# On 6G-enabled SDN-based mobile network User Plane with DRL-based Traffic Engineering (draft)

Robert Kołakowski*◇ [0000-0002-5451-8847], Lechosław Tomaszewski* [0000-0003-3836-7900], Sławomir Kukliński◇
*Orange Polska, ◇Warsaw University of Technology

**Abstract**

The emerging 6G use cases will pose new challenges for the mobile network User Plane (UP), requiring its rapid evolution in terms of flexibility and intelligent optimisation. To achieve the foremost, the exploitation of the Software-Defined Networking (SDN) concept is commonly considered due to the logically centralised network control and native support for Traffic Engineering (TE). A promising solution to embed intelligence in the network is using Deep Reinforcement Learning (DRL) methods, which are capable of flexible optimisation of complex environments without prior modelling. While there exist several state of the art concepts combining the above technologies pair-wise, there is no approach that integrates them into a unified 6G-ready solution. This paper presents the novel 3GPP-compliant SDN-based UP architecture enhanced by DRL-based TE to facilitate emerging 6$^{\text{th}}$ Generation (6G) use cases. The approach leverages hierarchical architectures to improve the scalability of operations, support decentralised 6G network deployments and enable DRL usage in carrier-grade mobile networks.

**Index Terms**

beyond 5G, 6G, User Plane, SDN, DRL, AI, UPF, traffic engineering, QoS, user-centricity, TN, NTN

## Acronyms

The following acronyms are used in this manuscript:

| | |
|---|---|
| 3GPP | 3$^{\text{rd}}$ Generation Partnership Project |
| 5G | 5$^{\text{th}}$ Generation |
| 5GS | 5G System |
| 5QI | 5G QoS Identifier |
| 6G | 6$^{\text{th}}$ Generation |
| 6G-UPF | 6G User Plane Function |
| AI | Artificial Intelligence |
| AMF | Access and Mobility Management Function |
| B5G | Beyond 5G |
| BAR | Buffering Action Rule |
| BN | Border Nodes |
| CAGR | Compound Annual Growth Rate |
| CN | Core Network |
| CP | Control Plane |
| CSPF | Constrained Shortest Path First |
| CU | Centralised Unit |
| CUPS | Control and User Plane Separation |
| DMOC | Domain Management and Orchestration Component |
| DNN | Deep Neural Network |
| DP | Data Plane |
| DRL | Deep Reinforcement Learning |
| DRLA | Deep Reinforcement Learning Agent |
| DT | Digital Twins |
| E2E | End-to-End |
| E2EF-DRLA | E2E Flow-DRLA |
| E2EMO | E2E Management and Orchestration |
| E2ENO | E2E Network Orchestrator |
| E2ESO | E2E Service Orchestrator |
| EU | European Union |
| F-DRLA | Flow-DRLA |
| FAR | Forwarding Action Rule |
| GBR | Guaranteed Bit Rate |
| gNB | Next Generation NodeB |
| GTP | GPRS Tunneling Protocol |
| HITL | Human In The Loop |
| HRLLC | Hyper Reliable Low-Latency Communication |
| INC | In-Network Computing |
| IP | Internet Protocol |
| ITU-R | International Telecommunication Union – Radiocommunication Sector |
| KPI | Key Performance Indicator |
| MANO | Management and Orchestration |
| MAPE-K | Monitor-Analyse-Plan-Execute based on Knowledge |
| MAR | Multi-Access Rule |
| MDP | Markov Decision Process |
| ML | Machine Learning |
| MNO | Mobile Network Operator |
| MSDNO | Master SDNO |
| mUPF | master-UPF |

ETHER project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101096526. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

| | |
|---|---|
| NF | Network Function |
| NGMN | Next Generation Mobile Networks |
| NOD | Network Orchestration Domain |
| NODI | NOD Infrastructure |
| NS | Network Slicing |
| NTN | Non-Terrestrial Network |
| OF | OpenFlow |
| OVS | Open vSwitch |
| OVSDB | OVS Database |
| P4 | Programming Protocol-independent Packet Processors |
| PDR | Packet Detection Rule |
| PDU | Protocol Data Unit |
| PFCP | Packet Forwarding Control Protocol |
| POMDP | Partially Observable Markov Decision Process |
| QER | QoS Enforcement Rule |
| QoE | Quality of Experience |
| QoS | Quality of Service |
| RAN | Radio Access Network |
| RL | Reinforcement Learning |
| SD-WAN | Software-Defined Wide Area Network |
| SDN | Software-Defined Network |
| SDN-UP | SDN-based UP |
| SDNC | SDN Controller |
| SDNO | SDN Orchestrator |
| SLA | Service Level Agreement |
| SMF | Session Management Function |
| SOD | Service Orchestration Domain |
| SODI | SOD Infrastructure |
| SotA | State of the Art |
| SRR | Session Reporting Rule |
| SUP-DTE | SDN-UP with DRL-TE |
| sUPF | sub-UPF |
| sUPF-A | sub-UPF Agent |
| sUPF-M | sub-UPF Manager |
| sXXR | sub-XXR |
| TE | Traffic Engineering |
| TE-A | TE Analytics Service |
| TE-C | TE Coordination Service |
| TE-D | Domain-level TE |
| TE-E | TE Execution Service |
| TE-EXP | TE Exposure Service |
| TE-G | Global-level TE |
| TE-M | TE Monitoring Service |
| TE-P | TE Planning Service |
| TE-R | TE Reconfiguration Service |
| UE | User Equipment |
| UP | User Plane |
| UPF | User Plane Function |
| URR | Usage Reporting Rule |
| VNF | Virtual Network Function |
| WIM | WAN Infrastructure Manager |
| XR | Extended Reality |
| XXR | Flow Processing Rule |

## I. Introduction

Software-Defined Network (SDN) is commonly considered an important component for modern telecommunication networks due to forwarding flexibility, fine-grained network traffic control, and facilitation towards application-driven Traffic Engineering (TE) that altogether constitute the overall service agility. Despite the technology potential and increased interest in the SDN market (the growth estimated at 47.3% Compound Annual Growth Rate (CAGR) [1] in years 2020-2031), the network-wide carrier-grade exploitation of SDN in the mobile User Plane (UP) is still at the infancy stage, mainly due to scalability issues.

Simultaneously, the 6G research accelerates. While the standardisation is at a very early stage, with only initial IMT-2030 document releases [2] and the first 3rd Generation Partnership Project (3GPP) study forecasted for 2025 [3], the 6G use cases such as holographic telepresence, Extended Reality (XR), robot/cobot network fabrics, or Non-Terrestrial Network (NTN)-driven applications [4], and 6G features have started to emerge. These will pose new UP requirements in terms of traffic control flexibility and granularity, user/application mobility and resilience. Moreover, intelligent, scalable and automated operation and optimisation are needed to address the rising networks' complexity and contribute to 6G targets. Given these circumstances, a promising solution to implement intelligence in UP is Deep Reinforcement Learning (DRL), which involves intelligent decision-making agents learning via interactions with the environment. While being a relatively old concept, the DRL was usually used for time-invariant tasks in stationary environments [5] i.e., the problems in which the isolated environment is considered and principles are not subjected to change during the agent operation. Despite successful DRL applications to continuous time-variant networking problems, their usage in carrier-grade networks poses several challenges, such as in-training violations, convergence, coordination of agents, etc.

The goal of the paper is to explore the perspectives and challenges regarding the exploitation of DRL-based TE in 6G SDN-based UP (SDN-UP). First, presented are the recent State of the Art (SotA) regarding the 3GPP standardisation (cf. Section II), Beyond 5G (B5G) SDN-UP and DRL-TE (cf. Section III) and key challenges regarding DRL exploitation in SDN-UP (cf. Section IV). Based on the analysis, a novel 3GPP-compliant SDN-UP architecture integrated with DRL-based TE is proposed, called SUP-DTE (Section V). Finally, the key benefits of the approach, remaining open issues, and solutions (cf. Section VI) are outlined.

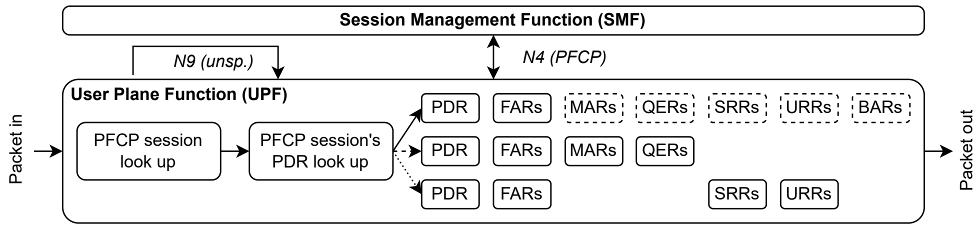


Fig. 1. 3GPP 5th Generation (5G) UPF packet forwarding model [7]

## II. User Plane in 5G System

The 5G System (5GS) architectural design is based on the fundamental principle of Control and User Plane Separation (CUPS)[6]. In the 5GS Core Network (CN), the user data flow and processing support is nested within the generically defined User Plane Function (UPF), which is terminated at Radio Access Network (RAN) node at one side and at the Internet Protocol (IP) data network at the other. UPF functionality can be implemented with several Network Functions (NFs), e.g., in the case of a roaming architecture or its implementation in multiple operator domains. Unlike in previous generations UP, UPF is not just a mobile user's packets pipe anchored at a fixed IP network point of presence. Now, according to the Network Slicing (NS) concept, it can include flexibly designed service-specific user traffic processing chain, e.g., packets inspection, classification, routing, marking, network address translation, etc., formerly supported within the "vestibule" between the anchor point of the user data tunnel and the IP network.

UPF embeds the control mechanisms to be exposed to the 5G Control Plane (CP) via the N4 interface [6], according to the CUPS principle. The UPF direct counter-partner in CP is Session Management Function (SMF) (except UPF monitoring events information exposure to other NFs) acting as a proxy for interactions with CP NFs, e.g., in the procedures of user data session establishment, termination, and modification (also for User Equipment (UE) handovers handling), Quality of Service (QoS) management, etc. For the latter, 5GS supports multiple parameters, which include flow type: non-, ordinary, or delay-critical Guaranteed Bit Rate (GBR), flow priority level, packet delay budget (including CN one), packet error rate, averaging window (for ordinary GBR and delay-critical GBR only), maximum data burst volume (for Delay-critical GBR only). The 5G QoS characteristics based on the above parameters are indicated through 5G QoS Identifier (5QI) classes. Other features are also supported, i.a., allocation and retention priority, reflective QoS, maximum (per-session, per-UE, per-UE+per-slice) bit rates (per-flow, aggregate) and packet loss rates.

The N4 interface implements the Packet Forwarding Control Protocol (PFCP) used by SMF to control the packet processing within UPF for each individual Protocol Data Unit (PDU) session via PFCP session. For each incoming packet, a look-up of a relevant PFCP session and then of a proper Packet Detection Rule (PDR) among ones provisioned for that session is performed (cf. Fig. 1). When a matching PDR is found, the instruction set is applied to a packet, which may include: Forwarding Action Rule (FAR), Buffering Action Rule (BAR), QoS Enforcement Rule (QER), Usage Reporting Rule (URR), or Multi-Access Rule (MAR). Additionally, a Session Reporting Rule (SRR) can request UPF to detect and report PFCP session events unassociated either with specific PDRs, or traffic usage measurements, e.g., per QoS flow and per UE QoS measurements.

The 5G UP design incorporates some of the SDN concepts (CUPS, packet processing approach, or programmatic control), but it lacks full alignment with SDN and most popular protocol [8], i.e., OpenFlow (OF) [9]. While OF can be used to implement some of the UPF functionalities, including PDR (packet matching to flow entries in flow tables), FAR (actions associated with flow entries), URR and SRR (by per-flow statistics provisioning and configuration of OF switch) there exist deficiencies regarding other rules. Primarily, the OF lacks QER support as it only allows mapping flow to the in-switch queues using identifiers – the queue configuration (priority, scheduling algorithm, maximum rate, etc.) cannot be set by OF. Also, packet delay measurement is not supported, further complicating 5QI enforcement. Finally, as OF lacks buffering control capabilities, implementation of BAR requires external support.

Nonetheless, the SDN can be advantageous for 6G UP. The primary benefits include the logically centralised and programmable control (enabling flexible path establishment and manipulation), the ability to create multiple paths (e.g., for seamless connectivity, increased resilience, etc.), mobility of two path ends (support for UE/application mobility), fast path operations (support for high mobility scenarios, e.g., NTN-based), per-flow granularity (e.g., per-UE), and ability to leverage external information on the network (predicted topology changes, network failures, etc.). Also, the SDN programmability allows for intelligent network optimisation via TE mechanisms, which enable load balancing, minimisation of congestion, E2E delay, packet loss and energy consumption, QoS/Quality of Experience (QoE) maximisation, or optimising resource utilisation. Centralisation, however, reduces scalability and raises the issues of a single point of failure, SDN Controller (SDNC) placement, and SDN nodes' synchronisation. However, the non-unified approach to traffic management and optimisation increases complexity, lowers inter-operability in integrated environments (e.g., using external solutions to connect data centres) and limits possibilities to implement End-to-End (E2E) QoS-aware TE on a per user basis. The E2E TE will require counterparts in network segments belonging to Mobile Network Operator (MNO) and external providers. The latter can implement TE mechanisms conflicting with UP QoS requirements. Such strong coupling can hamper the E2E TE in multi-domain and multi-stakeholder 6G UP.

## III. Related work

While for today, the 6G UP definition is vague, the envisioned 6G features can provide insights into potential directions of UP evolution. The International Telecommunication Union – Radiocommunication Sector (ITU-R) released initial materials on the trends for IMT-2030 system [2], [10], drafting six usage scenarios, namely, Immersive Communication, Artificial Intelligence (AI) and Communication, Hyper Reliable Low-Latency Communication (HRLLC), Ubiquitous Connectivity, Massive Communication, Integrated Sensing and Communication, falling under the overarching aspects of sustainability, connecting the unconnected, ubiquitous intelligence, security, and network resilience. The IMT-2030 technical performance requirements are expected to appear by Q1 2026 [11]. Nonetheless, academia proposes many UP concepts targeting the above use case categories and core 6G values. The 6G-enabled UP that introduces functional UPF split into CN, RAN, and common sub-functions has been proposed [12]. The E2E UP is formed by chaining elements composed of sub-functions controlled by SMF or Centralised Unit (CU). An "organic network" concept has been presented [13] where CP procedures are grouped into functionalities exposed to the UE and executed as processes. The approach shrinks CP interactions (e.g., Access and Mobility Management Function (AMF)/SMF/UPF reselection) allowing UPF implementation as a traffic control application atop forwarding fabric.

In the context of 6G UP, the incorporation of SDN with a different scope of responsibilities and positioning (SDNC atop or below NFs) is commonly considered. In [14], an SDNC acts as the centralised coordinator of all UPFs (OF-based switches) including Next Generation NodeBs (gNBs), which are connected with OF switches to enable E2E flow management. The Open vSwitch (OVS)-based UPF facilitating network slicing has been proposed [15]. UP slices are created using VLAN tunnels (instead of GPRS Tunneling Protocol (GTP)/IP) leading to lower and more stable latency. The Aether platform introduced Programming Protocol-independent Packet Processors (P4)-UPF [16], in which the SDNC operates below the UPF abstraction and executes commands of the upper-level applications. 3GPP compliance is ensured by PFCP agent acting as the CP termination and UP4 App translating the CP requests into SDN operations. The Intelligent UP (IUP), composed of switches with compute and data processing capabilities enabling In-Network Computing (INC), has been presented [17]. The concept adopts the OF-compliant approach by proposing Communication and Compute Flows to handle network and QoS management on a flow basis.

The autonomous UP optimisation requires intelligent TE. While classical TE in centralised SDN is well addressed [18], the AI-driven TE, especially DRL-based, is still investigated. The DRL-based framework for SDN traffic control enhanced with TE-aware exploration (using knowledge from baseline TE methods) and prioritised replay has been proposed to improve convergence and mitigate high-risk exploration [19]. The DRL-based routing for centralised SDN [20], which considers path-state metrics to adapt to variable traffic conditions, was introduced. The SDN-DRL self-management architecture has been presented [21], which optimises network services to meet QoS demands. Deep Reinforcement Learning Agents (DRLAs) are deployed across network sites, providing routing policies per site for each traffic class using monitoring data. A multi-agent DRL TE framework for a hybrid SDN has been proposed, in which DRLAs generate the routing policy for the SDN switches while legacy routers implement proprietary TE algorithms [22]. The hierarchical multi-controller SDN framework enhanced with TE, called Helix, has been presented [23]. The solution outperforms baseline TE solutions such as Constrained Shortest Path First (CSPF) routing, while improving system scalability, especially in distributed SDN.

While the above solutions involve the pivotal 6G UP technologies, there exists no approach that combines UP, SDN, and DRL-based TE into a unified E2E framework, while considering the critical issues, such as scalability and flexibility of operations, adaptability or coordination support.

## IV. DRL-based TE in SDN-UP

DRL is a branch of Machine Learning (ML) combining Reinforcement Learning (RL) and deep learning to provide decision-making for complex environments. A typical DRL setup (cf. Fig. 2) comprises the Deep Neural Network (DNN)-based Agent – the approximator for policy, value or both policy and value functions; and the Environment – the external system the Agent interacts with. The DRL problems are typically formulated as Markov Decision Process (MDP) [24] $M = (S; A; T; R)$, where $S$ is a set of environment states, $A$ a set of agent's actions, $T$ transition probability from state $s_t$ to $s_{t+1}$ at time $t$ under action $a_t$, and $R$ a reward function for transitioning between states. The Agent's goal is to learn the optimal policy $\pi : S \times A \to [0, 1]$, maximising the total reward over time. The DRLA learns through feedback by: i) obtaining the information on the current environment state $s_t$ and reward $r_t$ obtained for the last transition; ii) selecting the action $a_t$ based on $s_t$ using $\pi$; iii) executing $a_t$ leading environment to transition to $s_{t+1}$ (with reward $r_{t+1}$). The history of transitions, actions and rewards is used to iteratively improve DRLA policy $\pi$.

While DRLAs are successfully used in multiple fields requiring negotiation or dynamic decision-making (e.g., robotics, healthcare), DRL suffers from several issues, complicating its massive implementation in mobile networks. The key challenges concerning DRL-based TE in SDN-UP are described below.

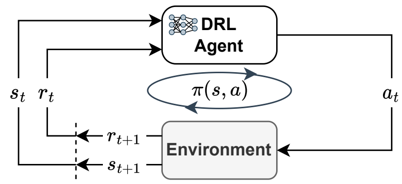


Fig. 2. Typical DRLA setup

**Sample efficiency:** As the DRLA knowledge is built upon experiences (i.e., visited states, executed actions, and their outcomes); improving sample efficiency to reduce training time is vital [25]. In conventional OF-based SDN setups, the environment state information is obtained via SDN CP channel [9], lowering SDNC's flow handling capacity and limiting the maximum Data Plane (DP) sampling frequency While particular solutions to improve data efficiency have been proposed, e.g., by increasing DRL model's comprehension of inter-dependencies among states [26], or long-horizon task splitting into shorter ones during training [25], their impact in the SDN context is not assessed, yet.
**Reward design:** Full flexibility of reward function definition, while highly beneficial in multi-criteria optimisation, requires prior assessment of potential undesired outcomes. In complex constrained environments, such as SDN-UP, the DRLA optimisation of network aspects can lead to the degradation of another Key Performance Indicator (KPI) (e.g., load distribution vs resource utilisation) and lead to potential issues such as decreased fairness across users or prioritisation of certain traffic types. Moreover, the DRL reward is intractable to change – any modification requires the deployment of a DRLA trained using the new reward definition. The promising solutions in the SDN-UP context are meta-learning approaches [27], which can be effectively used to coordinate the training across a set of DRL agents targeting different objectives [28].
**Transferability:** The DRLA policy (DNN model) is dependent fully on the DRLA setup – environment, available actions, reward function, etc. As the output policy is optimised in the context of specific MDP settings, the reusability of the DNN model to solve other MDPs is largely limited. Transferability is essential for SDN-UP to facilitate the creation of DRLAs for specific network segments (e.g., transport) or service types and support coordinated training (e.g., federated learning). Transferability can be improved by, i.a., reward shaping, learning from samples external to the environment, partial policy transfer, using mapping functions between source and target domains, or representations transfer [29]. A promising actor-critic approach has been proposed [30], in which the network graph embedding is used by a single-input DNN to evaluate routing decisions. A good generalisation for many network topologies allows transferring the critic model.
**Generalisation:** DRL generalises poorly on samples outside the training distribution, especially for continuous states and actions. Hence, in sensitive environments such as SDN-UP, it is vital to establish a relevant environment model and comprehensive training procedure to mitigate performance degradation during operation. It has been proved that generalisation can be largely improved by using representation learning for states [31] or both actions and states [32].
**Robustness:** Achieving DRL robustness to faults during long-term optimisation of real-life systems is problematic as the non-stationary environment can destabilise DRLA policy. A recently emerging field of Continual RL studies the aspects of joint agent and environment evolution to mitigate devastating phenomena such as, e.g., catastrophic forgetting [33]. The field is still forming, however, and lacks mature solutions that could be used to improve the DRLAs' robustness to failures, especially in SDN-UP environments.
**Observability:** MDP requires full observability of the environment states, which is often not feasible in real-life applications. In the SDN-UP, obtaining the actual environment state is impacted by latencies caused, i.a., by the spatial distribution of devices or SDNC placement. A potential solution to this issue involves Partially Observable Markov Decision Process (POMDP) in which the DRLA acts upon a belief state (a probability distribution over possible environment states). The POMDP problems, however, are largely intractable to be solved optimally [34] and suffer from poor convergence.
**Multi-stakeholder environment:** The 6G network complexity will require vast numbers of intelligent agents to optimise different network aspects. To maintain the stable operation of the network, advanced DRLAs coordination will be vital. Implementation of such mechanisms in a multi-stakeholder environment, e.g., cooperative learning support [35], while considering Service Level Agreement (SLA) constraints, remains a challenge.
**Interpretability:** DRL agents are implemented using DNNs, giving limited possibilities to track the reasoning and decision logic, giving limited visibility on causes of failures or fairness infringements. These come as a major drawback in the context of the recent EU legislation regarding the exploitation of AI-based technologies (i.e., the AI act) [36], which defines the set of requirements for the *high-risk* AI-based applications. While the telco systems are not directly listed in the document, they can be a part of the chain constituting such applications (e.g., critical infrastructure management), thus requiring compliance to the rules, also on the UP level. While some solutions improving DRL explainability and interpretability have already been proposed by academia [37], no work exists on their exploitation in SDN-UP TE context.

While specific solutions have already been proposed to the above-stated challenges, there exists no unified approach supporting their exploitation in SDN-UP. The architecture proposed in Section V aims to bridge this gap, as well as address the key SDN-UP and TE issues.

## V. INTEGRATED SDN-UP-DRL-TE FRAMEWORK

Tackling key DRL limitations and facilitating its exploitation for TE is crucial in terms of establishing intelligent, autonomous, and optimised future-proof UP. Hence, hereby, proposed is a 6G-enabled SDN-UP with DRL-TE (SUP-DTE) framework, supporting the envisioned multi-domain, multi-stakeholder and 3D character of the 6G network. The architecture design follows the principles proposed by the European Union (EU) ETHER project vision [38]. The cognitive capabilities in SUP-DTE are provided by the DRL-based TE performing network optimisation from both user and network operator perspectives. SUP-DTE comprises five hierarchical frameworks responsible for: i) SDN-based network orchestration; ii) service orchestration; iii) data-path setup and control – 6G User Plane Function (6G-UPF); iv) DRL-TE; v) TE data and DRL-TE models exchange between the network operator and external stakeholders – AI Layer (cf. Fig. 3). Each system operates at both domain and E2E levels.

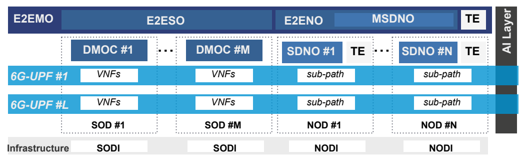


Fig. 3. The high-level view of SUP-DTE components

Two types of domains are considered: Service Orchestration Domains (SODs), devoted to the management and orchestration of services, i.e., Virtual Network Functions (VNFs); and Network Orchestration Domains (NODs) devoted to the establishment of connectivity between VNFs (both inter- and intra-SOD). Both NODs and SODs encompass different technological or administrative domains (transport, edge, RAN, etc.) in Terrestrial, Aerial or Satellite network segments and handle the underlying NOD Infrastructure (NODI) and SOD Infrastructure (SODI) resources. The E2E Management and Orchestration (E2EMO) in the system is handled by E2E Service Orchestrator (E2ESO) and E2E Network Orchestrator (E2ENO) that contains Master SDNO (MSDNO) – the supervisor of a federation of SDN Orchestrators (SDNOs). Both E2ESO and E2ENO delegate the UP orchestration processes to domain-level orchestrators Domain Management and Orchestration Components (DMOCs) (e.g., ETSI MANO [39]) and SDNOs (cf. Fig. 5).

In SUP-DTE, the UP is handled by a 6G-UPF, which exploits the MSDNO/SDNO interfaces to establish, maintain and reconfigure network paths (and their parts in NODs, further called *sub-paths*) according to the session-level configuration provided by the network CP (i.e., SMF). Both 6G-UPFs and SDN orchestration components are integrated with the DRL-based TE to allow concurrent and coordinated network and path optimisation. To make UP *user-centric*, 6G-UPF instances are deployed for individual PDU sessions. 6G-UPF implements a hierarchical structure (cf. Fig. 4) in which, master-UPF (mUPF) controls a set of sub-UPFs (sUPFs) to enforce path configuration on both E2E and domain levels. To ensure compliance with the 3GPP QoS model, mUPF contains the *PFCP agent*, which terminates the N4 interface and translates PFCP messages and session configuration, i.e., Flow Processing Rules (XXRs), to atomic UP operations. The CP commands are delivered by sub-UPF Manager (sUPF-M) to respective sUPF managers, i.e., sub-UPF Agent (sUPF-A), and MSDNO for execution. The foremost occurs for intra-domain UE mobility. On CP request, sUPF-M sends path-related commands to the sUPF-A, which executes them via SDNO interfaces. In the case of inter-domain mobility, sUPF-M triggers the E2ENO processes to both reconfigure the E2E path as well as orchestrate sUPFs to handle newly established sub-paths. sUPF-M is also responsible for the aggregation of sub-paths KPIs provided by sUPF-A to obtain E2E data and its verification against PFCP-driven session configuration (sUPF-M will trigger relevant reconfiguration actions in case of violations).

The E2E path establishment process is as follows. When UE initiates PDU establishment procedure, the E2ESO orchestrates mUPF in one of the managed SODs. The sUPF-M requests from the MSDNO to establish the abstracted E2E path composed of NODs and SODs Border Nodess (BNs) – i.e., gateways to NODs/SOD or nodes connected to the data source/sink – with session-specific requirements (QoS class, additional UP functions such as in-datapath analytics, firewalls, etc.). MSDNO splits the E2E path into sub-paths (based on the NOD membership) and delegates their orchestration to relevant SDNOs (including the information on the UP VNFs that have to be included in the sub-path). The SDNOs interact with SDNCs (SDN, or

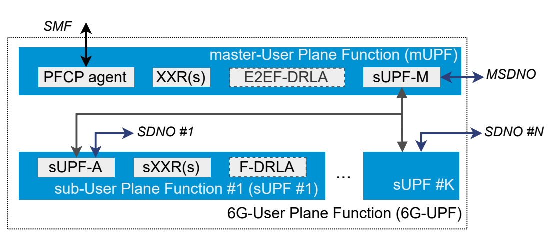


Fig. 4. 6G-UPF internals and interconnection with its sub-components, i.e., sUPFs

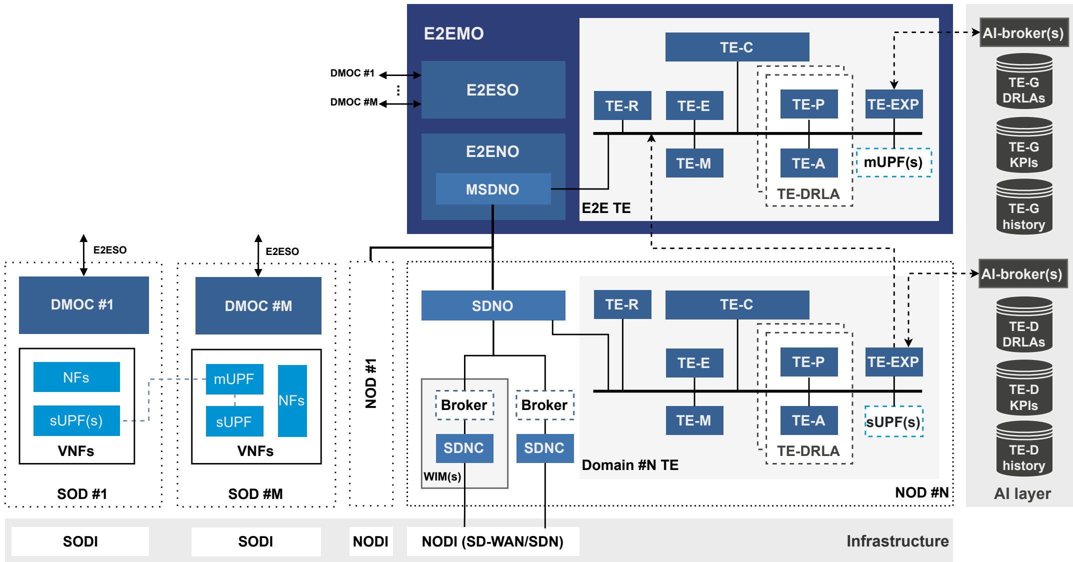

Fig. 5. The SUP-DTE architecture with 6G-UPF instances deployed in SODs and connected to TE frameworks message buses (marked with dashed line)

Software-Defined Wide Area Network (SD-WAN) in WAN Infrastructure Manager (WIM) case) or relevant brokers if the operator does not own the underlying infrastructure. When the path is established, E2ESO orchestrates sUPF for each sub-path (via DMOCs), with sub-paths metadata, privileges to connect to SDNO and NOD TE framework as well as the data needed to register each sUPF in mUPF.

The configuration for mUPF/sUPF is stored within the XXRs/sub-XXRs (sXXRs) – corresponding to 3GPP rules, e.g., URR, QER. sXXRs is a set of individual rules to be enforced by sUPFs, which contribute to overall XXR (e.g., GBR, delay). To surpass the SDN limitations described in Section II, the enhanced SDNC is envisioned, which, in addition to OF, implements parallel interfaces enabling dynamic queue and buffering management (e.g., in the case of OVS, connection to switch OVS Database (OVSDB) server). Moreover, sUPF-M and sUPF-As can interact with E2ENO and SDNOs to collect network-level metrics and suggest optimisation actions to improve QoS, based on the in-built optimisation engines E2E Flow-DRLA (E2EF-DRLA)/Flow-DRLA (F-DRLA). To this end, the 6G-UPF can be seen as the distributed agent (with sub-path granularity) responsible for user data path monitoring and management in case of QoS violations or mobility events. To decrease resource footprint, the 6G-UPF entities can be instantiated as lightweight software components (e.g., containers) or processes within MSDNO/SDNO frameworks. The 6G-UPF deployment within SUP-DTE architecture is depicted in Fig. 5.

In SUP-DTE, the network-level UP optimisation is performed by the E2E and domain TE frameworks – meshes of services devoted to network monitoring, analytics, decision-making and enforcement, data exposure or supplementary functions. The TE services are connected to the message bus and communicate using the publish/subscribe paradigm to synchronise actions across services (e.g., maintain common information state and network KPIs). Also, the TE services compose multiple parallel Monitor-Analyse-Plan-Execute based on Knowledge (MAPE-K) loops [40] concurrently optimising the network. The TE service types are described below.

**TE Monitoring Service (TE-M)** – responsible for network resources monitoring (physical or virtual links) managed by SDNO/MSDNO, assembling common environment state information (e.g., link utilisation, network topology, etc.) and KPIs calculation for both infrastructure and flows.

**TE Analytics Service (TE-A)** – providing analytics using the TE-M data, e.g., state representation recalculation, calculation of rewards for the DRL agents, flow- and network-related predictions, etc.

**TE Planning Service (TE-P)** – responsible for deriving optimisation decisions based on the analytics data regarding flows and network reconfiguration. The DRLAs comprise both TE-A and TE-P functionality.

**TE Coordination Service (TE-C)** – responsible for the coordination of actions proposed by DRLAs, TE-Ps and sUPFs/UPFs connected to the Domain/E2E TE frameworks.

**TE Execution Service (TE-E)** – translating commands/policies selected by TE-C into SDNO/MSDNO requests and their execution.

**TE Reconfiguration Service (TE-R)** – component exposing the TE framework to enable reconfiguration of TE components or network – i.e., Human In The Loop (HITL) – to mitigate instabilities or improve the DRLAs performance.

**TE Exposure Service (TE-EXP)** – a gateway to the TE framework enabling: i) connection with AI-brokers for DRLA models selection to use within TE framework and network data/KPIs/reconfiguration history monetisation/purchase; ii) network data and KPIs exposure to the E2E TE framework. Due to different dynamics and scope, dedicated data and model databases for

Domain-level TE (TE-D) and Global-level TE (TE-G) are assumed. The aspects of charging and secure data/model exposure are out of TE scope.

It must be noted that NOD and E2E TE frameworks operate with different time scales and can target conflicting optimisation objectives. Whereas the domain TE uses the link-level metrics and performs optimisation in a fast manner, it is assumed that the E2E TE focuses on NOD and metrics abstractions and slow-scale optimisation.

## VI. Benefits and open issues

The core SUP-DTE benefits in the context of 6G UP include:

- **User-centricity** – 6G-UPFs is focused on individual session, which allows for *user-centric* QoS management and rapid path reconfiguration in case of QoS degradation. Also, the fine-grained QoS monitoring (sub-path and flow granularity) is provided, which enables E2E path fine-tuning.
- **Service chaining** – SDN enables flexible path split to include in-datapath VNFs to form service chains or perform INC. Moreover, the 6G-UPF abstraction eliminates the need for GTP tunnelling to reduce related DP overhead.
- **Autonomic optimisation** – using DRLAs both at the network and user-session level facilitates autonomous QoS-aware network-wide optimisation. The TE framework can include information on the flow QoS requirements and constraints into optimisation processes.
- **Lightweightness** – the 6G-UPF lifespan equals the session duration limiting the static resource consumption during UE inactivity periods. The dynamic orchestration of 6G-UPF instances, however, will lead to increased resource consumption by Management and Orchestration (MANO). Hence, it would be necessary to empirically derive optimal 6G-UPF lifespan to mitigate resource overspending on its life cycle management operations.
- **Scalability** – hierarchical approach improves the scalability of SDN (due to SDN CP distribution) and DRL-TE (due to reduced network state sizes).
- **DRL-TE support** – the hierarchical distributed architecture contributes to: i) faster DRLAs convergence due to better observability of network (more accurate and up-to-date monitoring data due to the collocation of SDNC and DP devices) and reduced state sizes; ii) contribution to data efficiency problem due to more frequent DP sampling (i.e., smaller impact on the SDN CP) and sampling adaptation to the NOD specifics; iii) improved transferability of the DRL models across the same domain types.
- **Cooperative learning and coordination** – the SDNO acts as the umbrella abstracting the resources of SDN/SD-WAN providers. Such an approach enables the implementation of network-wide cooperative learning in TE, e.g., multi-agent DRL, without considering the provider-level optimisation mechanisms. Moreover, the inclusion of mUPF/sUPF into the reconfiguration process allows for improved optimisation decisions while conforming to session-level restrictions.

While SUP-DTE provides multiple benefits regarding the 6G UP implementation, several open issues remain to be addressed. The critical issues, together with potential solutions, are listed below.

**Country-wide deployments** – while the proposed hybrid approach significantly improves the scalability of SDN and TE, centralised E2ENO might not be scalable enough to accommodate millions of active UEs in country-wide deployments. The potential solution would be to create a mesh of interconnected SUP-DTE responsible for handling the traffic in certain parts of the country and higher-level supervising components to implement the E2E behaviour and maintain E2E QoSs.

**Network abstraction** – hierarchical approach imposes the creation of the network and KPIs abstractions that provide an insightful view of the actual network state. While network graph abstraction approaches are well-studied [41], their application and efficiency in SDN context have not been well explored yet and remain a serious concern.

**UE mobility handling** – inter-domain UE mobility might require the reconfiguration of the 6G-UPF, migration of mUPF/sUPF across SODs or deployment of additional sUPF. Frequent handovers can result in large resource overhead regarding E2EMO operations.

**DRLAs coordination** – concurrent deployment of several DRLAs optimising various conflicting objectives can potentially result in poor algorithms' convergence and performance. To this end, effective coordination mechanisms have to be developed to verify the actions proposed by DRLAs and introduce the reconfiguration prioritisation to allow reaching key optimisation objectives.

**SLA maintenance** – the DRLA training can lead to severe SLA degradation. Therefore, it is needed to provide a relevant representation of the UP environment for the DRLAs pre-training and evaluation, i.e., UP Digital Twins (DT), before deployment in the production network. Moreover, the supporting solutions that control the agents' behaviour and implement corrective actions in case of instabilities have to be instantiated.

**Environment state representation** – the integration with NTNs requires the creation of network state representations that encapsulate the aspects of time-variant network topology, including link/node activity and dynamic topology changes. In distributed networks, this issue is aggravated by the cross-domain mobility of nodes and the resulting variable size of the network states.

Overall, there is a lack of comprehensive approach to DRL-TE in SDN-UP. While there exist solutions that address specific networking problems, there are no "commercial off-the-shelf" TE solutions that could be effectively applied to carrier-grade deployments. The SUP-DTE approach aims to facilitate the development and integration of TE solutions by introducing a hierarchical distributed architecture, service-based TE framework, and integration with external AI data and model providers.

## VII. Summary and conclusions

In this paper, analysed is the current SotA regarding key 6G UP components, namely, SDN, 3GPP NFs and DRL-TE, with respect to architectural features, challenges, open issues and their integrations. On that basis, the novel 6G-enabled UP architecture called SUP-DTE is proposed, which integrates SDN, 6G-UPF, DRL-based TE and MANO into a hierarchical 6G-ready UP architecture. The key benefits of the SUP-DTE include *user-centric* data-path provisioning, improved operations scalability, embedded intelligent optimisation and support for DRL-based TE, and reduced UPF resource footprint. Finally, presented are the remaining open issues to be addressed to enable the implementation of fully autonomous, flexible and intelligent UP network-wide.

## References


[1] Business Research Insights, "SDN orchestration market size, share, growth, and industry analysis, by type (solutions, & services), by application (cloud service providers, telecom service providers, & others) and regional forecast to 2031," [Website], Jan 2024, Last accessed 09 Feb 2024. [Online]. Available: https://www.businessresearchinsights.com/market-reports/sdn-orchestration-market-109427

[2] ITU-R, "Future technology trends of terrestrial International Mobile Telecommunications systems towards 2030 and beyond," International Telecommunication Union – Radiocommunication Sector, Report M.2516-0 (11/2022), Nov 2022.

[3] Huawei, "6G: The next horizon from connected people and things to connected intelligence," Huawei: Shenzen, China, White Paper, 2021, https://www-file.huawei.com/-/media/corp2020/pdf/tech-insights/1/6g-white-paper-en.pdf?la=en.

[4] NGMN, "6G use cases and analysis," Next Generation Mobile Networks (NGMN) Alliance, White Paper ver. 1.0, Feb 2022. [Online]. Available: https://www.ngmn.org/wp-content/uploads/NGMN-6G-Use-Cases-and-Analysis.pdf

[5] G. Tesauro, "Temporal difference learning and TD-Gammon," *Commun. ACM*, vol. 38, no. 3, pp. 58—68, Mar 1995, doi: 10.1145/203330.203343.

[6] 3GPP, "System architecture for the 5G System (5GS)," 3rd Generation Partnership Project, Technical Standard TS 23.501, ver. 18.5.0, Mar. 2024.

[7] ——, "Interface between the Control Plane and the User Plane nodes," 3rd Generation Partnership Project, Technical Standard TS 29.244, ver. 18.5.0, Mar. 2024.

[8] R. Yadav, R. Kamran, P. Jha, and A. Karandikar, "Applying SDN to mobile networks: A new perspective for 6G architecture," 2023, doi: 10.48550/arXiv.2307.05924.

[9] ONF, "OpenFlow switch specification; version 1.5.1 (protocol version 0x06)," Open Networking Foundation, Specification ONF TS-025, Mar 2015. [Online]. Available: https://opennetworking.org/wp-content/uploads/2014/10/openflow-switch-v1.5.1.pdf

[10] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond," International Telecommunication Union – Radiocommunication Sector, Recommendation M.2160-0 (11/2023), Nov 2023.

[11] ——, "IMT towards 2030 and beyond," [Website], 2024, Last accessed 20 Feb 2024. [Online]. Available: https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2030/Pages/default.aspx

[12] J. Cha, Y. Moon, S. Cho, D. Kim, J. Choi, J. Jung, J. Lee, and S. Choi, "RAN-CN converged user-plane for 6G cellular networks," pp. 2843–2848, 2022, doi: 10.1109/GLOBECOM48099.2022.10001487.

[13] M. Corici, E. Troudt, T. Magedanz, and H. Schotten, "Organic 6G networks : Decomplexification of software-based core networks," in *2022 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 2022, pp. 541–546, doi: 10.1109/EuCNC/6GSummit54941.2022.9815730.

[14] A. Abdulghaffar, A. Mahmoud, M. Abu-Amara, and T. Sheltami, "Modeling and evaluation of software defined networking based 5G core network architecture," *IEEE Access*, vol. 9, pp. 10 179–10 198, 2021, doi: 10.1109/ACCESS.2021.3049945.

[15] J. Costa-Requena, A. Poutanen, S. Vural, G. Kamel, C. Clark, and S. K. Roy, "SDN-based UPF for mobile backhaul network slicing," in *2018 European Conference on Networks and Communications (EuCNC)*, 2018, pp. 48–53, doi: 10.1109/EuCNC.2018.8442795.

[16] Open Networking Foundation, "Using P4 and programmable switches to implement a 4G/5G UPF in Aether," [Website], Sep 2021, Last accessed 10 Apr 2024. [Online]. Available: https://opennetworking.org/news-and-events/blog/using-p4-and-programmable-switches-to-implement-a-4g-5g-upf-in-aether/

[17] S. Schwarzmann, R. Trivisonno, S. Lange, T. E. Civelek, D. Corujo, R. Guerzoni, T. Zinner, and T. Mahmoodi, "An intelligent user plane to support in-network computing in 6G networks," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 1100–1105, doi: 10.1109/ICC45041.2023.10279652.

[18] M. R. Abbasi, A. Guleria, and M. S. Devi, "Traffic engineering in software defined networks: A survey," *Journal of Telecommunications and Information Technology*, vol. 2016, no. 4, pp. 3—14, Dec 2016. [Online]. Available: https://jtit.pl/jtit/article/view/757

[19] H. An, Y. Ji, N. Zhang, W. Hu, Y. Peng, and Y. Wang, "Dynamically split the traffic in software defined network based on deep reinforcement learning," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, Jun 2020, pp. 806–811, doi: 10.1109/IWCMC48107.2020.9148369.

[20] G. Kim, Y. Kim, and H. Lim, "Deep reinforcement learning-based routing on software-defined networks," *IEEE Access*, vol. 10, pp. 18 121–18 133, 2022, doi: 10.1109/ACCESS.2022.3151081.

[21] J. C. Altamirano, M. A. Slimane, H. Hassan, and K. Drira, "QoS-aware network self-management architecture based on DRL and SDN for remote areas," in *2022 IEEE 11th IFIP International Conference on Performance Evaluation and Modeling in Wireless and Wired Networks (PEMWN)*, 2022, pp. 1–6, doi: 10.23919/PEMWN56085.2022.9963841.

[22] Y. Guo, Q. Tang, Y. Ma, H. Tian, and K. Chen, "Distributed traffic engineering in hybrid software defined networks: A multi-agent reinforcement learning framework," 2023, doi: 10.48550/arXiv.2307.15922.

[23] N. F. Zaicu, M. Luckie, R. Nelson, and M. Barcellos, "Helix: Traffic engineering for multi-controller SDN," in *Proceedings of the ACM SIGCOMM Symposium on SDN Research (SOSR)*, ser. SOSR '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 80—87, doi: 10.1145/3482898.3483354.

[24] R. Bellman, "A markovian decision process," *Journal of Mathematics and Mechanics*, vol. 6, no. 5, pp. 679–684, 1957. [Online]. Available: http://www.jstor.org/stable/24900506

[25] W. Feng, C. Han, F. Lian, and X. Liu, "A data-efficient training method for deep reinforcement learning," *Electronics*, vol. 11, no. 24, 2022, doi: 10.3390/electronics11244205.

[26] Z. Wang and M. Jiang, "Enhancing data efficiency in reinforcement learning: a novel imagination mechanism based on mesh information propagation," 2023, doi: 10.48550/arXiv.2309.14243.

[27] Z. Bing, D. Lerch, K. Huang, and A. Knoll, "Meta-reinforcement learning in non-stationary and dynamic environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3476–3491, 2023, doi: 10.1109/TPAMI.2022.3185549.

[28] Z. Zhang, Z. Wu, H. Zhang, and J. Wang, "Meta-learning-based deep reinforcement learning for multiobjective optimization problems," 2022, doi: 10.48550/arXiv.2105.02741.

[29] Z. Zhu, K. Lin, A. K. Jain, and J. Zhou, "Transfer learning in deep reinforcement learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13 344–13 362, 2023, doi: 10.1109/TPAMI.2023.3292075.

[30] P. Almasan, J. Suárez-Varela, K. Rusek, P. Barlet-Ros, and A. Cabellos-Aparicio, "Deep reinforcement learning meets graph neural networks: Exploring a routing optimization use case," *Computer Communications*, vol. 196, pp. 184–194, 2022, doi: 10.1016/j.comcom.2022.09.029.

[31] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[32] T. Zhao, Y. Wang, W. Sun, Y. Chen, G. Niub, and M. Sugiyama, "Representation learning for continuous action spaces is beneficial for efficient policy learning," 2022, doi: 10.48550/arXiv.2211.13257.

[33] D. Abel, A. Barreto, B. V. Roy, D. Precup, H. van Hasselt, and S. Singh, "A definition of continual reinforcement learning," 2023, doi: 10.48550/arXiv.2307.11046.

[34] G. E. Monahan, "A survey of partially observable Markov decision processes: Theory, models, and algorithms," *Management Science*, vol. 28, no. 1, pp. 1–16, 1982. [Online]. Available: http://www.jstor.org/stable/2631070

[35] A. M. Ibrahim, K.-L. A. Yau, Y.-W. Chong, and C. Wu, "Applications of multi-agent deep reinforcement learning: Models and algorithms," *Applied Sciences*, vol. 11, no. 22, 2021, doi: 10.3390/app112210870.

[36] European Parliament and Council, "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS," Official Journal of the European Union, 52021PC0206, 21 Apr 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

[37] T. Hickling, A. Zenati, N. Aouf, and P. Spencer, "Explainability in deep reinforcement learning, a review into current methods and applications," 2023.

[38] L. Tomaszewski, R. Kołakowski, A. Mesodiakaki, K. Ntontin, A. Antonopoulos, N. Pappas, M. Fiore, M. Mosahebfard, S. Watts, P. Harris, C.-K. Lin, A. Santiago, F. Lazarakis, and S. Chatzinotas, "ETHER: Energy-and cost-efficient framework for seamless connectivity over the integrated terrestrial and non-terrestrial 6G networks," in *Artificial Intelligence Applications and Innovations. AIAI 2023 IFIP WG 12.5 International Workshops. AIAI 2023. IFIP Advances in Information and Communication Technology*, I. Maglogiannis, I. Lazaros, P. Antionios, and I. Chochliouros, Eds., vol. 677. Springer, Jun. 2023, pp. 32–44, doi: 10.1007/978-3-031-34171-7_2.

[39] ETSI, "Network Functions Virtualisation (NFV); Management and Orchestration," European Telecommunications Standards Institute, Group Specification, GS NFV-MAN 001 V1.1.1, Dec 2014, https://www.etsi.org/deliver/etsi_gs/nfv-man/001_099/001/01.01.01_60/gs_nfv-man001v010101p.pdf.

[40] IBM, *An architectural blueprint for autonomic computing*. IBM Autonomic Computing White Paper, Fourth Edition, Jun. 2006.

[41] F. Zhou, "Methods for network abstraction," Ph.D. dissertation, University of Helsinki, Finland, 2012.