platforms Factory in Fukushima, Japan

- Partnership with NEC
- **HW Design** completed
- **Mass Manufacturing** started in Japan
- **SW is being developed** for-Sub-6 Antenna

Automation of Network Slicing



Automation of Network Slicing



Journey of Innovation:
Building a highly stable, demand elastic and future-proof telco platform

Platform

Cloud Native Telco Platform

Highly stable. Demand Elastic, Cloud Native Telco Platform

RAN

4G/5G OpenRAN Platform

4G/5G Radio Architecture on Containers designed

Operations

Automation & AI-Ops

Ops efficiency with RMOPs everyday being improved

Devices

Rakuten Casa

Fully virtualized LTE and Wi-Fi Devices deployed

Core

Cloud Native CORE

Collaboration with NEC to build OpenCore

Cloud

Multi Edge Computing

COTS based MEC Infrastructure deployed

Services

RCS Rakuten Link App

Single App for all Comm needs deployed

Devices

Rakuten Mini

Top selling Rakuten Handset in market

Open RAN [cont.]

- Open RAN is usually introduced as a "disaggregation" attempt:

"An Open Radio Access Network (ORAN) is a nonproprietary version of the Radio Access Network (RAN) system that allows interoperation between cellular network equipment provided by different vendors."

CISCO

"Open RAN is a hot topic in mobile telecoms. At its simplest, the concept is for a more open radio access network architecture than provided today. Many different claims have been made as to the potential of Open RAN to improve competition, network flexibility and cost."

ERICSSON

- Actually, Open RAN is also an enabler of intelligent radio access networking.

Open RAN [cont.]

ETSI releases first O-RAN specification

ETSI - Sophia Antipolis/France, O-RAN ALLIANCE - Bonn/Germany, 15 September 2022

ETSI and O-RAN are pleased to announce that ETSI has adopted the first O-RAN specification as ETSI TS 103 859, namely 'O-RAN Fronthaul Control, User and Synchronization Plane Specification v7.02'. The document focuses on Open Fronthaul, one of the interfaces in the O-RAN Architecture for open and intelligent Radio Access Networks. It specifies the control plane, user plane and synchronization plane protocols used over the fronthaul interface linking the O-DU (O-RAN Distributed Unit) and the O-RU (O-RAN Radio Unit) for the lower layer functional splits. The scope of TS 103 859 includes both LTE and NR (5G).

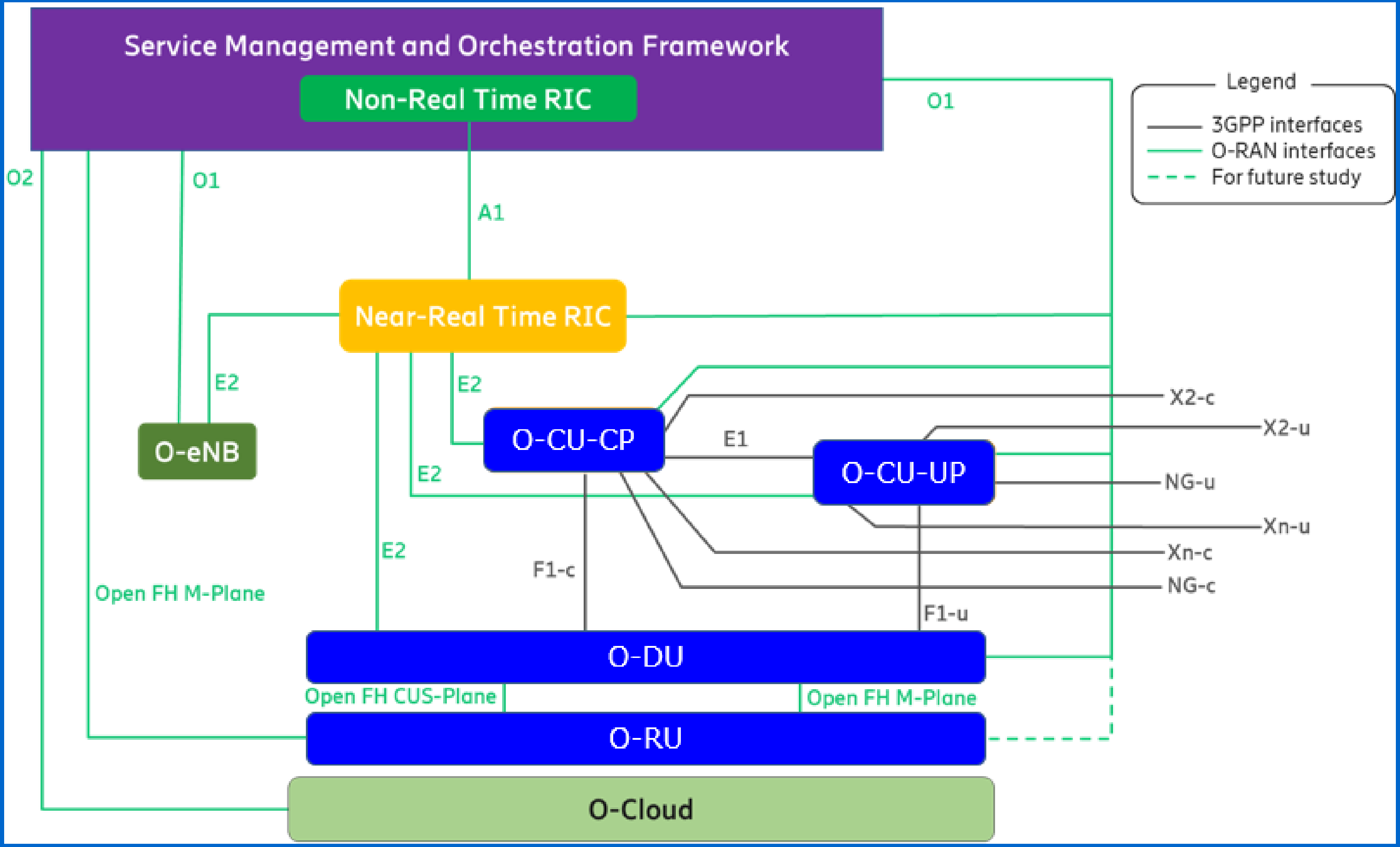
"*Recognition of O-RAN specifications by ETSI is another major step in enabling broad adoption of Open RAN,*" said Claire Chauvin, O-RAN Board member and Strategy Architecture and Standardization Director at Orange. "*Having the O-RAN specification available as an ETSI specification adds further endorsement desired by commercial and public sector entities in a range of countries.*"

"*The O-RAN specification has been approved as an ETSI specification after a thorough review and requested revisions by our experts,*" said Dominique Everaere, Chair of the ETSI Mobile Standards Group Technical Committee. "*When specifications go through the ETSI PAS process, they need to comply with the ETSI rules, and the ETSI committee in charge of these specifications works with the organization to ensure they align with existing procedures for approval as ETSI standards.*"

The ETSI Publicly Available Specification (PAS) process enables an ETSI partner to submit one or more of its PAS for adoption by ETSI. It can then become an ETSI Technical Specification (TS) or ETSI Technical Report (TR). The organization asking for this process must be a legal entity, have an IPR policy compatible with ETSI's or have exceptionally accepted to apply ETSI's IPR policy, and have signed a Cooperation Agreement.

O-RAN is preparing to submit more of its specifications to the ETSI PAS process to recognize additional parts of the O-RAN Architecture as ETSI specifications.

Open RAN



Service Management and Orchestration Framework
Non-Real Time RIC
O1
Legend
3GPP interfaces
O-RAN interfaces
For future study
O2
O1
A1
Near-Real Time RIC
E2
E2
O-eNB
O-CU-CP
E1
O-CU-UP
X2-c
X2-u
NG-u
Xn-u
Xn-c
NG-c
E2
E2
F1-c
F1-u
Open FH M-Plane
O-DU
Open FH CUS-Plane
Open FH M-Plane
O-RU
O-Cloud

Another example of Network Automation

Operators' network automation strategies

What is still missing?
April 2020

Adaora Okeleke
Senior Analyst, Service Provider & Comms
Adaora.okeleke@omdia.com

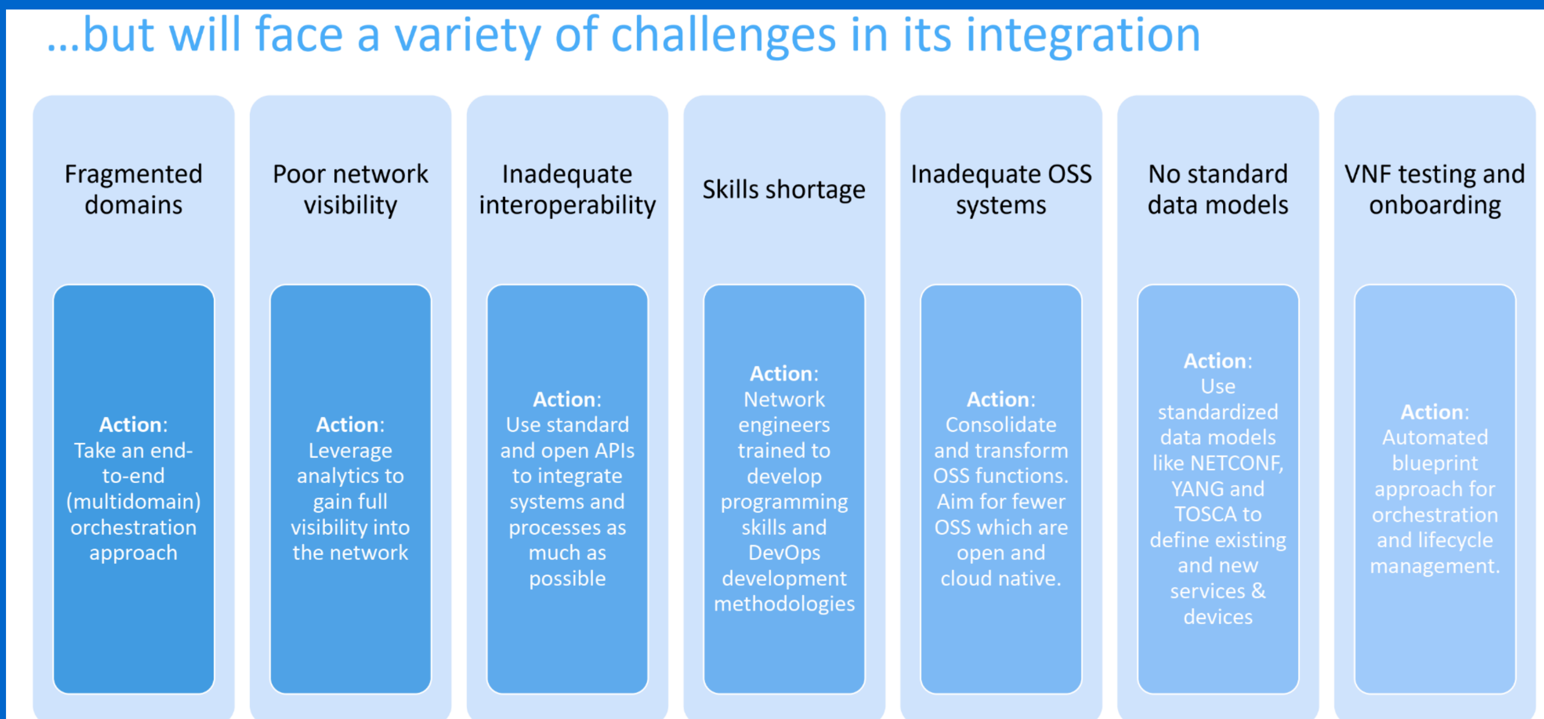
Stephanie Gibbons
Principal Analyst, Service Provider & Comms
Stephanie.gibbons@omdia.com

Brought to you by Informa Tech

OMDIA

Network Automation

- Operators will benefit from automation:
  - End-to-End monitoring in real time
  - Networks and Services Agility
  - Cost savings and new revenues from network optimization



Network programmability

- SDN/NFV rollout
- Standardized data models and protocols
- Open APIs for interoperability, ease of scale

Big data, analytics & AI

- Closed loop automation
- Real-time telemetry of hybrid network
- Predictive and prescriptive analytics from AI for self-healing networks

Build DevOps skills

- Network automation drives IT and operations changes
- Network engineers' reliance on software
- DevOps skills and acceptance is critical

vMWC announcements emphasize the role of AI and cloud automation

Telco cloud automation

vmware

VMWare Telco Cloud Automation

Accelerates network agility for telco multi-cloud environment

IBM

IBM Telco Network Cloud Manager

Simplifies and optimizes .end to end network transformation

amdocs

Amdocs Network and Service Automation Solution

Streamlines operations across a hybrid cloud networks

Operation specific automation

EXFO

Nova Adaptive Service Assurance

Automates service and network monitoring across domains; predicts, reports and identifies root causes of service impacting degradations

mycomosi

New AI/ML solution for predictive service assurance

Implements intelligence from AI to improve service and network performance

SANDVINE

Cloud native ActiveLogic software

Utilizes ML to based application classification engine to automate traffic monitoring

Automating Network Operations

ERICSSON

Network Intelligence

Applies insights from ML models to deliver pre-emptive support services to CSPs

HUAWEI

Huawei Intelligent 5G Operations Solution

Uses centralized network monitoring, anomaly detection and root cause analysis to automate network operations and maintenance functions

NOKIA

Nokia AVA 5G Cognitive Operations

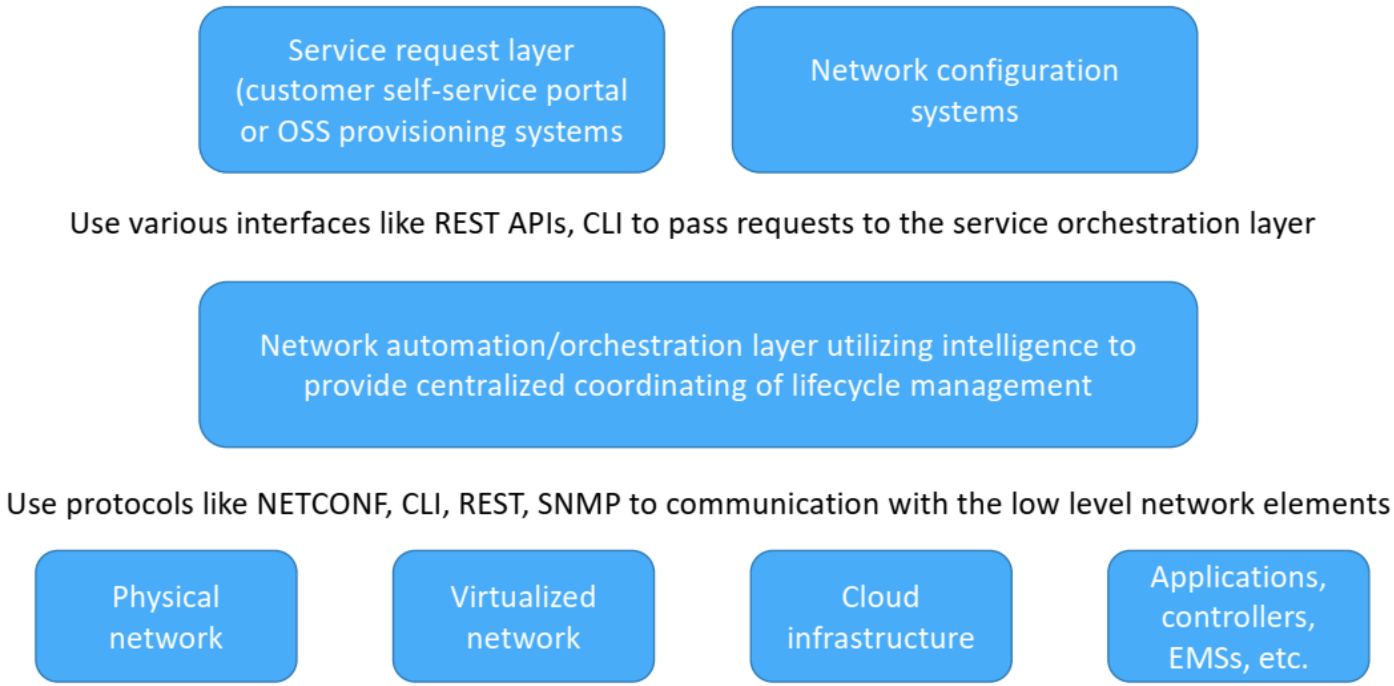
Enables CSPs to transform and manage their networks using AI and automation

CISCO

Cisco Network Insights

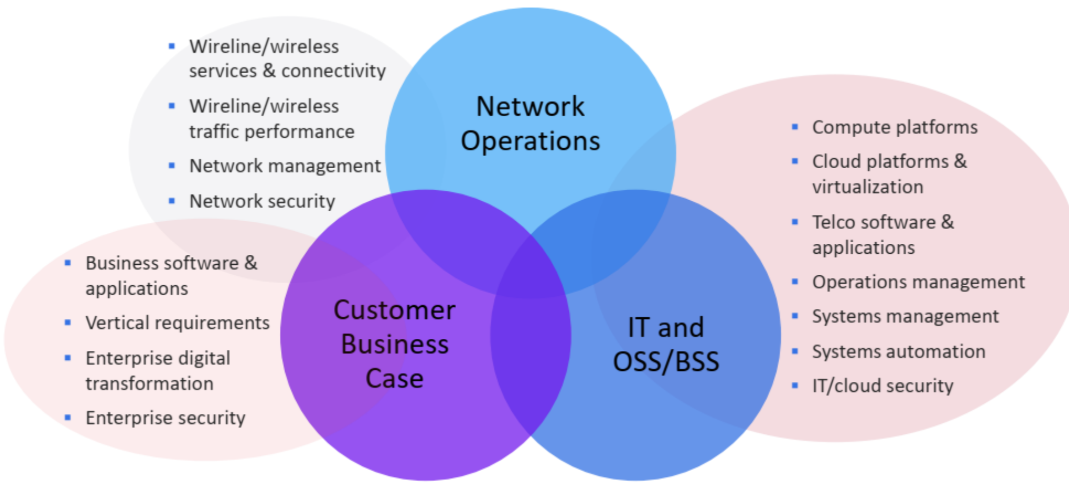
New network assurance capabilities added to data center portfolio for proactive networking operations proactive

Multi-domain service orchestration is critical to make this all work



What’s missing? Talent shortage can slow network automation

Fields of Expertise for the New Telco



What’s missing? Addressing data, interoperability and organizational issues

INTEROPERABILITY

Create open interfaces and APIs to ease interactions within a multi-vendor multi-technology network environment

DATA

Data quality issues need to be addressed to ensure that accurate insights are used to drive network automation

ORGANIZATIONAL CHANGE

Adopt DevOps and agile operational practices to support automated network environment

Automating Multi-Vendor Networks: A Practical Approach in Telefonica's Network

Fernando Muñoz Macaya, Senior Manager, Systems Engineering, Global Service Provider Sales, Cisco Systems

May 27, 2020

A Partnership

Telefónica + cisco

Solutions | Turnkey | Integrated | Customized | Multi Domain

Products

Open API | Extensible |
Cloud Delivered |
Multi Vendor

CX Lifecycle Services

Advisory | Design & Implement
Optimization | Support | Integration

Cisco Crosswork Automation Platform

intent based

proactive resolutions

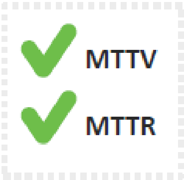
Multi vendor

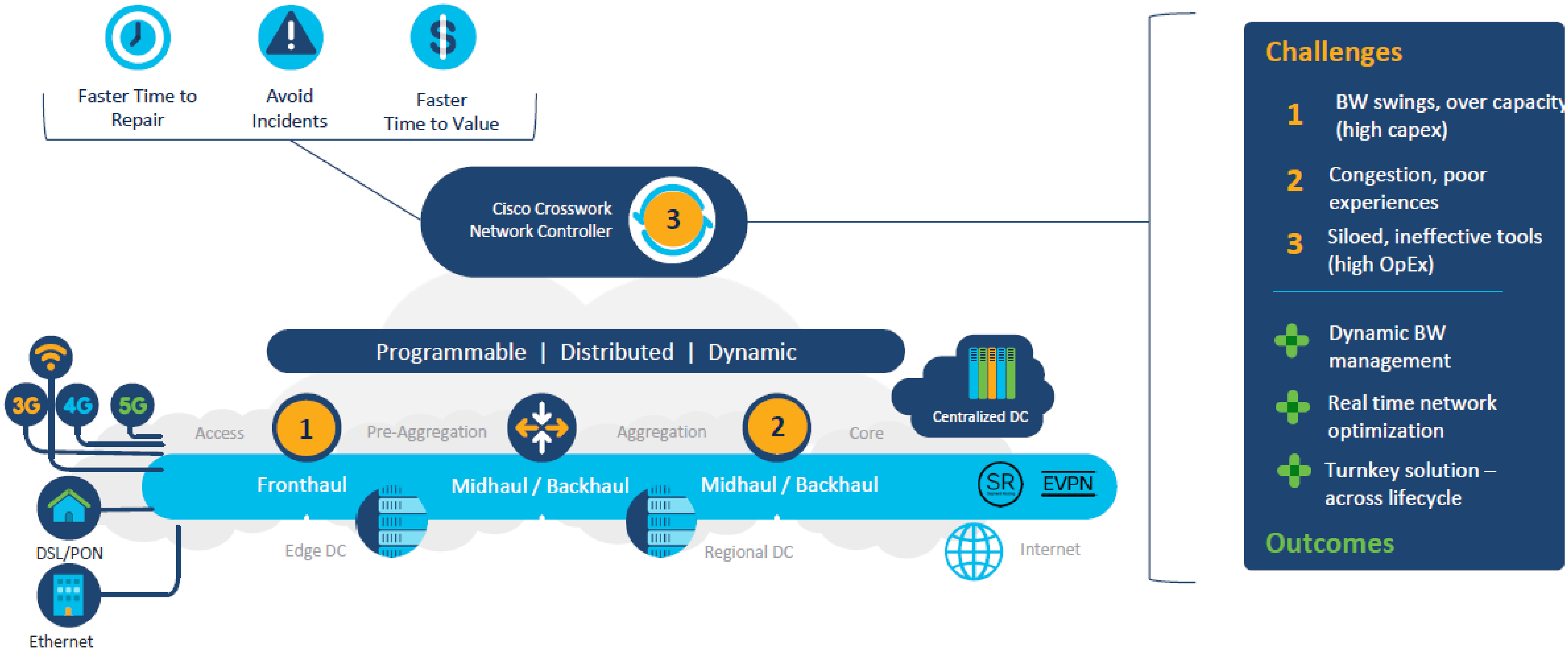
machine learning

closed loop

Use Case: Converged SDN Transport Automation

Simplify Operational Lifecycle





Some Network AI Use Cases

- Congestion-aware fairness
- Network-aware scheduling of content
- Load- and policy-aware multi-RAN selection
- Outage alerting, avoidance and reporting
- Network-wide security, anomaly and intrusion detection
- NFV automated network growth/degrowth
- Inter-datacenter congestion mitigation
- Dynamically inspect traffic and predict network performance
- AI-Based Content Processing and Management
- …

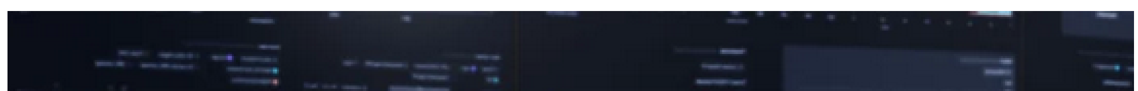
AI for Root Cause Analysis

HOME › ERICSSON BLOG › CAN AI SPEED UP THE ROOT CAUSE ANALYSIS OF NETWORK SECURITY INCIDENTS?

Can AI speed up the root cause analysis of network security incidents?

When a security incident happens in the network, it's critical to nail down the problem before any real damage can be done. However, today's distributed networks, with multiple virtual levels, can make finding the root source acutely difficult. But can AI hold the key? Below, in our latest post from Ericsson Research, we get to the very root of the issue.

SEP 27, 2022 | 5 min.



Causal Inference in Telecommunications [cont.]

HOME › ERICSSON BLOG › FROM HOW TO WHY: AN OVERVIEW OF CAUSAL INFERENCE IN MACHINE LEARNING

From how to why: An overview of causal inference in machine learning

Artificial intelligence is good at predicting outcomes, but how do we go one step further? Here, we discuss how AI can use causal inference and machine learning to measure the effects of multiple variables – and why it's important for technological progression.

FEB 06, 2020 | 6 min.

Ravi Pandya

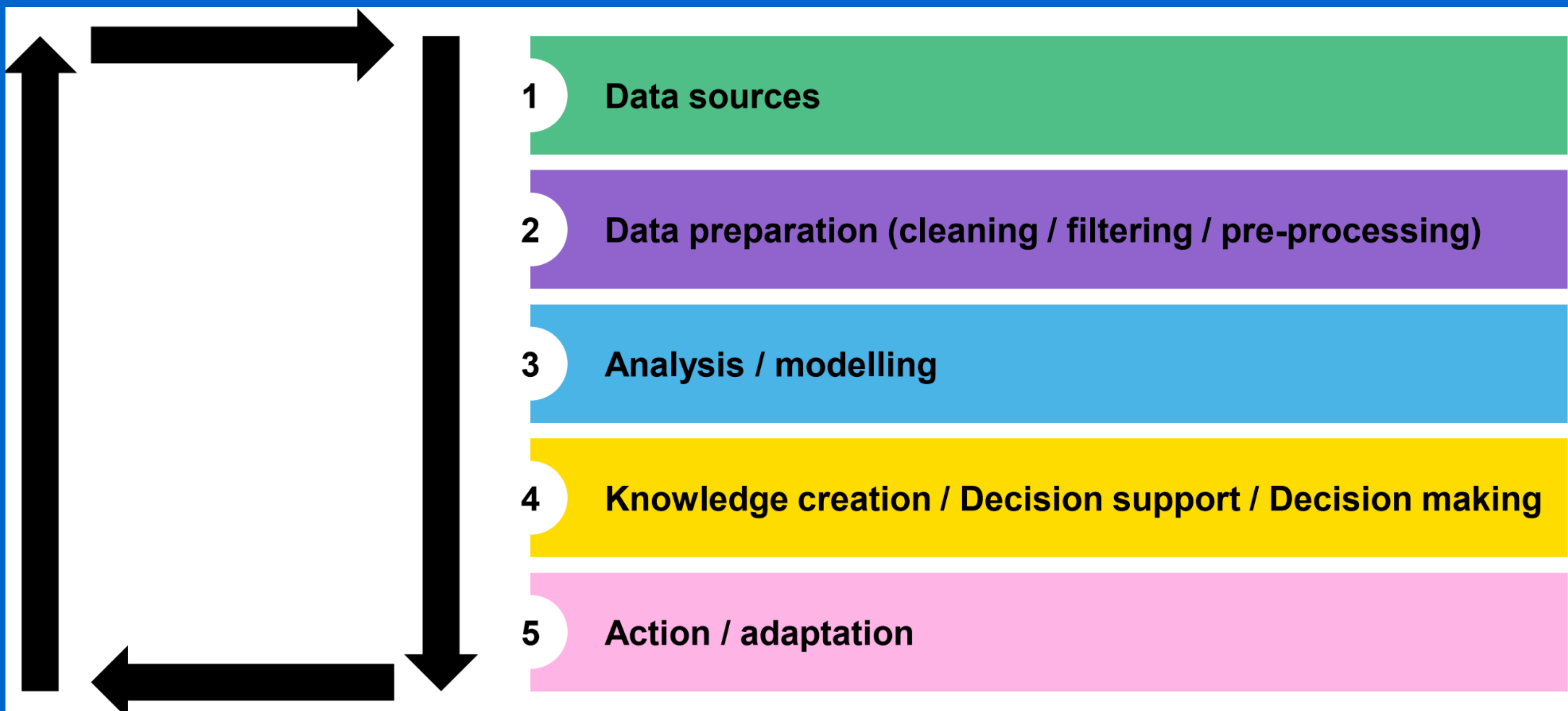
Causal Inference in Telecommunications

- From the Ericsson article:

  “In a major operator’s network control center complaints are flooding in. The network is down across a large US city; calls are getting dropped and critical infrastructure is slow to respond. Pulling up the system’s event history, the manager sees that new 5G towers were installed in the affected area today.
  Did installing those towers cause the outage, or was it merely a coincidence? In circumstances such as these, being able to answer this question accurately is crucial for Ericsson.
  Most machine learning-based data science focuses on predicting outcomes, not understanding causality. However, some of the biggest names in the field agree it’s important to start incorporating causality into our AI and machine learning systems.”

Data: The fuel for AI!

No “good” data: AI can not help!



Data Collection

- Network data can be categorized along four basic dimensions:
  - traffic-based attributes
  - network/subscriber state
  - topology/location
  - time/history
- Network performance monitoring system collects **metrics** such as latency, packet loss and throughput on a per-link or per-path basis for every 10-15 minutes, representing huge amount of data to be analyzed.
- Packet-traffic-based attributes include raw packet router/switch or link-specific metrics such as throughput, packet loss, latency, packet length (short or long), burstiness and queue fill

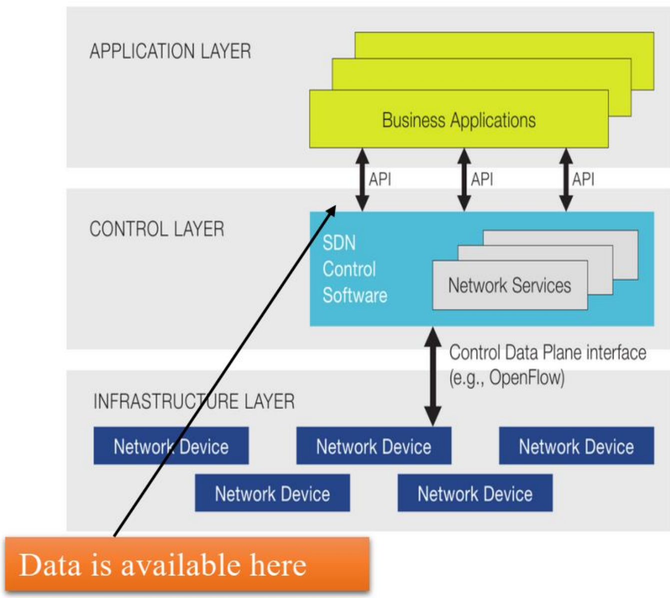
Where to collect data?

- Data may be collected from network elements such as edge routers and gateway elements, control and operations systems, as well as application specific function to understand network topology and potentially the context of the data.
- Collect data from active probes, traffic terminated internal to the network (e.g., DNS traffic) and other identifying control information.
- More information may be available through deeper packet header inspection, but widespread use of TLS/HTTPS on transport flows limits the utility of this approach

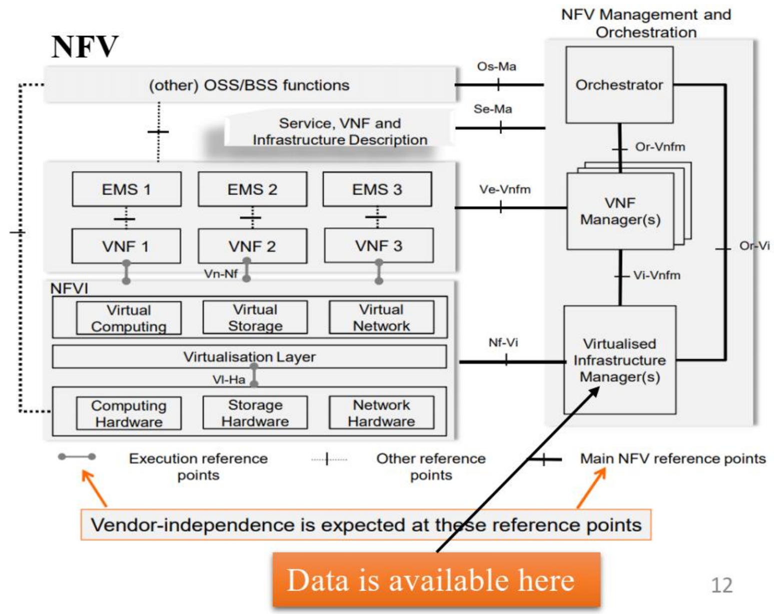
Show me the Data!

- SDN and NFV provide APIs to access data and centralized controls and orchestrators to provide the network (learning control loop).
- Therefore, two modules of data gathering and action implementation are ready.

SDN



NFV

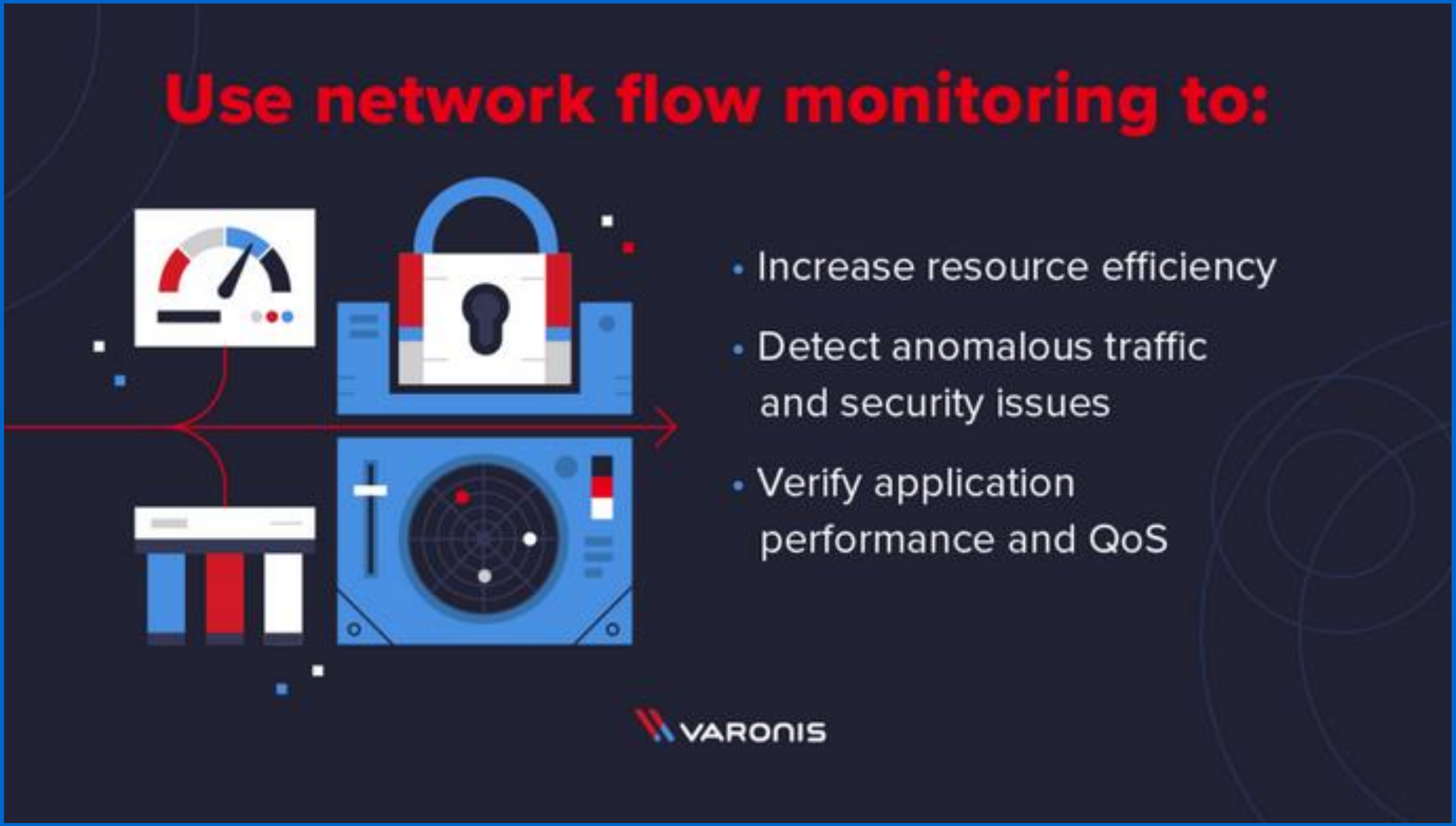


12

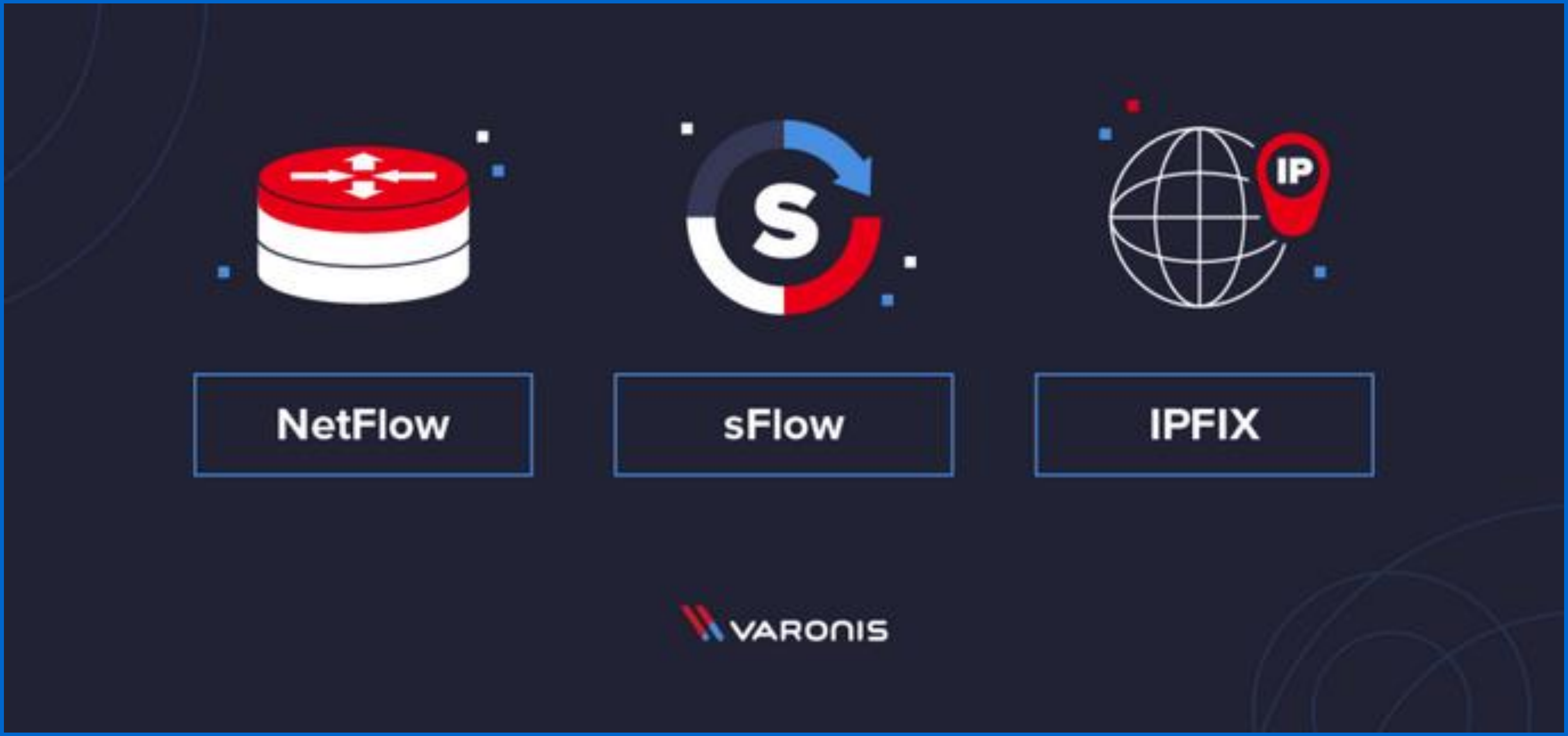
- Dynamic network configuration based on SDN-controlled traffic routing

Network Flow Monitoring

- The traffic traversing a data network should be collected and analyzed for network management …



Some Network Flow Monitoring Solutions



NetFlow [cont.]

- A feature introduced around 1996 for Cisco routers, NetFlow is a protocol for IP traffic flow monitoring.
- In the context of NetFlow, a “flow” is defined as a unidirectional stream of packets which have some values in common.
- For instance, in NetFlow v5 a flow is defined by:
  - Ingress interface
  - Src IP
  - Dst IP
  - IP Protocol Number
  - Src UDP/TCP port; 0 if other transport protocols
  - Dst UDP/TCP port, type and code for ICMP; 0 if other protocols
  - IP ToS

NetFlow [cont.]

- NetFlow operates based on 3 components:
  - NetFlow **exporter**: a NetFlow-enabled network device runs an exporter which collects flow information. Exporter generates records using indicators extracted from the aggregation of packets pertaining to a flow. Records are periodically communicated to collector(s) via UDP. For datagram structure in NetFlow v9 check https://www.cisco.com/en/US/technologies/tk648/tk362/technologies_white_paper09186a00800a3db9.html
  - NetFlow **collector**: receives records communicated by the exporter(s), performs pre-processing and stores the data
  - NetFlow **analyzer**: a tool that analyzes the records received and stored by collector(s) in order to extract insights into network traffic, bandwidth consumption, security, etc.

NetFlow

- Other proprietary flow monitoring protocols (Cisco NetFlow counterparts):
  - Jflow for Juniper Networks
  - NetStream for Huawei
  - AppFlow for Citrix

Internet Protocol Flow Information Export (IPFIX) [cont.]

[RFC Home] [TEXT|PDF|HTML] [Tracker] [IPR] [Info page]

INFORMATIONAL

Network Working Group
Request for Comments: 3917
Category: Informational

J. Quittek
NEC Europe Ltd.
T. Zseby
Fraunhofer FOKUS
B. Claise
Cisco Systems
S. Zander
Swinburne University

October 2004

Requirements for IP Flow Information Export (IPFIX)

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This memo defines requirements for the export of measured IP flow information out of routers, traffic measurement probes, and middleboxes.

Table of Contents

Internet Protocol Flow Information Export (IPFIX) [cont.]

- IETF standard for network flow monitoring
- Based on NetFlow v9
- In contrast with the proprietary NetFlow, IPFIX is widely supported by many vendors.
- IPFIX uses a similar definition of “flow” to NetFlow

Internet Protocol Flow Information Export (IPFIX) [cont.]

- IPFIX uses a particular terminology:
  - **Observation Point**: a point where IP packets can be observed; e.g. router interfaces, firewalls, etc.
  - **Metering Process**: responsible for generating flow records (this functionality is assigned to the exporter in NetFlow)
  - **Exporter Process**: responsible for sending records to the collecting process(es)
  - **Collecting process**: receives flow records from one or more exporter process(es)

Internet Protocol Flow Information Export (IPFIX)

```
   packet header capturing
             |
        timestamping
             |
             v
      +----->+
      |      |
      |  classifying
      |      |
      +------+
             |
  maintaining flow records
             |
             v

 Figure 1: Functions of the metering process
```

(from RFC 3917 memo)

```
All examples are composed of one or more of the following elements:
observation point (O), metering process (M), exporting process (E),
and collecting process (C).  The observation points shown in the
figure are always the most fine-granular ones supported by the
respective device.

      +---+     +-----+    +---------+    +---------+
      | E-+->   |  E--+->  |    E----+->  <-+--E   E--+->
      | | |     |  |  |    |   / \   |    |  |    |  |
      | M |     |  M  |    |  M   M  |    |  M    M  |
      | | |     | /|\ |    | /|\ /|\ |    | /|\  /|\ |
      | O |     | OOO |    | OOO OOO |    | OOO  OOO |
      +---+     +-----+    +---------+    +---------+
      Probe      Basic      Complex        Multiple
                 Router     Router         Exporting
                                           Processes

    +---+     +---+    +---+
    | E-+->   | E-+->  | E-+-------------->---+
    | | |     | | |    | | | +---+         ++-----+
    +-+-+     | M |    | M | | E-+------->-+-C-M-E-+->
      |       | | |    | | | | | | +---+   ++-----+
    +-+-+     +-+-+    | O | | M | | E-+->---+
    | | |       |      +---+ | | | | | |
    | M |     +-+-+          | O | | M |
    | | |     | | |          +---+ | | |       +-----+
    | O |     | O |                | O |     ->-+-C-E-+->
    +---+     +---+                +---+       +-----+

   Protocol  Remote           Concentrator     Proxy
   Converter Observation

                Figure 2: IPFIX-related Devices
```

(from RFC 3917 memo)

sFlow

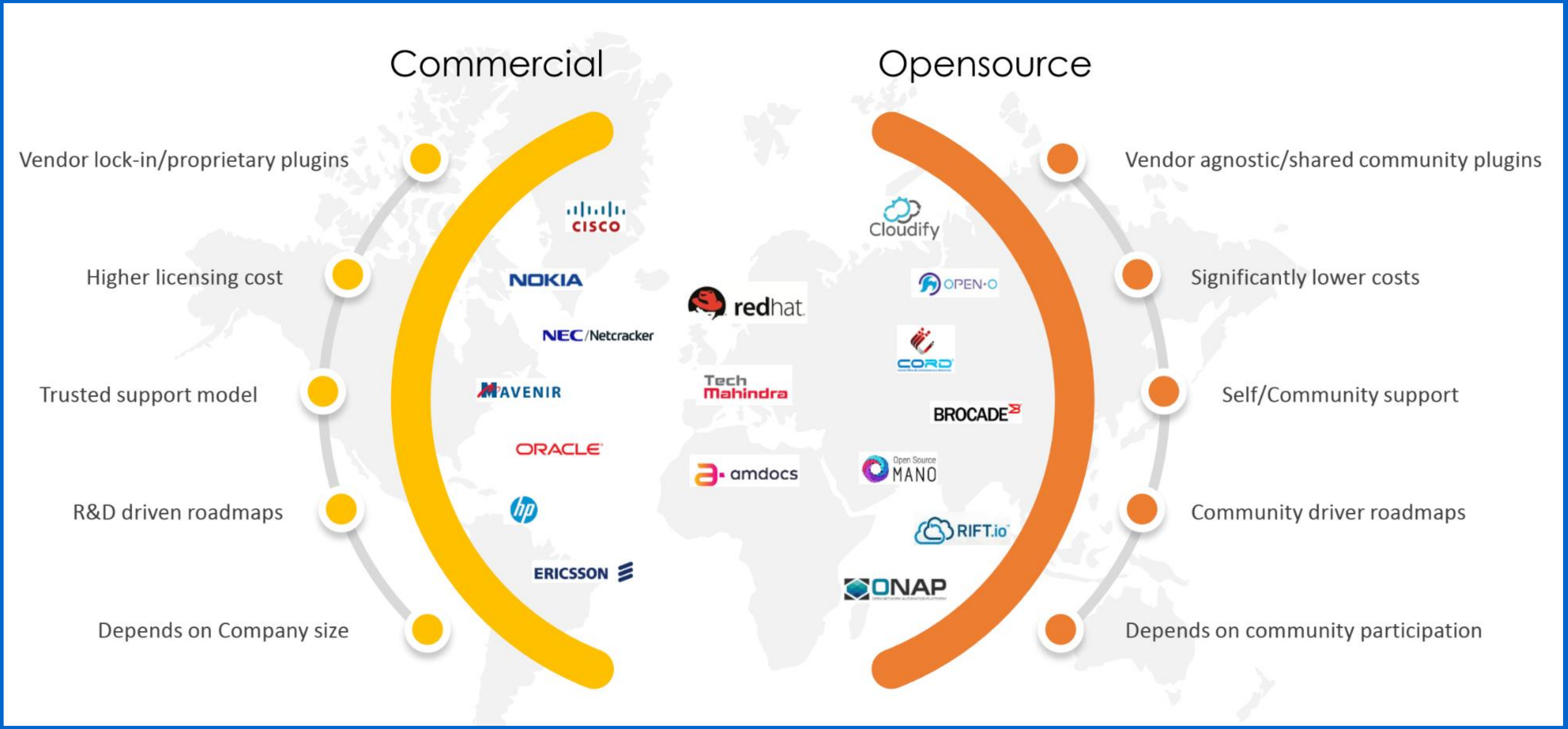
- The approach taken by Sampled Flow (sFlow) is unlike NetFlow and IPFIX, as it randomly samples full headers of packets. The choice of sampling rate plays a major role in the resulting efficacy.
- sFlow is advantageous as:
  - less bandwidth and CPU is required on the network devices compared to IPFIX and NetFlow
  - it captures and keeps the full packet headers
  - it needs no cache and operates in a real-time manner. Thus, it is beneficial for high-speed networks.
- However, there are some challenges:
  - as the right choice of sampling rate is difficult to figure out and there exists a trade-off
  - as sampling can negatively impact accuracy in some sense

Passive/Active Monitoring

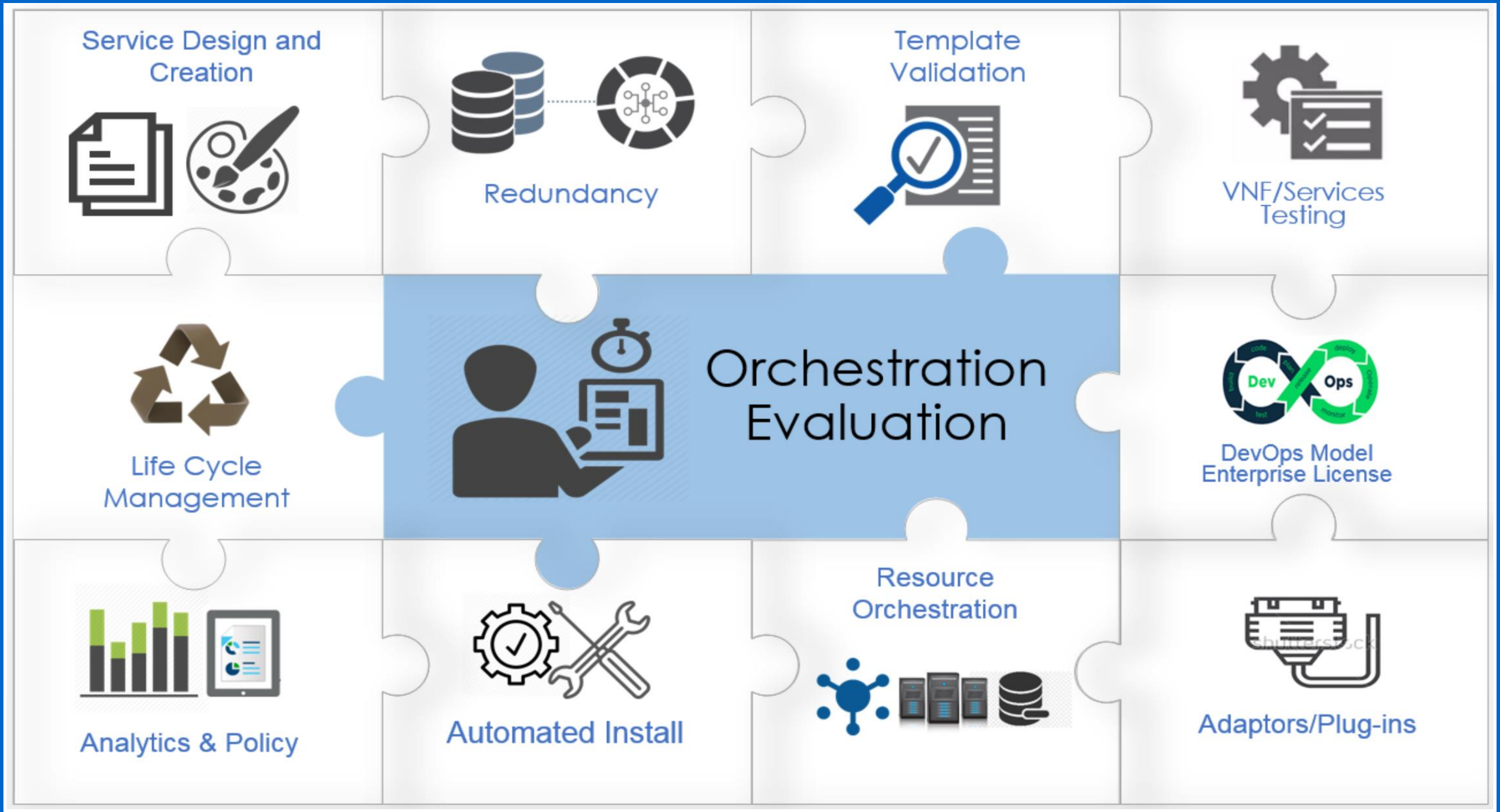
- Passive monitoring: observing the natural traffic that traverses the network
  - does not inject additional test traffic to the network
  - **real** data is collected and analyzed
  - however, a large volume of data is usually collected

- Active monitoring: generating test traffic to have it passed through the network, and measuring network's response
  - makes it possible to emulate arbitrary error scenarios before occurrence
  - small volumes of traffic suffice for achieving effective measurements

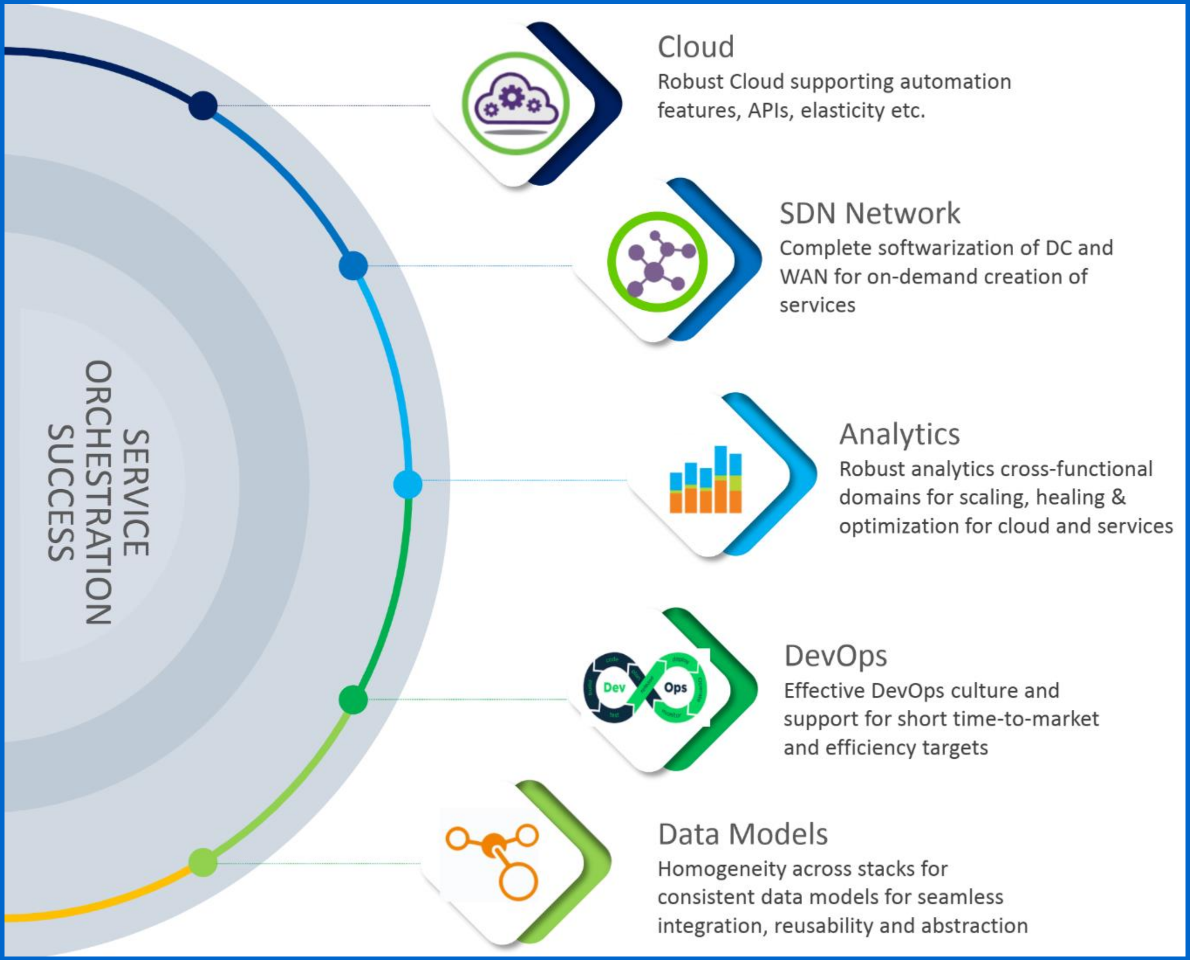
Orchestration: Commercial or Open Source



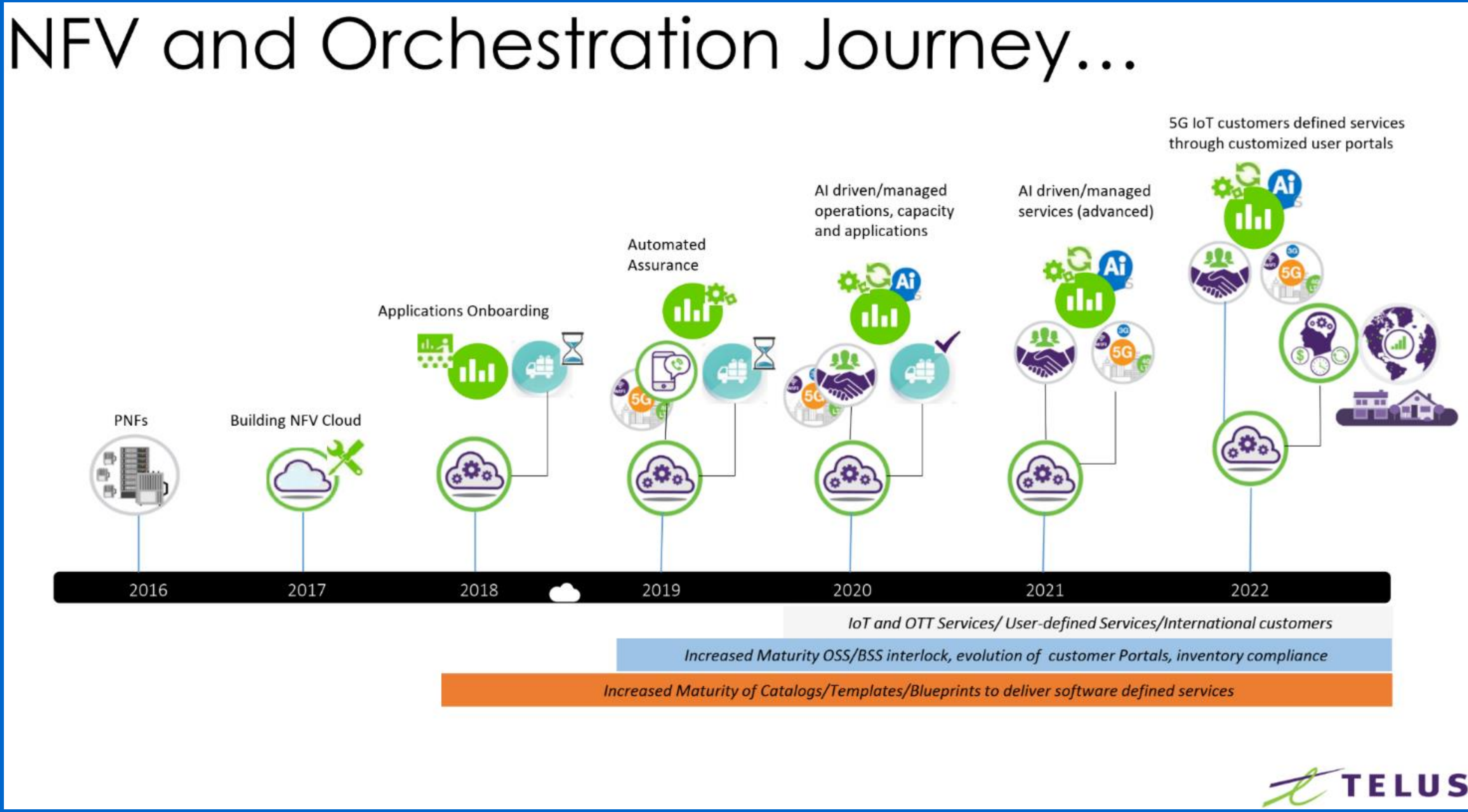
Orchestration Functional Elements



New Orchestration EcoSystem

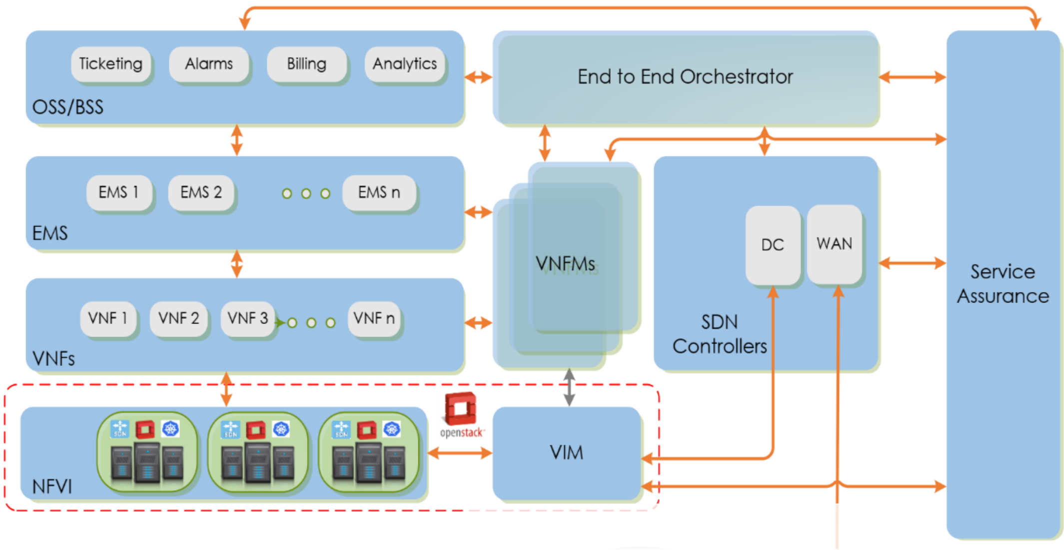


Management and Operation Use Cases of AI

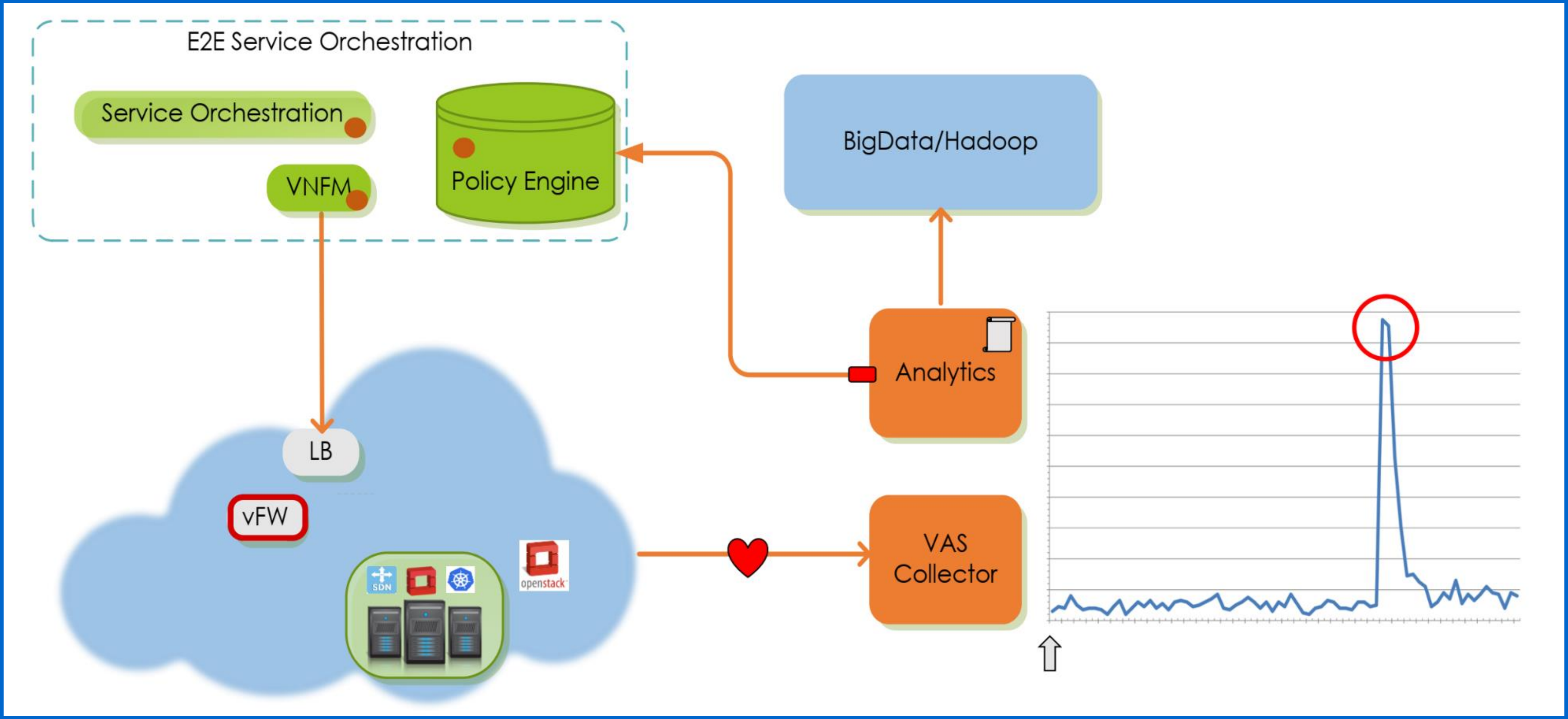


NFV and Orchestration Journey…

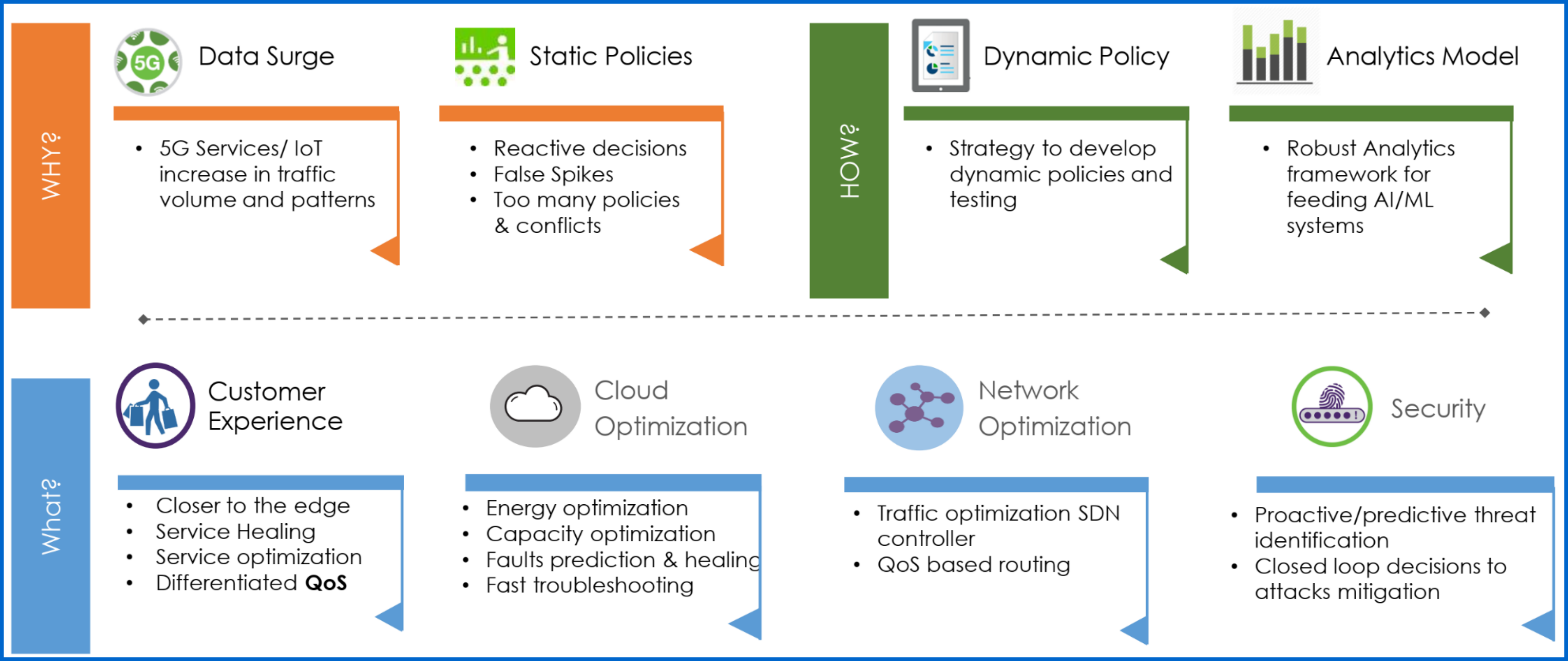
- Building robust cloud infrastructure
- Virtualizing Applications
  - Provisioning workflows
  - Assurance
  - Automated scaling
- Traffic steered through SDN network
- Monitoring through holistic service assurance
- Integration with OSS/BSS
- E2E Service Orchestration plays are major "orchestration" role



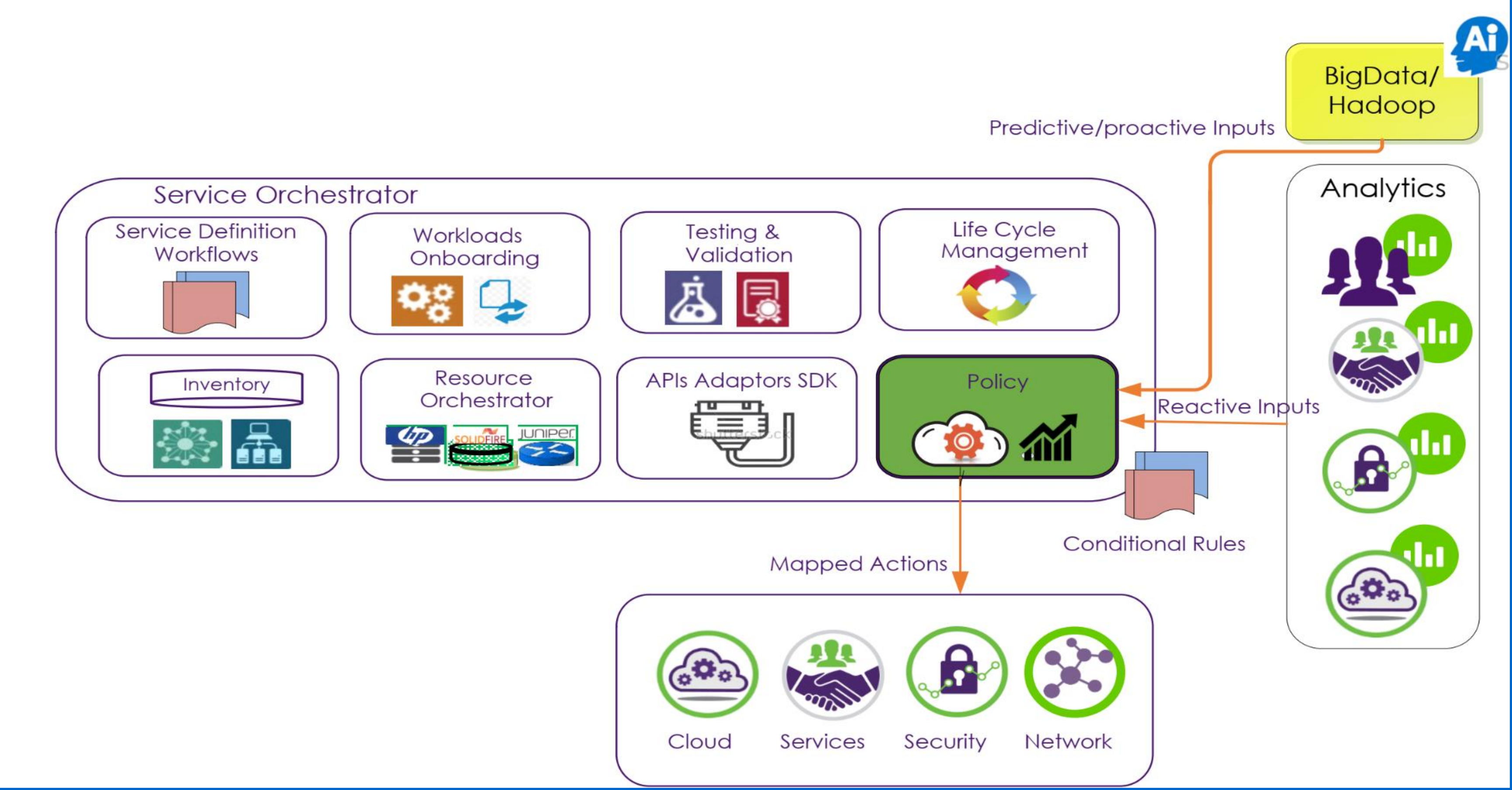
Traditional Closed Loop Orchestration



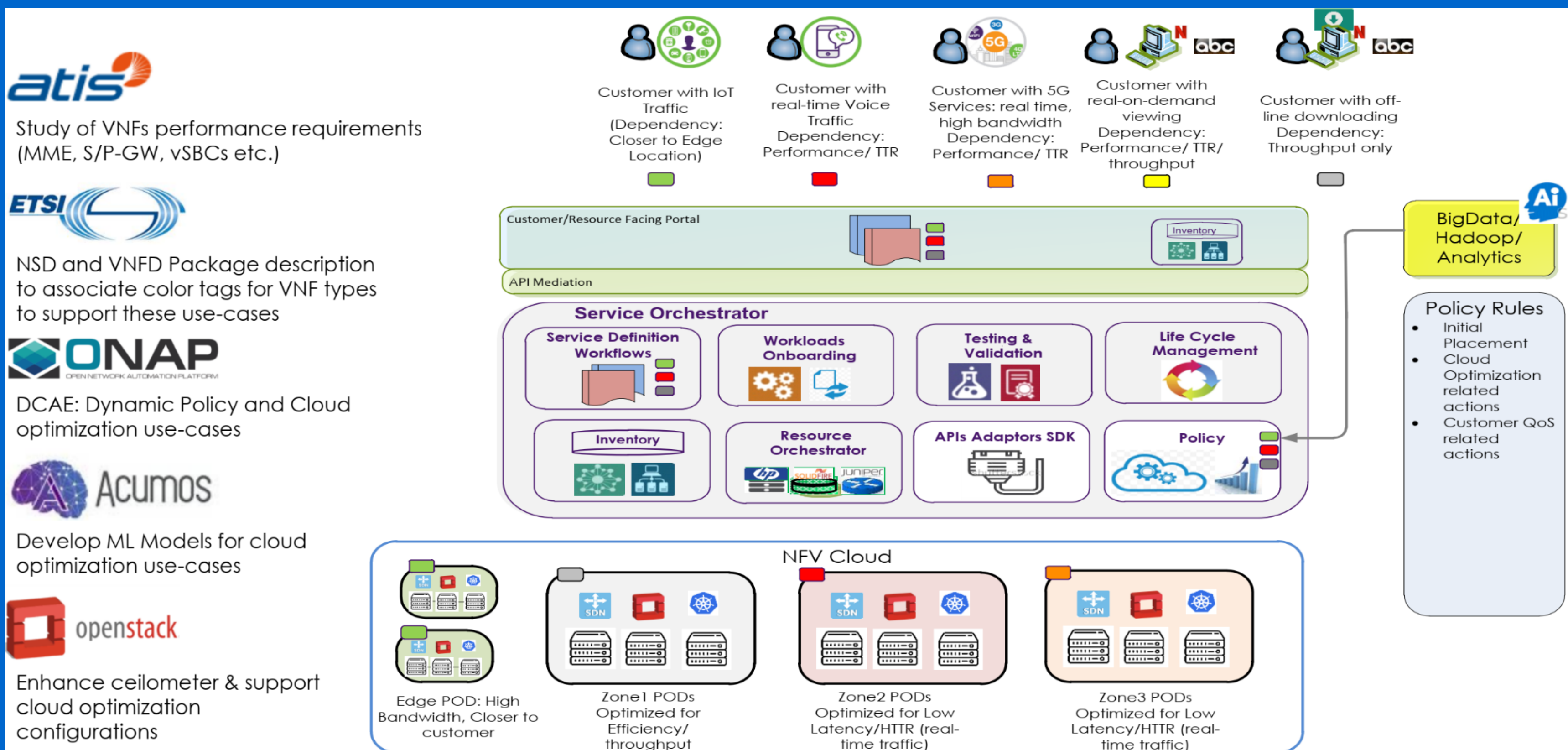
AI/ML Orchestration



Closed Loop Orchestration: AI/ML



AI/ML Orchestration Industry Verticals



Ecosystem for AI Platforms

	Unified view	Interaction assurance	Sandbox
AI-aaS	√	√	Networking & OTT
ML-pipeline [3]	√	×	Networking & OTT
ENI [6]	Networking	×	×
NWDAF [15]	Networking	×	×
ONAP [10]	Networking	×	×
ACUMOS [9]	Computing	×	OTT

- *Experiential networked intelligence (ENI)* is introduced by ITU-T to bring a closed-loop of AI mechanisms in network management
- *Network data analytic function (NWDAF)* is introduced by 3GPP to collect and analysis data of one network operator to offer data driven slice management.
- *ACUMOS* is an open source AI machine learning platform to develop any AI based applications in "*containerized microservices*" and modular manner.
- *Open network automation platform (ONAP)* is a microservice platform to bring a full network management automation in 5G for both physical and virtual network functions. In ONAP, there is a closed loop automation management platform (CLAMP) to provide any AI-based operation within networks.

ETSI Experiential Networked Intelligence (ENI)

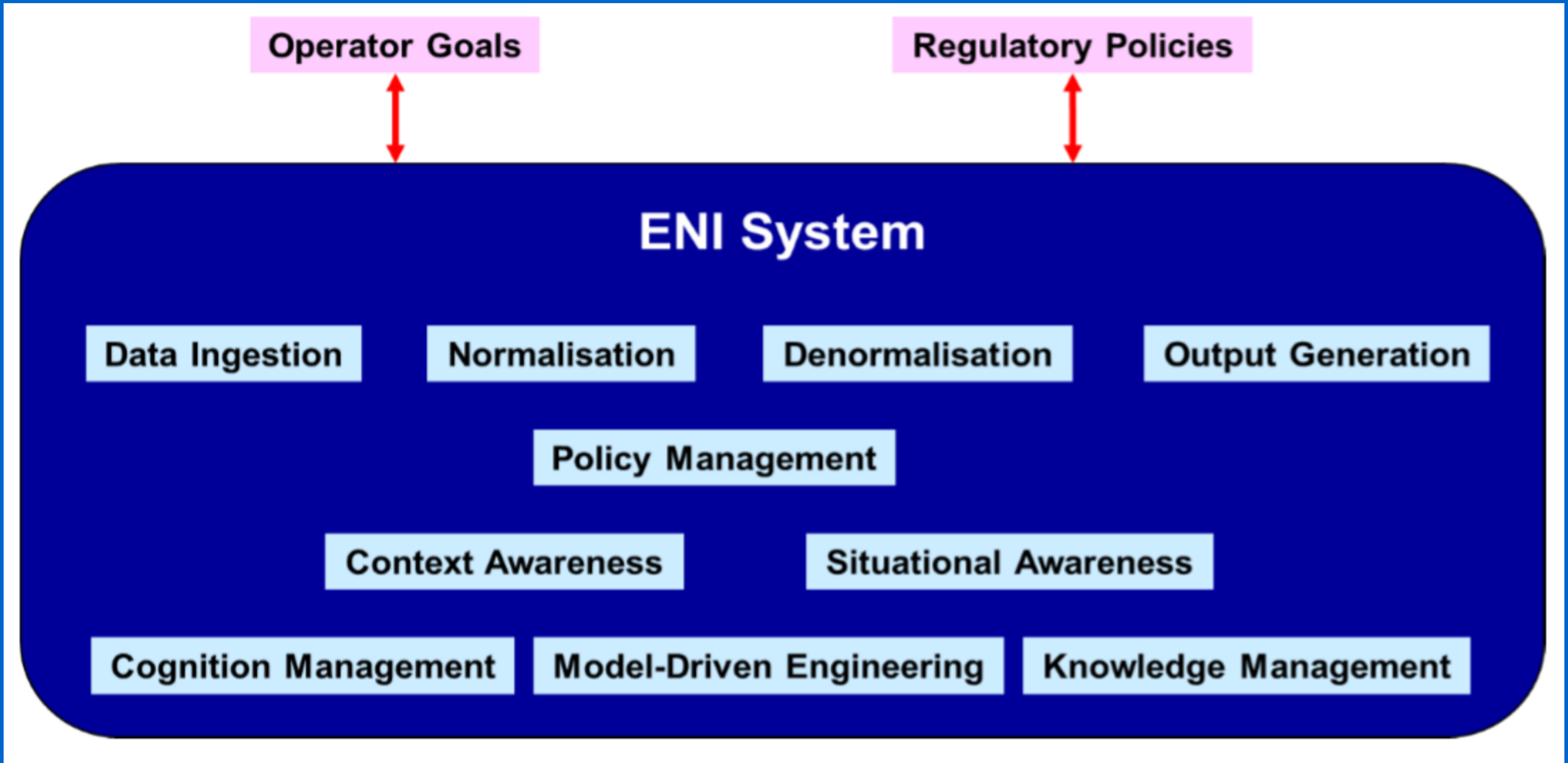
- Goal: **Improved operator experience**
- Based on the "observe-orient-decide-act" control model and Intelligent Policy-based management
- This encompasses open intelligent functionality for network configuration and management
- Propose a data collection and analyzing mechanism to provide End to-End (E2E) network diagnosis ability
- There is a need to automate the SDN/NFV:

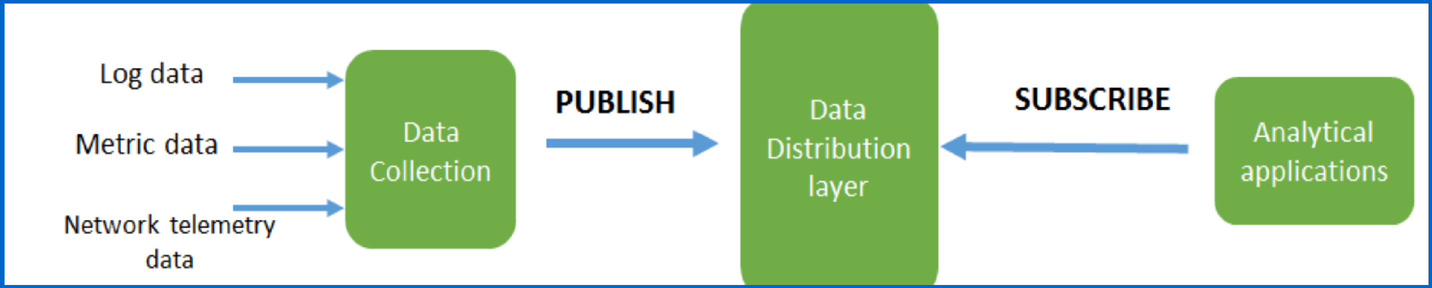
  policy - service - resource

Example Scenarios and Use Cases

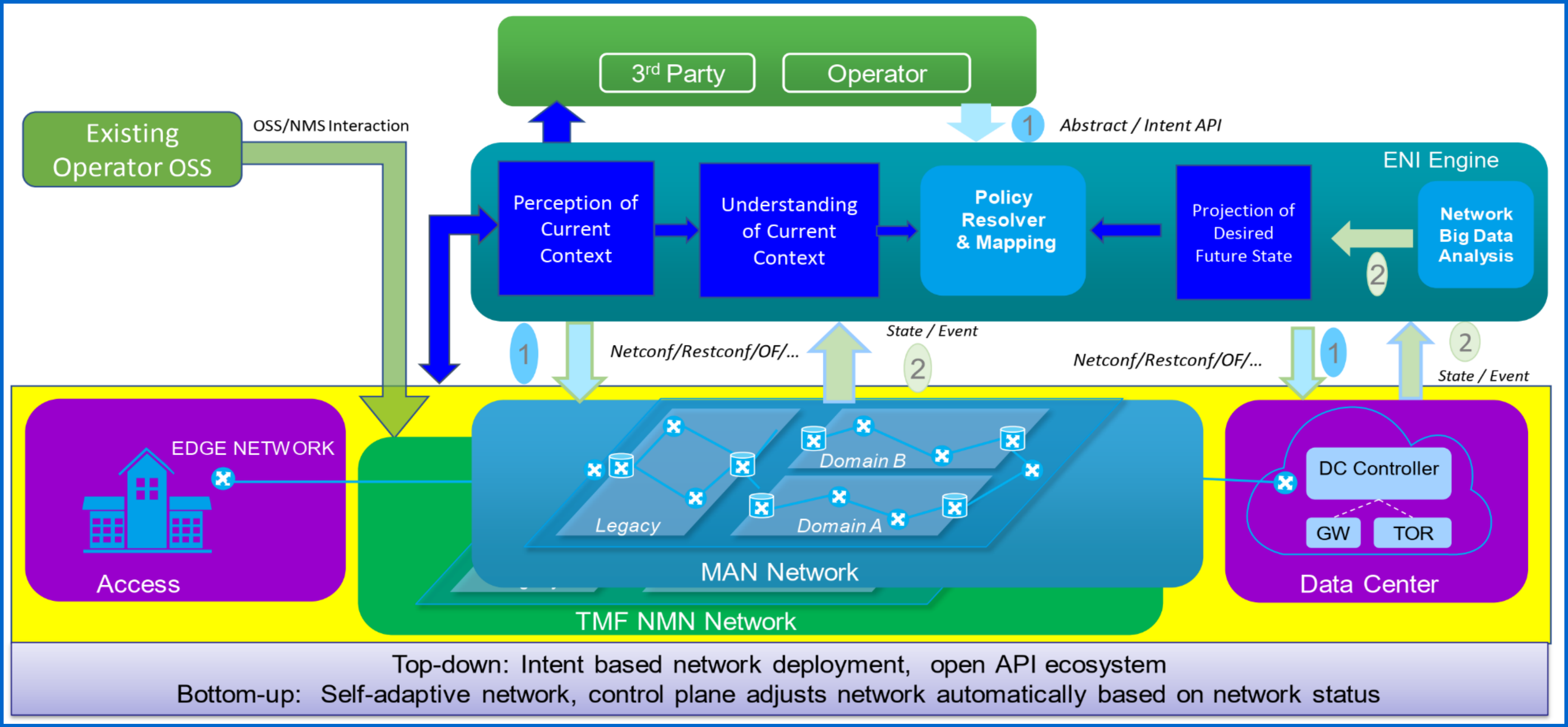
- **Policy-driven Inter-Data Center traffic steering**
  - With policy-driven traffic steering, operator experience is improved by making the network more resilient, and services are improved in terms of maintaining QoS and QoE.
- **Policy driven IP managed networks**
  - Manual IP address allocation is cumbersome; and scripts are fragile and cannot adjust to dynamic network conditions. Policy enables more intelligent usage of address pools and automates the address allocation.

ENI High Level Conceptual Model





ENI Conceptual Architecture



- Policies can be used to manage how services and resources interact with the environment to achieve a defined target.
- For example, policies can control the transition to a new state in a state machine, and the overall state machine is managed using a closed control loop.

The Standards People

Artificial Intelligence for networks: understanding it through ETSI ENI use cases

Presented by: **Dr. Luca Pesando (TIM, ISG ENI vice-chairman)**

Dr. Yue Wang (Samsung, ISG ENI secretary & rapporteur -Use Cases-)

April 17th 2020

ENI Use Cases of AI in Networking

Use Cases

Infrastructure Management
Policy-driven IDC traffic steering
Handling of peak planned occurrences
Energy optimization using AI

Network Assurance
Network fault identification and prediction
Assurance of Service Requirements
Network Fault Root-cause Analysis and Intelligent Recovery

Network Operations
Policy-driven IP managed networks
Radio coverage and capacity optimization
Intelligent software rollouts
Intelligent fronthaul management and orchestration
Elastic Resource Management and Orchestration
Application Characteristic based Network Operation
AI enabled network traffic classification
Automatic service and resource design framework for cloud service
Intelligent time synchronization of network
Intelligent Content-Aware Real-Time Gaming Network

Service Orchestration and Management
Context aware VoLTE service experience optimization
Intelligent network slicing management
Intelligent carrier-managed SD-WAN
Intelligent caching based on prediction of content popularity

Network Security
Policy-based network slicing for IoT security
Limiting profit in cyber-attacks

Source: ETSI GS ENI 001 ; ENI Use Cases

7

Definition of Categories for AI Application to Networks

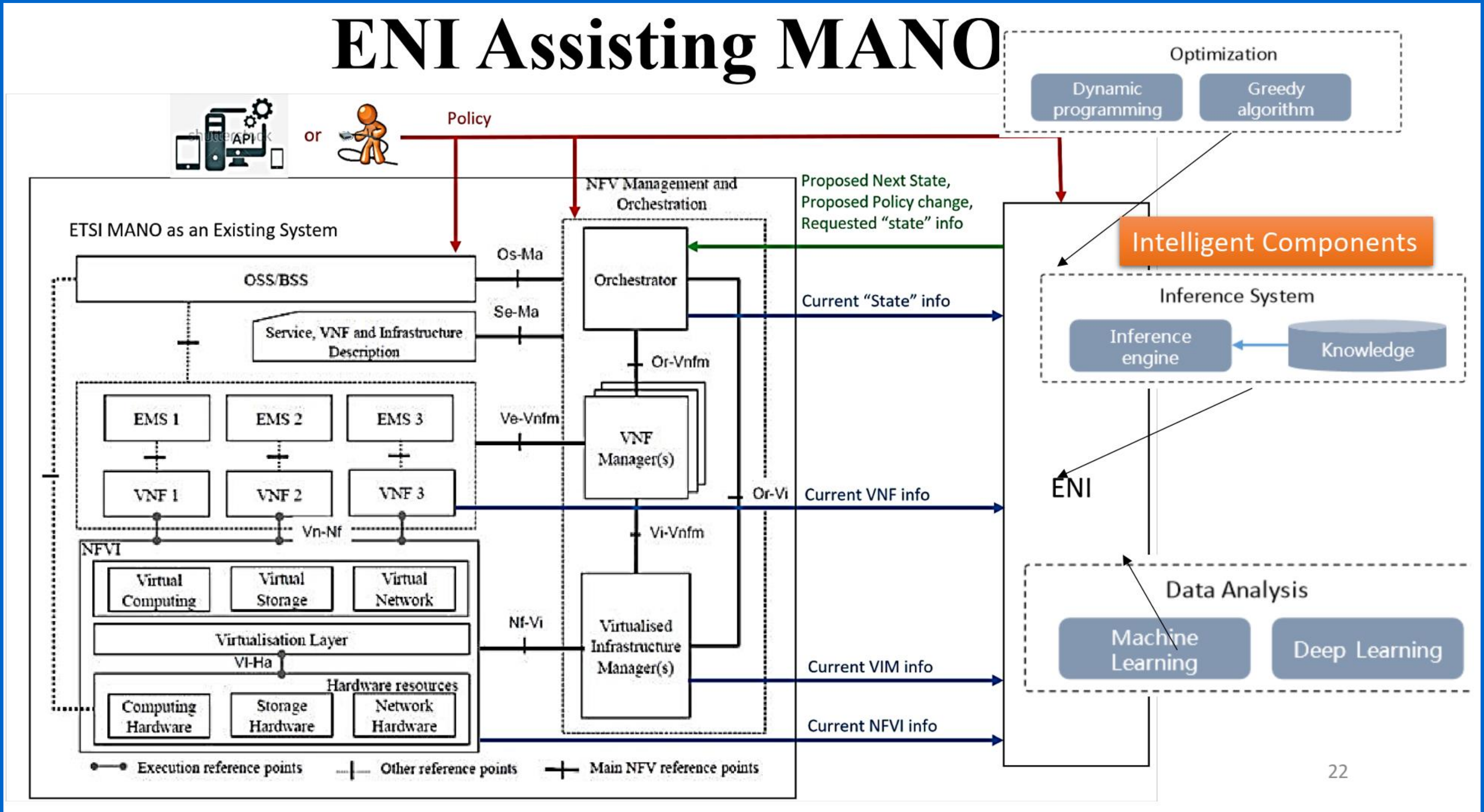
Category	Name	Definition	Man-Machine Interface	Decision Making Participation	Data Collection and Analysis	Degree of Intelligence	Environment Adaptability	Supported Scenario
Category 0	Manual O&M	O&M operators manually control the network and obtain network alarms and logs	How (command)	All-manual	Single and shallow awareness (SNMP events and alarms)	Lack of AI based understanding (manual management and control)	Fixed	Single scenario
Category 1	Assisted O&M	Automated scripts are used in service provisioning, network deployment, and maintenance. Shallow perception of network status and machine suggestions for decision making	How (command)	Provide suggestions for machines or humans and help decision making	Local awareness (SNMP events, alarms, KPIs, and logs)	Limited analysis capability	Limited adaptability to changes	Selected scenarios
Category 2	Partial automation	Automation of most service provisioning, network deployment, and maintenance Comprehensive perception of network status and local machine decision making	How (declarative)	The machine provides multiple opinions, and the machine makes limited decisions	Comprehensive awareness (basic telemetry data)	Deep analysis capability	Limited adaptability to changes	Selected scenarios
Category 3	Conditional automation	In specific environmental and network conditions there is automatic network control and adaptation	How (declarative)	Most of the machines make decisions	Comprehensive and adaptive sensing (such as data compression and optimization technologies)	Comprehensive analysis and knowledge; Short-term forecast capability	Adaptability to significant changes	Multiple scenarios
Category 4	Partial autonomicity	Deep awareness of network status; in most cases the network performs autonomic decision-making and operation adjustment	What (intent)	Optional decision-making response	Adaptive posture awareness	Comprehensive analysis and knowledge Long-term forecast capability	Adaptability to significant changes	Multiple scenarios
Category 5	Full autonomicity	In all environmental and network conditions, the network can automatically adapt	What (intent)	Machine autonomous decision	Adaptive optimization as a consequence of quality of service deterioration	Autonomic evolution and knowledge reasoning	Adaptability to any change	Any scenario

Categories of network autonomicity from a technical point of view

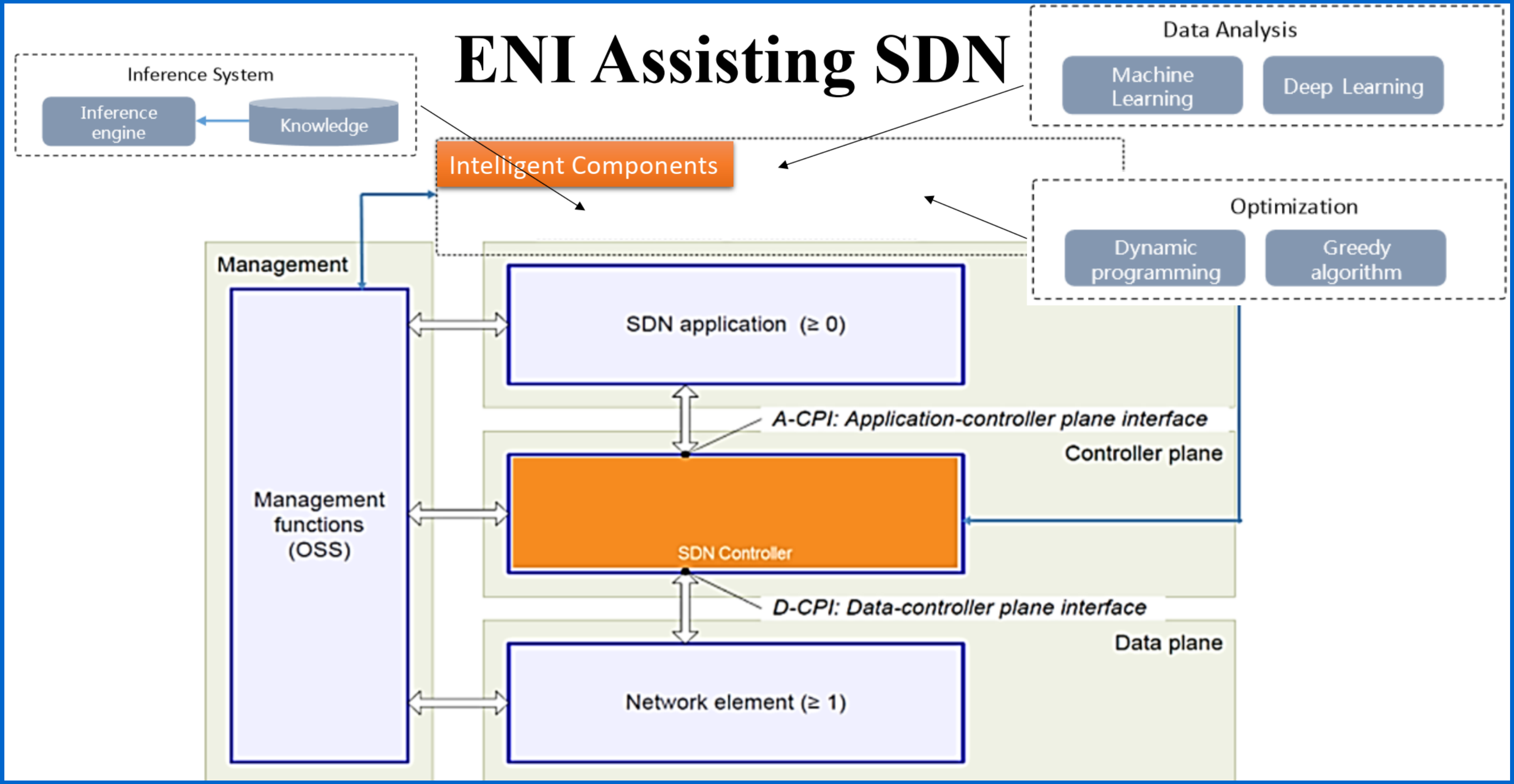
Source: ETSI GR ENI 007 , ENI Definition of Categories for AI Application to Networks

12

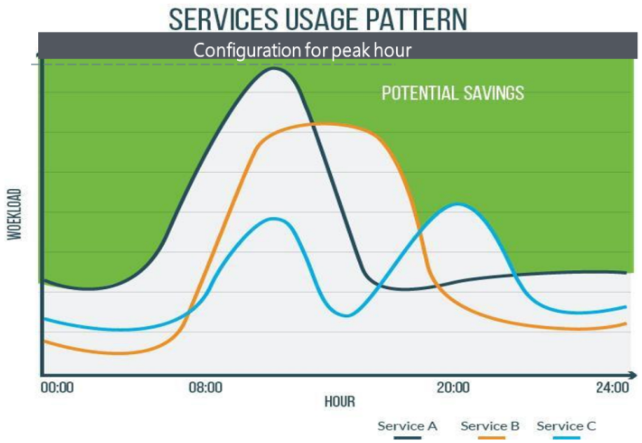
ENI Use Cases of AI in Networking



ENI Use Cases of AI in Networking



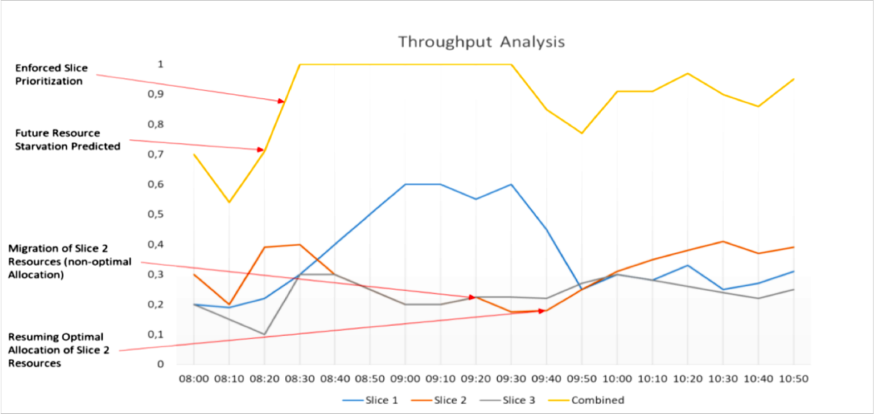
Energy Optimization Using AI



Challenges
- The servers in a DC take 70% of the total power consumption
- Servers are deployed and running to meet the requirement of peak hour service – 100% powered-up full time
- Huge amount of energy consumption
- Reliability is essential

The role of AI and ENI
- Learn and update the usage pattern of the services
- Autonomously turn the low traffic servers to idle state
- Predict the peak hours and wake up the necessary number of servers
- Predict workload of data centers
- Proactively determine on-off data centers

Assurance of Service Requirements



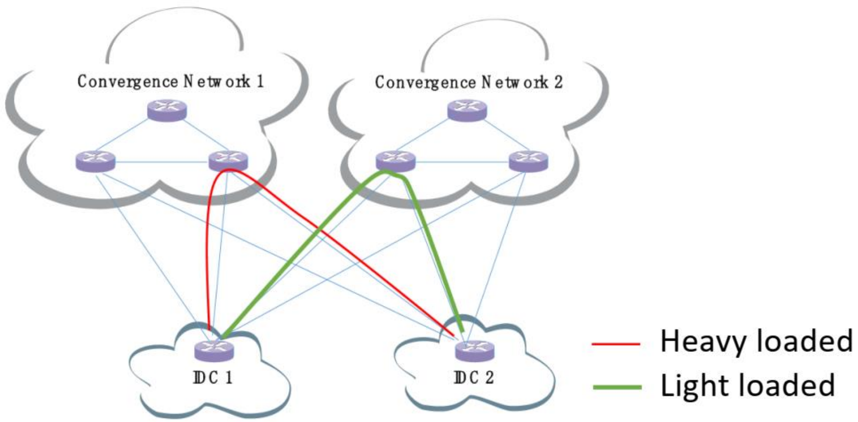
The role of AI and ENI

- Predict situations where multiple slices are competing for the same physical resources and employ preventive measures
- Predict or detect requirements change and make decisions (e.g., increasing the priority of a given network slice over neighboring slices) at run time

Challenges:

- Dedicated slice to each vertical (e.g., banking, energy provider company) pose highly restricted SLAs in network management
- There always exist resource allocation conflicts between competing network slices deployed on top of a shared infrastructure

Policy Driven IDC Traffic Steering



Challenges:

- Multiple links between IDCs, allocated by network administrator, e.g., shortest path strategy
- Link load not sufficiently considered when calculating the traffic path
- Bandwidth allocated to a tenant is not always fully used all time

The role of AI and ENI

- Autonomous service volume monitoring
- Network resource optimization through historical data and prediction in real-time
- Network traffic via different links is balanced
- Network resource, such as bandwidth, will be used more efficiently

25

ENI PoC (Proof of Concept) List

Title	PoC Team Members	Main Contact	Start Time	Current Status (Feb-2020)
PoC #1: Intelligent Network Slice Lifecycle Management	**China Telecom** Huawei, Intel, CATT, DAHO Networks, China Electric Power Research Institute	Haining Wang	Jun-2018	Completed
PoC #2: Elastic Network Slice Management	**Universidad Carlos III de Madrid** Telecom Italia S.p.A., CEA-Leti, Samsung R&D Institute UK, Huawei	Marco Gramaglia	Nov-2018	Completed
PoC #3: SHIELD, security through NFV	**Telefonica** Space Hellas, ORION, Demokritos (NCSR)	Diego R. Lopez Antonio Pastor	Jan-2019	Completed
PoC #4: Predictive Fault management of E2E Multi-domain Network Slices	**Portugal Telecom/Altice Labs** SliceNet Consortium	António Gamelas Rui Calé	Mar-2019	Ongoing
PoC #5: AI Enabled Network Traffic Classification	**China Mobile** Huawei, Intel, Tsinghua University	Weiyuan Li	Jun- 2019	Ongoing
PoC #6: Intelligent caching based on prediction of content popularity	**China Unicom** Beijing University of Posts and Telecommunications, Samsung, Cambricon, Huawei	Bingming Huang	Sep-2019	Ongoing
PoC #7: Intelligent time synchronization of network	**China Unicom** Beijing University of Posts and Telecommunications, Samsung, Cambricon, Huawei	Bingming Huang	Sep-2019	Ongoing
PoC #8: Intent-based user experience optimization	**China Telecom/Huawei Technologies** China Telecom, Huawei Technologies, AsiaInfo, Beijing University of Posts and Telecommunications	Dong Li	Jan-2020	Started
PoC #9: Autonomous Network Slice Management for 5G Vertical Services	**Nextworks** TIM, Nextworks, Samsung, WINGS, UC3M	Gino Carrozzo	Jan-2020	Started
PoC #10: Intelligent Telecom Network Energy Optimization	**China Mobile** China Mobile Research Institute, Intel, Quanta Cloud Technology, Hong Kong ASTRI (Applied Science and Technology Research Institute)	Yan Yang	Jan-2020	Started

ENI PoC project #1: Intelligent Network Slice Lifecycle Management

Host/Team Leader — China Telecom

Team members — Huawei, Intel, State Grid, 中国电力科学研究院有限公司, DAHO 大河云联

AI-based predictor:

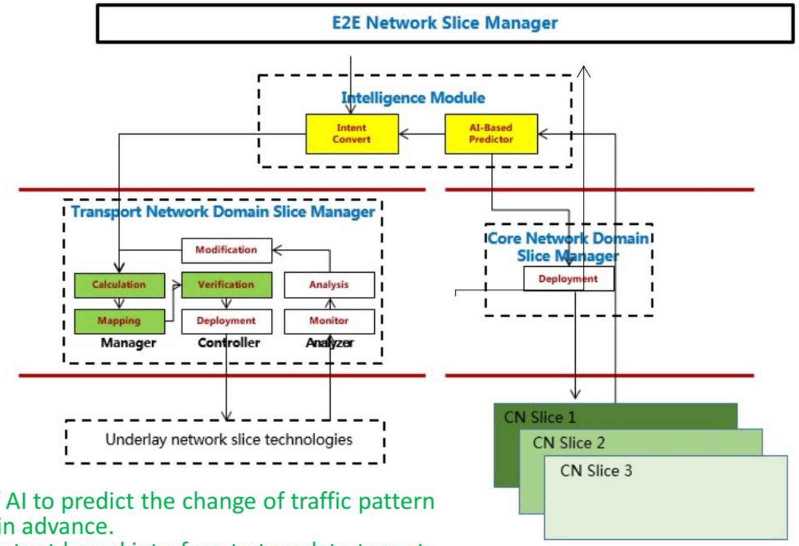
- For generating new scale up/down and converting the intent to suggested configuration.
- LSTM is used for traffic prediction.

TNSM:

- Provides underlay network control to satisfy the network slice requests.
- FlexE and a FlexE-based optimization algorithm are used for underlay network slice creation and modification.

CNSM:

- Provides core network control to satisfy the network slice requests



Showcases:
Beijing, Sep 19-20, 2018
Nanjing, Nov 14-16, 2018
Warsaw, Apr 10, 2019

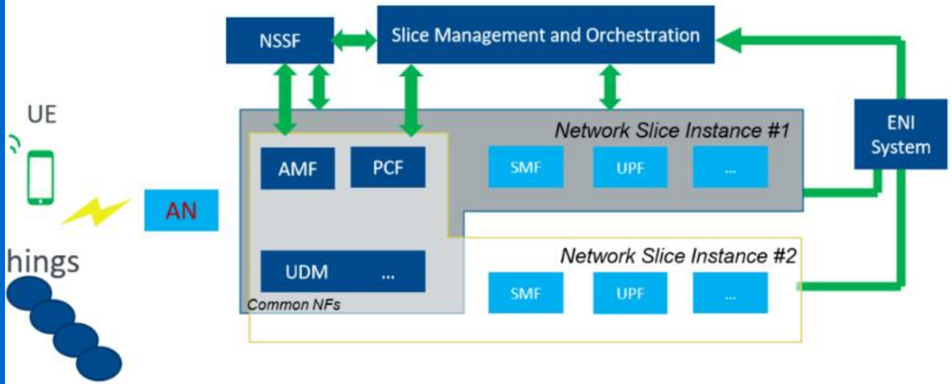
✓ PoC Project Goal #1: Demonstrate the use of AI to predict the change of traffic pattern and adjust the configuration of network slice in advance.

✓ PoC Project Goal #2: Demonstrate the use of intent based interface to translate tenant requirements to network slice configuration and intelligent network slice lifecycle management on demand.

https://eniwiki.etsi.org/index.php?title=PoC#PoC.231_:_Intelligent_Network_Slice_Lifecycle_Management

17

LSTM: Long short-term memory for time series prediction



The role of AI and ENI

- Analyze collected data associated to e.g., traffic load, service characteristic, VNF type and constraints, infrastructure capability and resource usage, etc.
- Produce a proper context aware policy to indicate to the network slice management entity when, where and how to place or adjust the network slice instance (e.g., reconfiguration, scale in, scale-out, change the template of the network slice instance)
- Dynamically change a given slice resource reservation

Challenges:

- Highly dynamic traffic pattern
- Allocation of network resources intra/inter slices (VNF scale in/out, up/down)

ENI PoC project #3: Securing against Intruders and other threats through a NFV-enabled Environment (SHIELD)

Status: PoC public demo 23-25th Feb. 2019, and **finished**.

Host/Team Leader: Telefónica

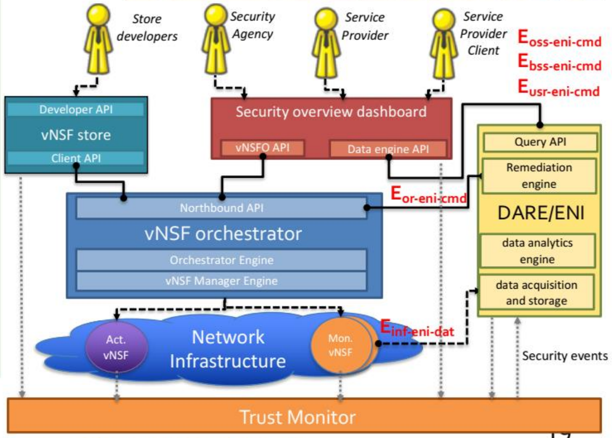
Team members:: SPACE ORION

SHIELD

https://www.shield-h2020.eu/

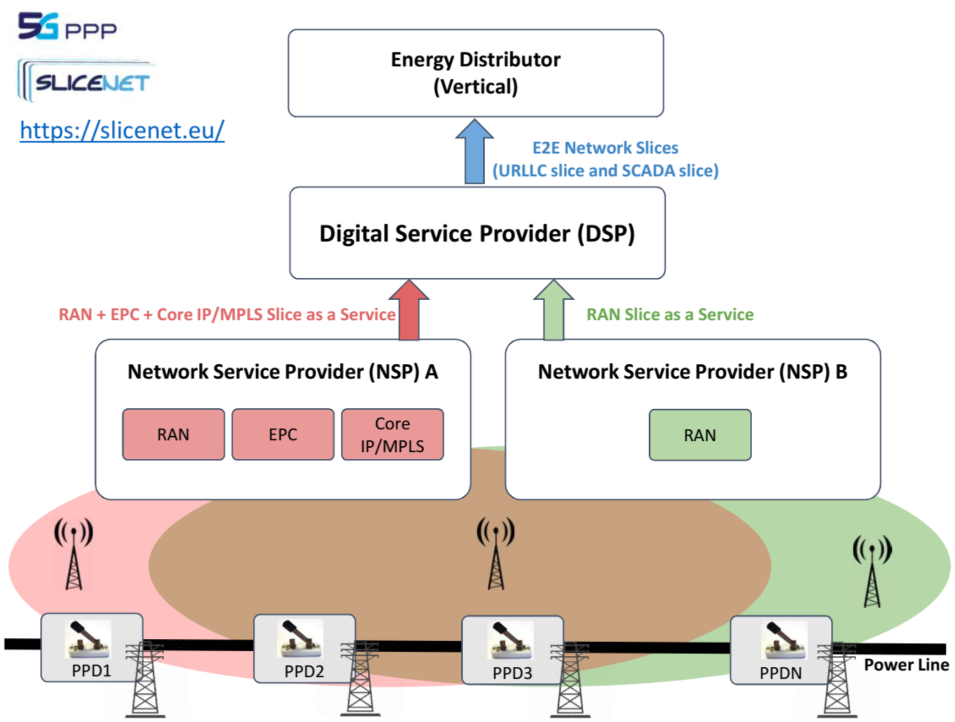
Goals	Gaps identified	Contribution

• PoC Goal #1: Demonstrate an AI framework able to detect network attacks over NFV network combining several ML algorithms
• PoC Goal #2: Recommend intent-based security policy
• PoC Goal #3: Remote attestation for data collectors.
	• Coordination between AI models / ENI systems.
	  • Support 3rd party AI models
	  • Information sharing between ENI systems
	• Data collection integrity and security with Trust Monitoring
	• Synchronization between intent-based and configuration policies API
		• New type of use cases related with malware in ENI 001



https://eniwiki.etsi.org/index.php?title=PoC#PoC.233:_Securing_against_Intruders_and_other_threats_through_a_NFV-enabled_Environment

ENI PoC project #4: Predictive Fault management of E2E Multi-domain Network Slices



- PoC scenario is a power grid vertical, that uses 5G to provide time sensitive communications for grid protection mechanisms. A Network Slice is provided by a DSP for that effect.

- PoC is focused on the **DSP** functions
  - NSPs provide Sub-slices
  - DSP monitors all Sub-slices behaviour
  - DSP predicts Sub-slice failure
  - DSP decides best failover sub-slice alternative
  - DSP triggers Subslice/NSP switching

PoC Project Goal #1: Network Slice Fault Prediction. Demonstrate the use of AI on performance data to be able to accurately predict failure situations on Network Slices and estimate their impact on an E2E multi-domain slice performance.

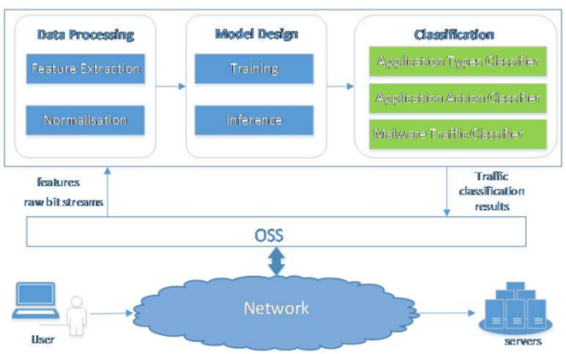
PoC Project Goal #2: Policy-based Network Slice Management. Evaluate the use of a policy-based structure for slice composition decisions, as well as the mechanisms for policy definition on that same context.

https://eniwiki.etsi.org/index.php?title=PoC#PoC.234:_Predictive_Fault_management_of_E2E_Multi-domain_Network_Slices

ENI PoC project #5: Intelligent Traffic Profiling

Host/Team Leader:

Team members:

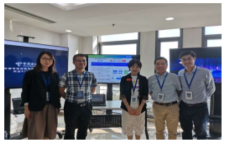


This PoC aims to verify the feasibility and the benefits of the use of AI/ML for network traffic classification, including the encrypted traffic, and demonstrate in a testbed environment that how ENI system can support intelligent traffic profiling and mechanism generalization.

This PoC consists of three scenarios:

- ✓ **The first scenario** demonstrates that the AI/ML-based technique enables network traffic to be categorized into classes of application types, e.g. high-throughput data, real-time interactive, multimedia streaming, low-priority data.
- ✓ **The second scenario** shows that traffic flows generated by a specific application can be classified into classes of subactions types , e.g. query action, call action etc.
- ✓ **The third scenario** shows that the malware traffic and normal traffic can be identified based on the AI/ML algorithms.

Showcases:
- ☐ Beijing, Sep 27, 2019
- ☐ Guangzhou, Nov 14-16, 2019
- ☐ Sophia Antipolis France, **Dec**9-10, 2019



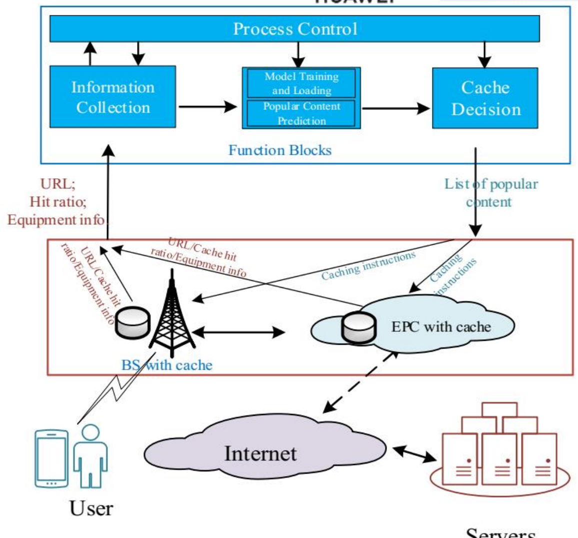
21

https://eniwiki.etsi.org/index.php?title=PoC_05:_Intelligent_Traffic_Profiling

ENI PoC project #6: Intelligent caching based on prediction of content popularity

PoC Team

This PoC is meant to show the feasibility and the benefits of an intelligent mobile edge cache, especially when there are large amount users request the same content, and demonstrate in a testbed environment that how ENI system can support intelligent caching based on prediction of content popularity.



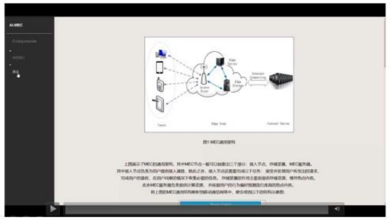
Scenarios of this PoC

Goal #1:

- ✓ The first scenario demonstrates how the system, which is called intelligent caching based on prediction of content popularity, predicts popular content.

Goal #2:

- ✓ The second scenario shows the functionality of this system in the network, specifically to reduce the latency and reduce the core network load.

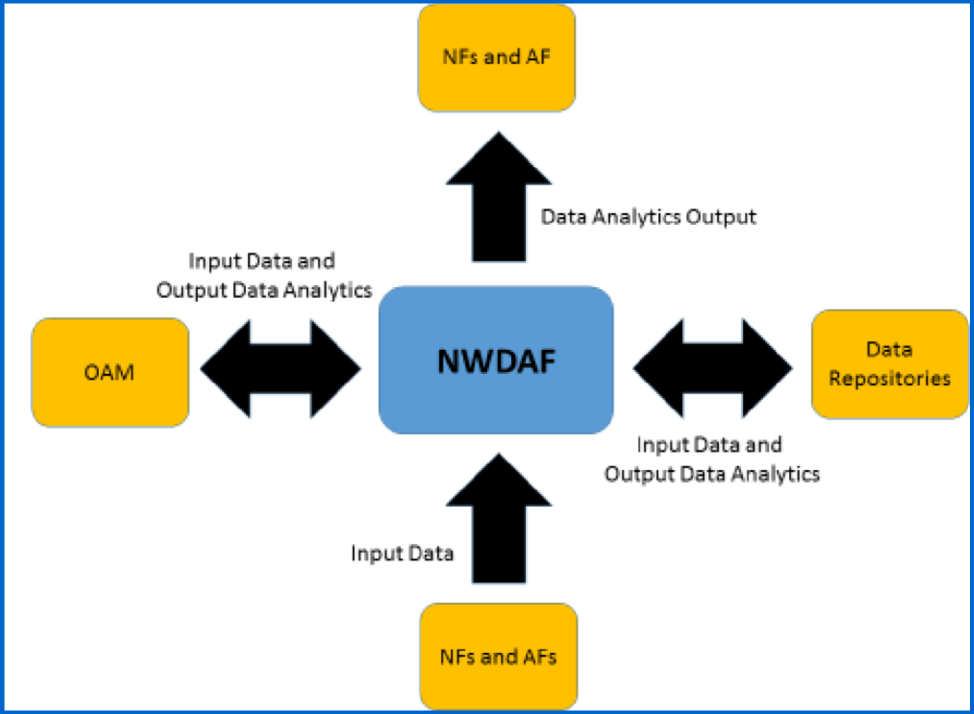


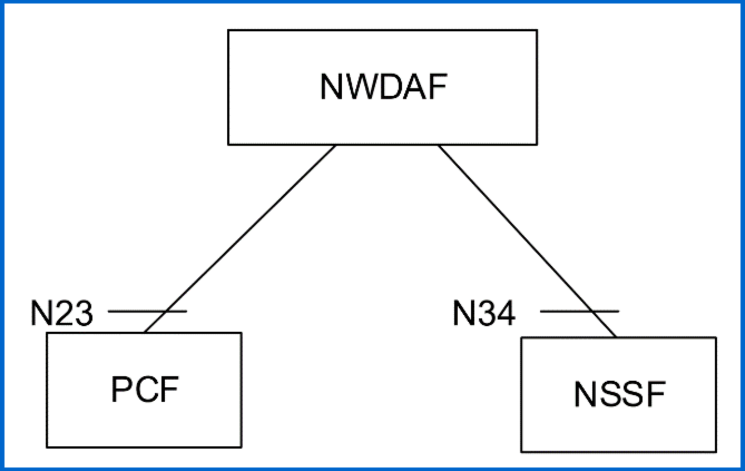
Showcases:
Sophia Antipolis
France, **December** 9-10, 2019

https://eniwiki.etsi.org/index.php?title=PoC_06:_Intelligent_caching_based_on_prediction_of_content_popularity

NWDAF: Network Data Analytics Function 3GPP Data Collection for AI/ML

- 3GPP TS 23.501, 502, 503, TS 129.520 v15.3, Technical Report TR 23.791
- NWDAF used for assistance based on location analytics, congestion analytics, network performance prediction, QoS provisioning and adjustment, selection of NF instances, etc.
- Both PCF and NSSF are consumers of NWDAF. The PCF may use that data in its policy decisions. NSSF may use the load level information provided by NWDAF for slice selection.
- NWDAF can collect UE mobility-related information.
- It provides output of analytics on mobility data to the PCF.





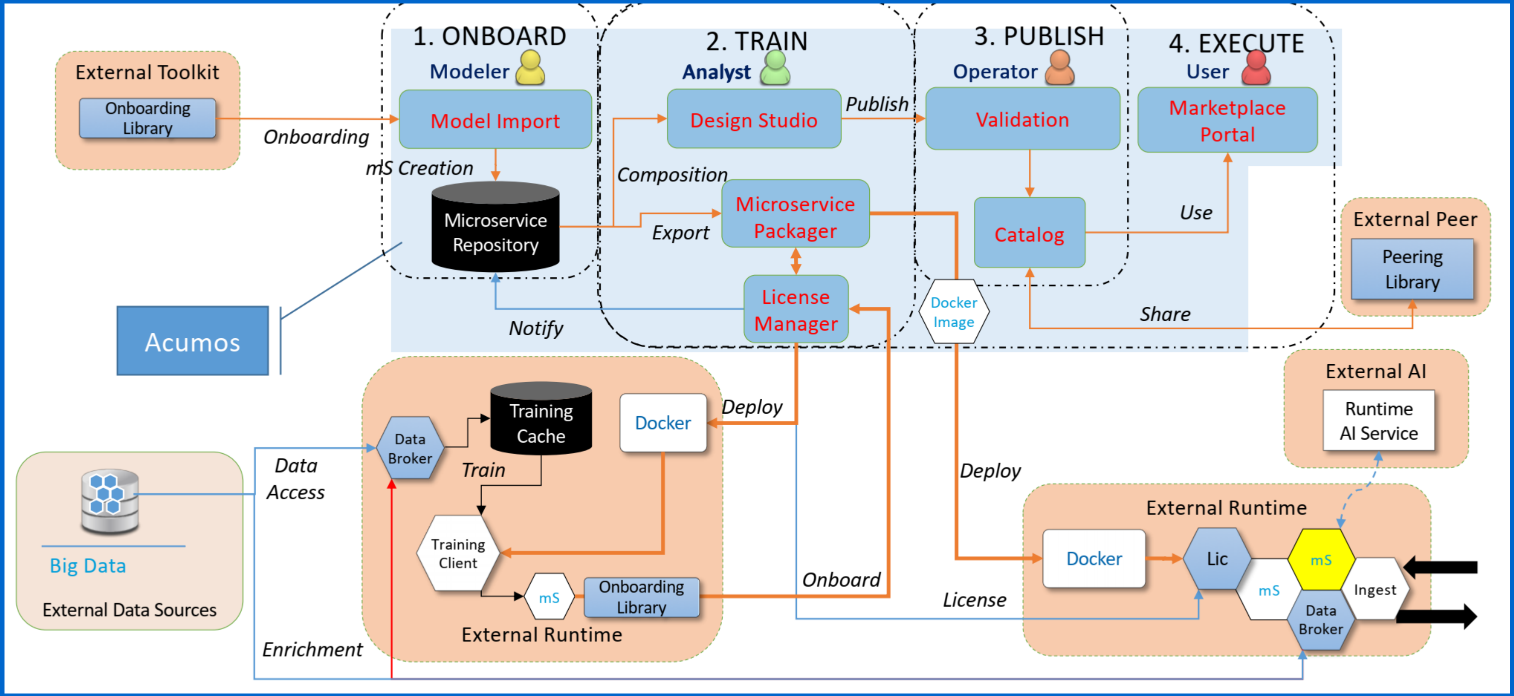
Acumos in Cloud Platfrom

- The final step in building a working AI solution, is to deploy a working application into a runtime system.
- For instance, it is hard to imagine an autonomous vehicle making an HTTP request to a central brain to make real-time decisions such as avoiding pedestrians.
- Acumos packages solutions into Docker image files, which can then be deployed into any Docker environment and managed through a set of container management tools, such as Kubernetes.
- Acumos provides tools to package any set of components, including predictors, adapters and other microservices, as needed, to any runtime environment and to create a compatible, deployable image file.

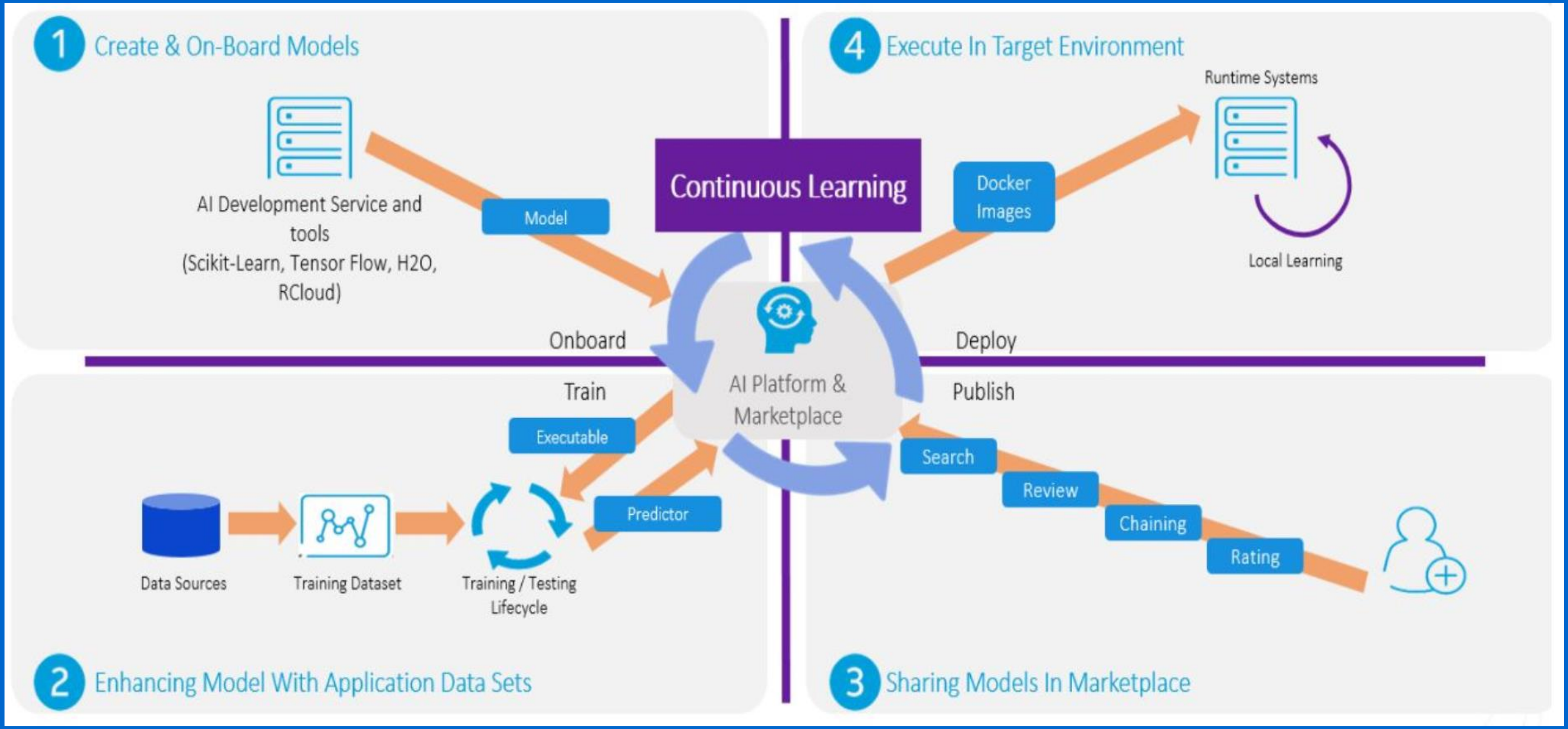
Acumos AI Management Platform

- Acumos breaks the development flow into four distinct steps:

1. Models are onboarded to an Acumos-compliant platform and packaged as distinct microservices with a component describing the APIs
2. Model is packaged into a training application that can be deployed to a training environment from which data can be cached for later retraining
3. A predictor, is published into a catalog that can be shared where other developers can find it, discuss it, review it and create a full solution by using the predictor and "chaining" to other components employing the predictor.
4. Finally, the entire solution is packed a Docker container where it can be executed.

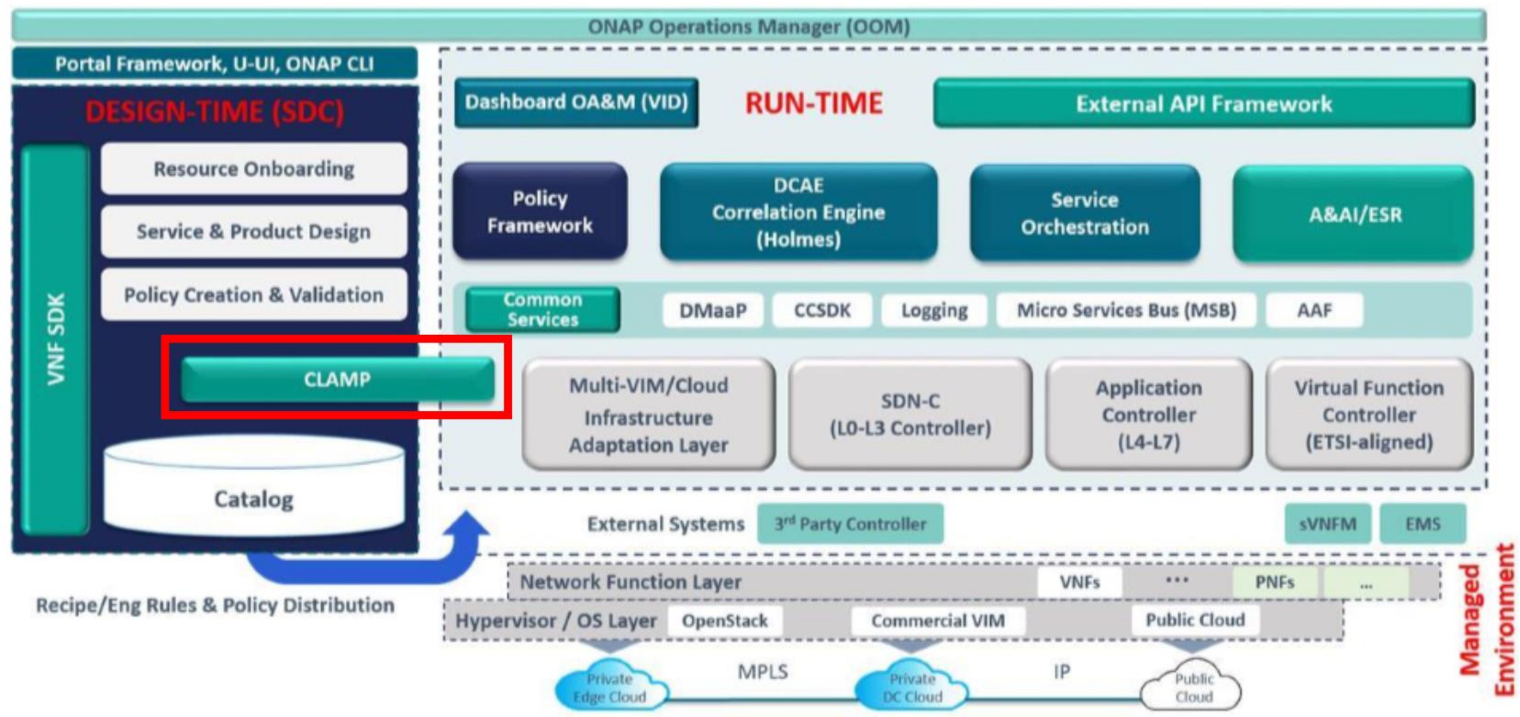


Acumos AI Management Architecture



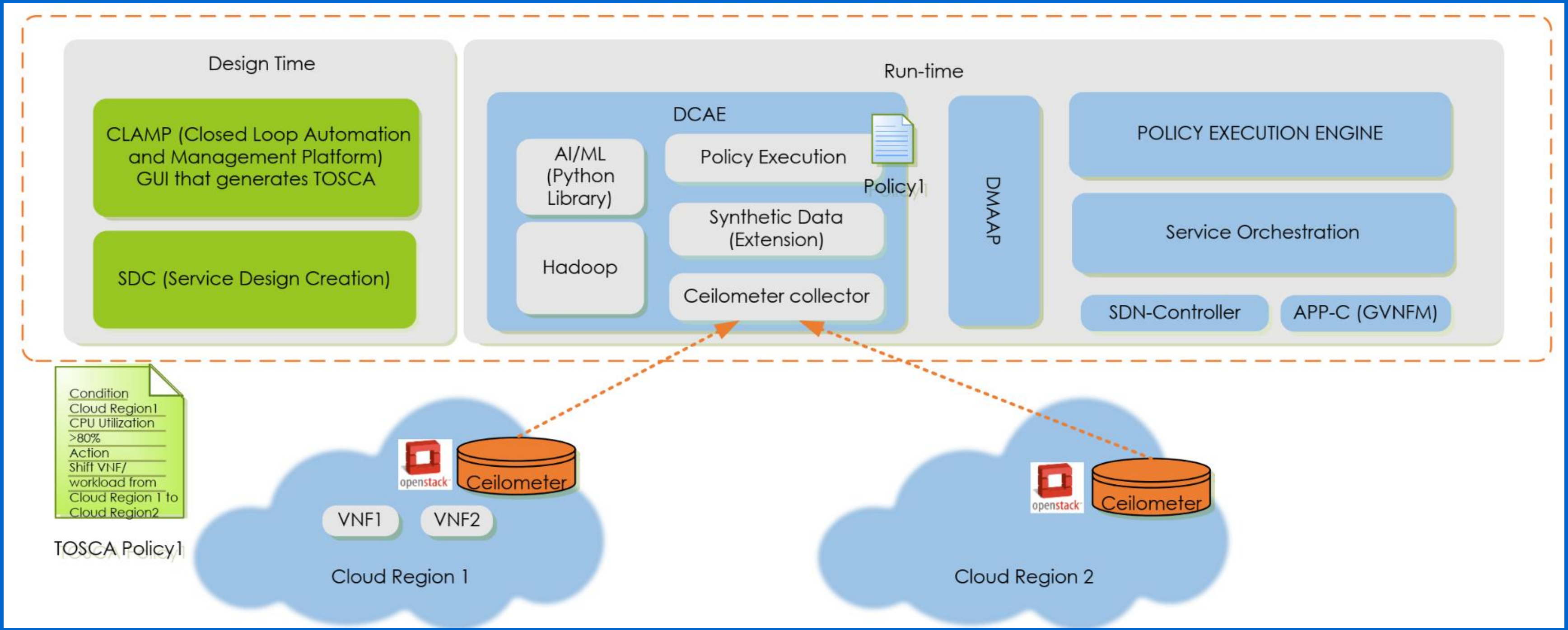
ONAP (Open Networking Automation Platform)

An open source networking project hosted by the Linux Foundation
ONAP platform uses SDN and NFV to orchestrate and automate physical and virtual network services



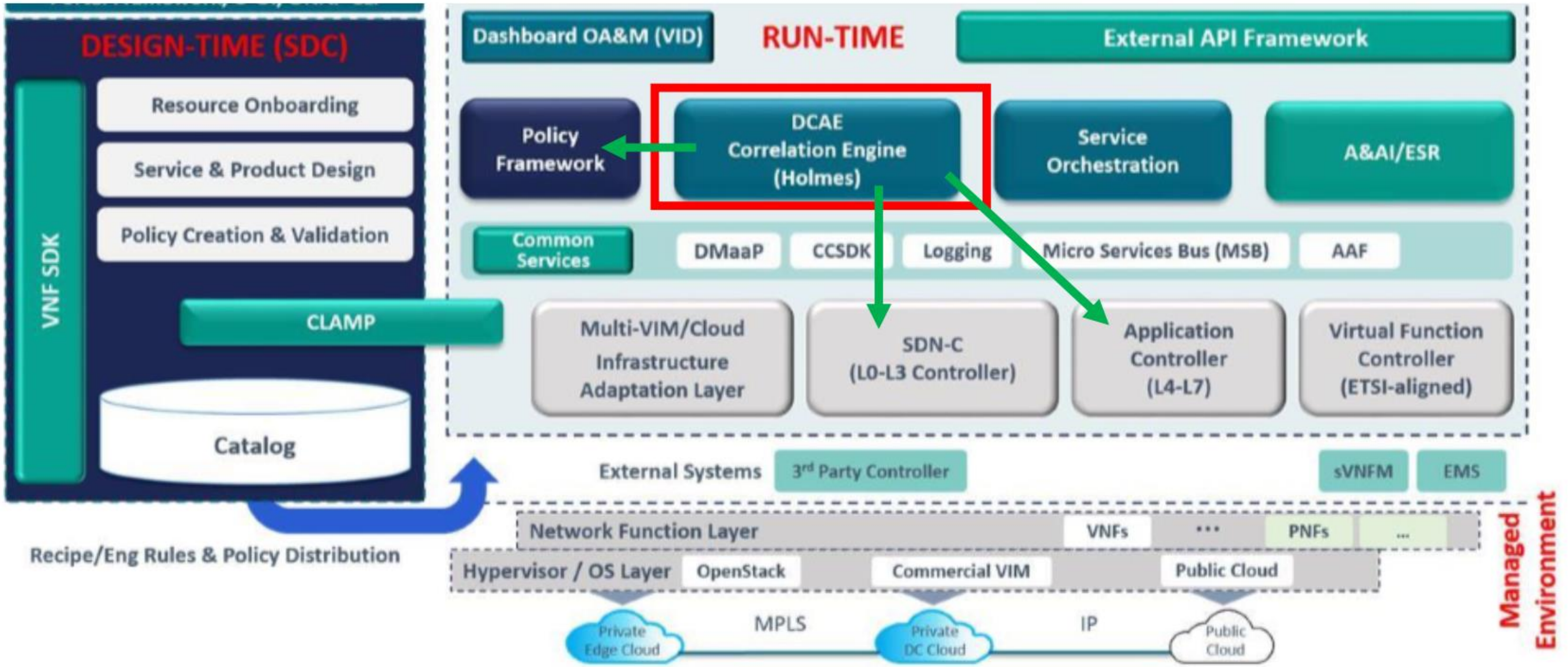
CLAMP: platform for designing and managing control loops. Used to visualize a control loop, configure it with specific parameters for a particular network service, then deploying and undeploying it.

ONAP Intelligent Closed Loop Architecture

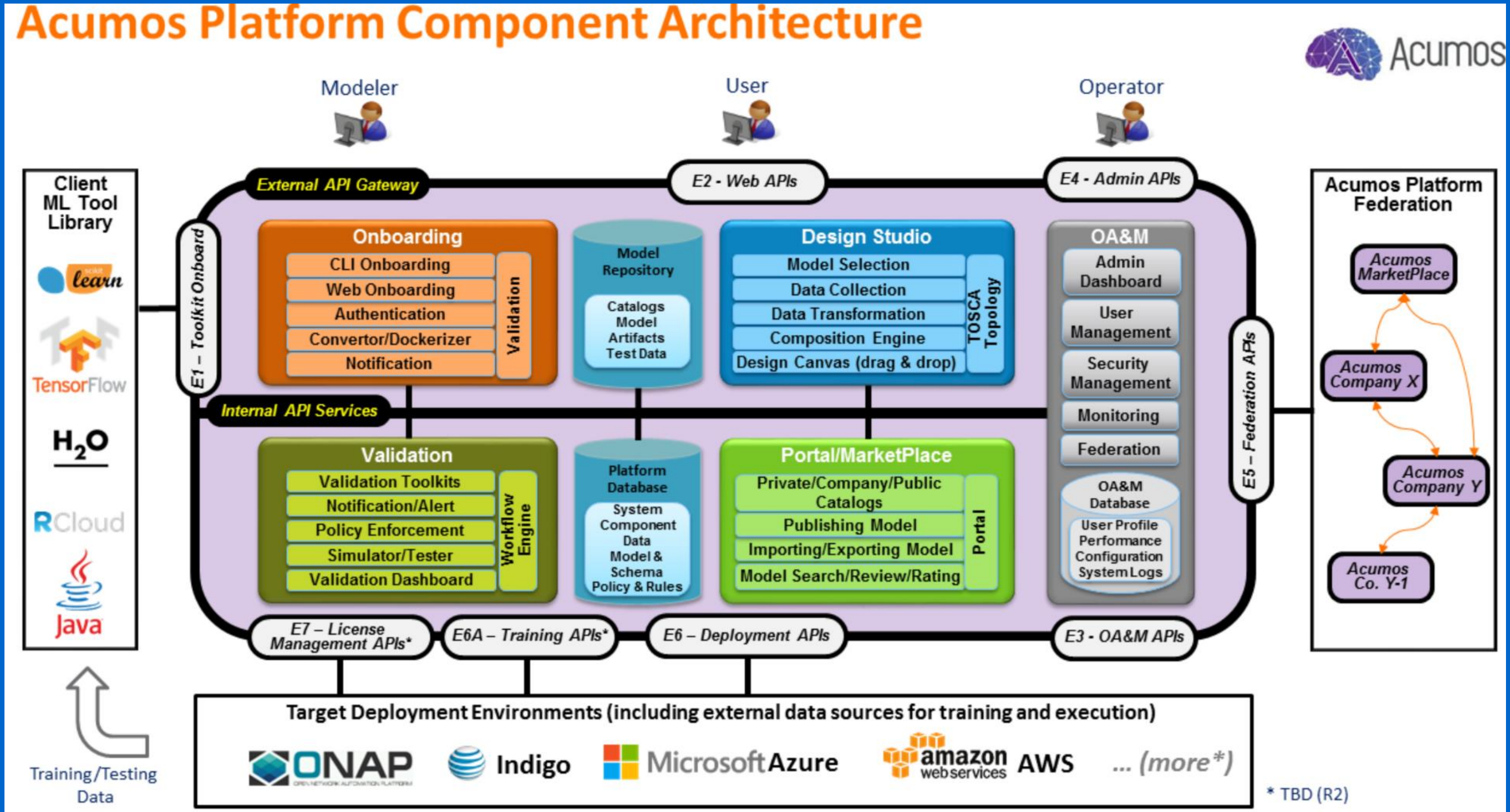


Data Collection Analytics and Events: data collection and analysis subsystem:

- Collecting measurement, fault, status, configuration, and other data from network entities & infrastructure
- Applying analytics and generating intelligence (i.e. events) for ONAP components such as Policy, APPC, and SDNC to operate upon; hence completing the ONAP's close control loop for managing network services and applications.



ONAP and Acumos Architecture



(source: https://wiki.acumos.org/)

Other groups

ITU-T

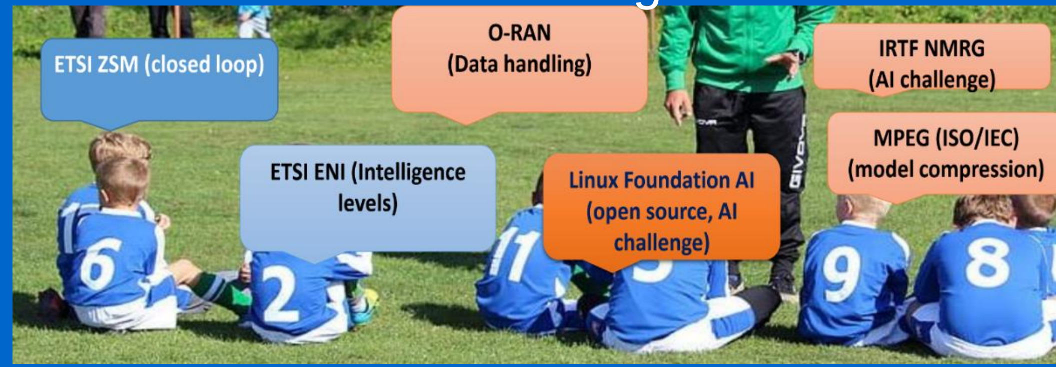
TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU

Series Y
Supplement 55
(10/2019)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

ITU-T Y.3170-series – Machine learning in future networks including IMT-2020: Use cases

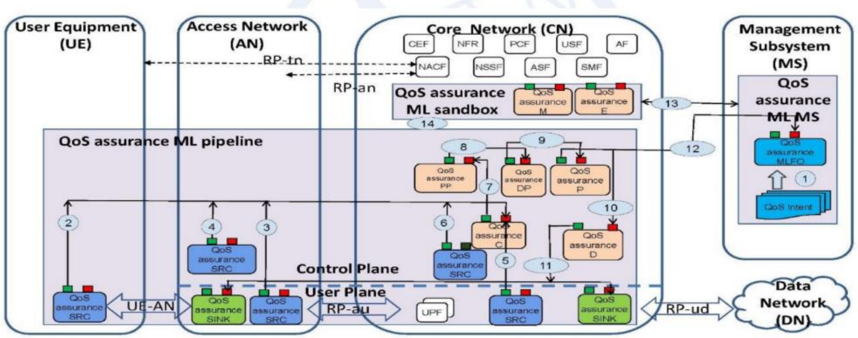
ETSI Zero Touch Network and Service Management



Q6/13 achievement: Y.3175 [consented on 13 March 2020]

"Functional architecture of machine learning based quality of service assurance for the IMT-2020 network"

- It applies the high level architecture defined in [ITU-T Y.3172] to fulfil the requirements for ML based QoS assurance for the IMT-2020 network



- QoS assurance Service for egress
- QoS assruance Service for ingress

SRC: QoS source of data
C: QoS data collector
PP: QoS data pre-processor
DP: QoS anomaly detection and prediction
SINK: QoS assruance target of ML output
P: QoS policy
D: QoS policy distributor
M: QoS assurance ML modelling
E: QoS assurance model evaluating

Architecture of ML based QoS assurance

ITU-T

Technical Specification

TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU

(01/2019)

ML Pipe Line, ITU-T Focus group on Machine Learning for Future Networks including 5G (FG-ML5G) proposes a set of logical entities to be chained in order to provide the AI functionalities and closed loop control.

Focus group on Machine Learning for Future Networks including 5G (FG-ML5G)

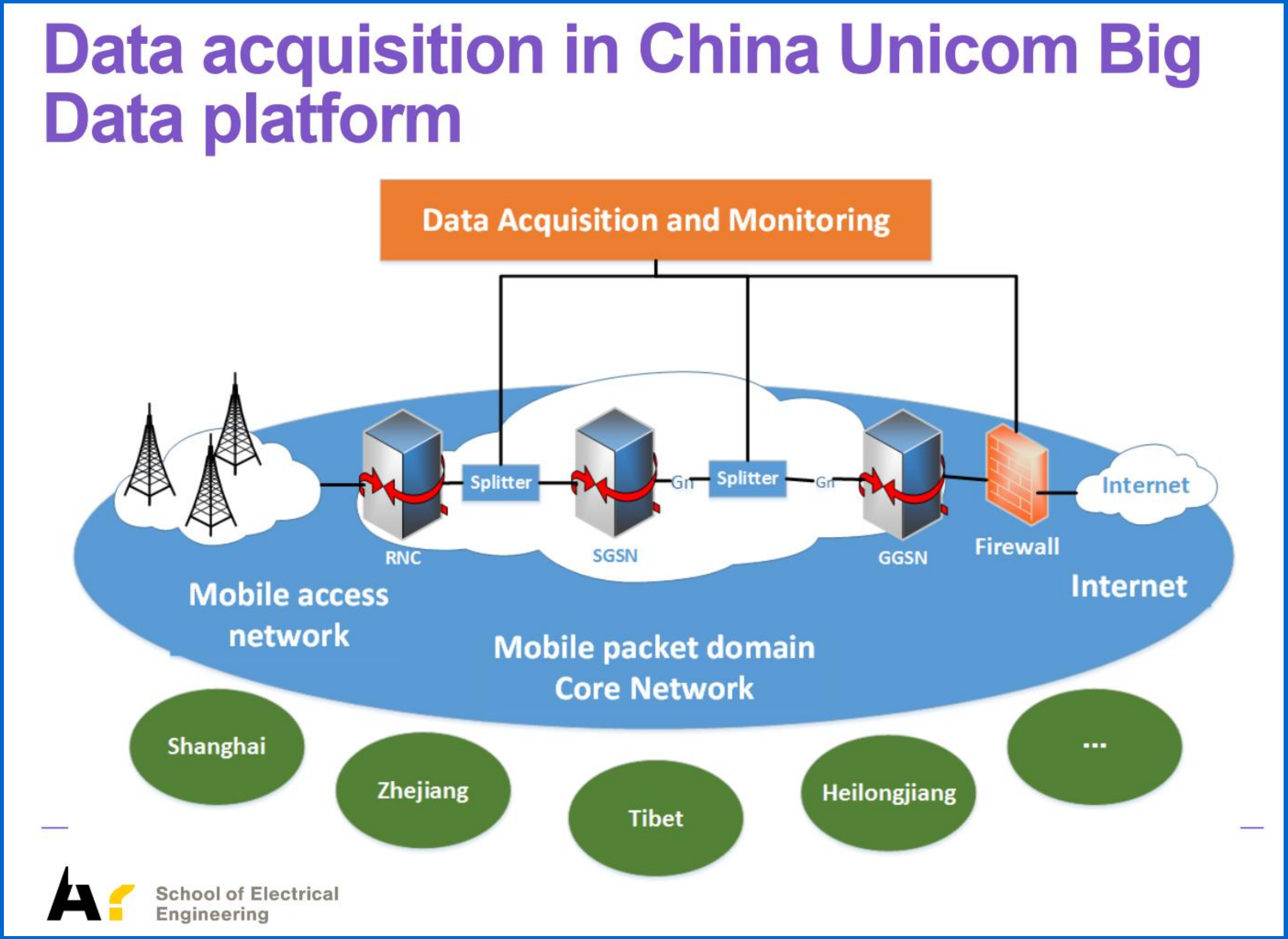
FG-ML5G-ARC5G

Unified architecture for machine learning in 5G and future networks

Orange Experience

- Develop internal runtime data platforms and external design platforms (e.g. Acumos) for more “industrial” experimentation of AI on actual operational data
- Positioning of AI use cases in Open Network Automation Platform (ONAP)
  - Data Collection, Analytics and Events (DCAE) to embed knowledge creation and decision support use cases
  - Closed Loop Automation Management Platform (CLAMP) for decision making use cases
  - Possible role of AI also in the update of policies as well as service orchestration
- Major general obstacles to implementation of AI use cases in networks
  - Dependence on data, lack of labelled data, readability and debuggability, …
  - Difficulty to match business objectives, optimization objectives of the AI model, and relevant data
- Dedicated software platforms with suitable APIs are necessary
  - To design, train, deploy and compose the machine learning/AI-based models
  - Such AI software platform(s) (e.g. Acumos) have to be combined / plugged with the runtime automation framework (e.g. ONAP) for effective use case implementation

Big Data Analysis in China Unicom



Big Data Analysis in China Unicom

Data acquisition: some numbers

Traffic acquisition from Gn point

- More than 100 GGSN elements in the network

Traffic acquisition capture all IP packets and aggregates each user packets => valid user data.

Data is collected to 200Mb packets, each containing 700000 records

Data files containing records are transmitted to data warehouse in Beijing

Traffic record format 1/2

Number	Field	Remark
1	Cell phone number	Not contain a prefix such as +86, 0086, 86
2	Location area code	LAC
3	CI number	Select the first CI when a network switches
4	Terminal type	IMEI
5	Traffic type	
6	Start time	YYYY-MM-DD HH:MM:SS.1234567, Accurate to 0.1 Microsecond
7	End Time	YYYY-MM-DD HH:MM:SS.1234567, Accurate to 0.1 Microseconds
8	Duration (in seconds)	
9	Upstream traffic (in bytes)	
10	Downstream traffic (in bytes)	
11	Total traffic (in bytes)	
12	RAT type	1 represents 3G, 2 represents 2G

Traffic record format 2/2

13	Terminal IP	
14	IP Visited	Not exist the IP information: null, Multiple IP information or the record of combined traffic: The first IP
15	Status code	
16	**User agent**	**Collect all information**
17	APN	3gwap,3gnet,uniwap,uninet,cmwap,cmnet
18	IMSI	
19	SGSN IP	The first access IP
20	GGSN IP	
21	**Content-type**	
22	Source port	
23	Destination port	
24	Record logo	0: The records unconsolidated and not split. 1: The records consolidated but not split. 2: The records unconsolidated but split. 3: The records consolidated and split.
25	Merge records	1,3: The number of combined records 0,2:null
26	URL/feature information	Business with URL/URI: the information of URL/URI Business without URL/URI: Specific information

Aalto University
School of Electrical Engineering

Two Examples from NEC

5G network slice broker

- Mapping per slice service requirements onto available resources
- Reinforcement learning just adds a small component to the overall solution
- Implemented on top of commercial components
- Published at Infocom 2017
  - V. Sciancalepore, K. Samdanis, X. Costa-Pérez, D. Bega, M. Gramaglia, A. Banchs: *Mobile traffic forecasting for maximizing 5G network slicing resource utilization*

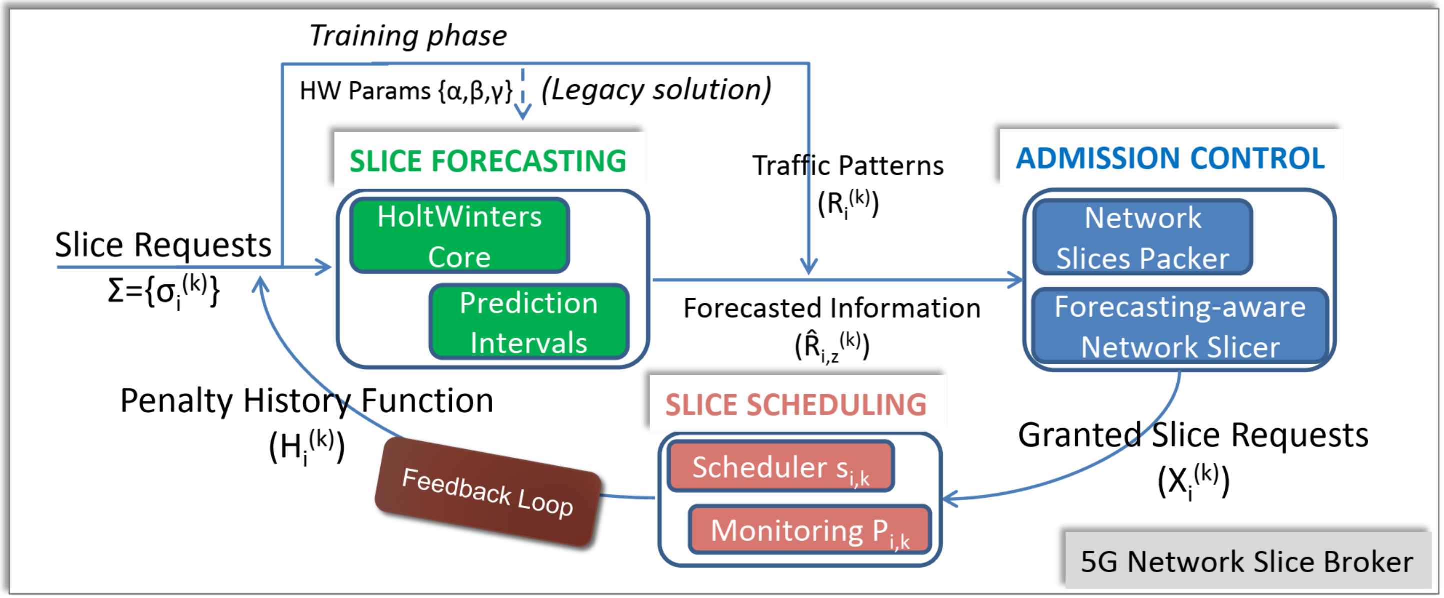
Net2Vec telecom carrier analytics system

- AI engine for various analytics and prediction applications
- Uncovers the vast potential of AI in this area
- Components of NEC's network management solutions
- Example application: Development of user profiling application

5G Network Slice Broker

5G Network Slice Broker features:

- Resource monitoring: e.g., resource blocks, MCSs
- Machine Learning operations for traffic forecasting: online reinf. learning
- Admission Control for network slice requests (based on forecasting info)
- Support for multiple classes of Network Slices SLAs
  - Heterogeneous QoS traffic requirements (data rate, latency, …)



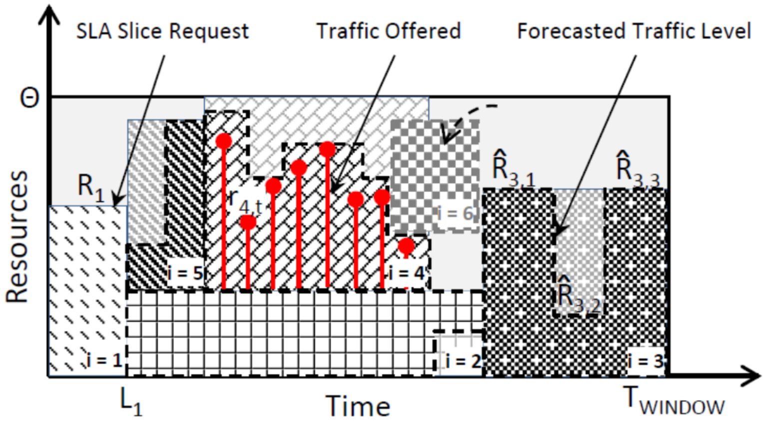
Maths..

Admission Control:

- *Geometric Bin-Packing* problem:
  - Maximizing the overal system resource utilization
  - Optimization problem MILP → high complexity

Problem ADM-CONTROL:

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in \mathcal{I}} c_i \cdot x_i \\ \text{subject to} \quad & \sum_{i \in \mathcal{I}} w_i \cdot x_i \leq W; \qquad \text{(relaxed)} \\ & \mathcal{S}(x_i) \cap \mathcal{S}(x_k) = \emptyset, \quad \forall i \neq k; \\ & \mathcal{S}(x_i) \subset \mathbb{S}, \quad \forall i \in \mathcal{I}; \\ & x_i \in [0,1], \quad \forall i \in \mathcal{I}; \end{aligned}$$



Slice Traffic Scheduling

- Minimizing the traffic scheduled per slice (while meeting the QoS constraints) in order to leave more room for other network slices.

Problem SLICER-SCHEDULING:

$$\begin{aligned} \text{minimize} \quad & s_{i,j}^{(k)} \\ \text{subject to} \quad & \left(\sum_{j=zk+\bar{t}}^{zk+\bar{t}+T^{(k)}} s_{i,j}^{(k)} \right) \geq r_{i,z}^{(k)} x_i^{(k)}, \ \forall z \in [0, \lceil \tfrac{L_i}{T^{(k)}} \rceil - 1]; \\ & \sum_{i \in \mathcal{N}} s_{i,j}^{(k)} \leq \Theta + P_{i,j}^{(k)}, \quad \forall j \in \mathcal{L}; \\ & s_{i,j}^{(k)} \in \mathbb{R}_+, \quad \forall i \in \mathcal{N}, j \in \mathcal{L}, k \in \mathcal{K}; \end{aligned}$$

Mixed traffic classes with different QoS requirements:

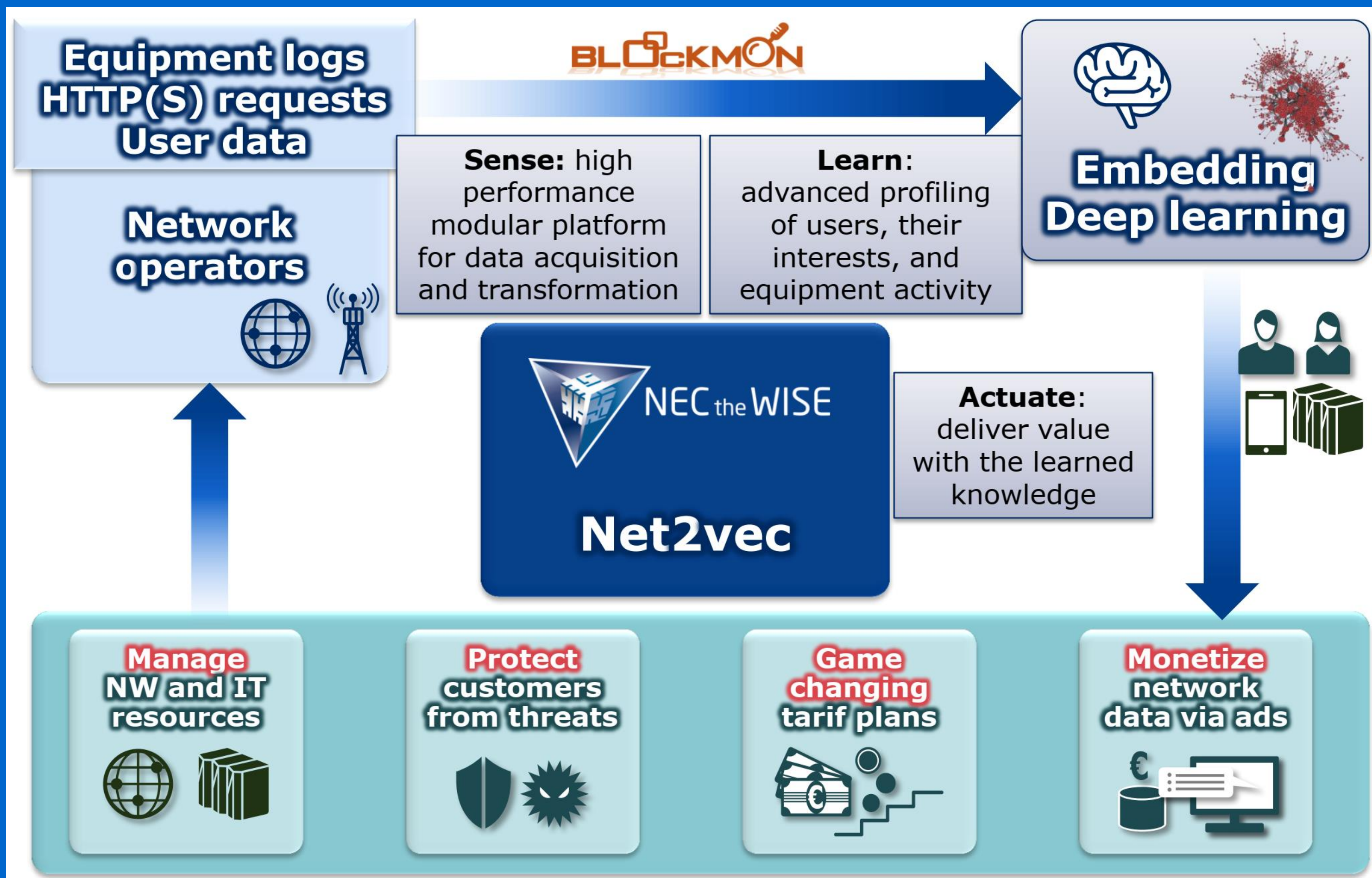
- Mission Critical (guaranteed bit rate);
- Interactive Gaming;
- TCP-Based (FTP, e-mails, p2p).

Monitoring and Feedback

- Holt-Winters technique to predict traffic pattern for pair {tenant, traffic class}.
- Online Reinforcement Learning to dynamically adjust the forecasting interval accuracy:

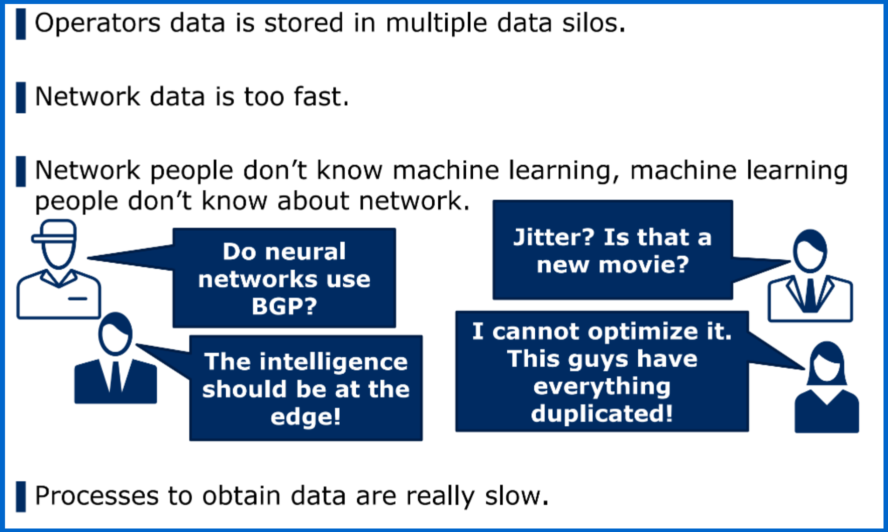
$$\boxed{\chi_i^{(k)} : h_i^{(k)} \Omega_\chi \sqrt{Var(e_{i,z}^{(k)})} = \hat{d}_i^{(k)}}$$

Net2vec: AI Platform for Carrier Analytics



Summary: AI is becoming easier..

Challenges



- AI algorithms may produce incorrect results in unexpected ways
- Defining best way to visualize how network services are provided and managed to improve network maintenance and operation
- Providing an **experiential architecture** that uses AI to improve its understanding of the environment and operator experience, over time

Open Problems

- These are early days in application of AI in networks
- Few real AI network applications in operation
- Very thin experience base in networking, data collection, and algorithm design
- High barrier in going from algorithm to operations
- Focus on training, deployment & operation, and continuous learning