Optimizing Key Value Indicators in Intent-Based Networks through Digital Twins aided service orchestration mechanisms

Federica de Trizio[a], Giancarlo Sciddurlo[a,b], Ilaria Cianci[a], Giuseppe Piro[a,b], Gennaro Boggia[a,b]

[a] *Department of Electrical and Information Engineering, Politecnico di Bari, Bari, Italy*
[b] *CNIT, Consorzio Nazionale Interuniversitario per le Telecomunicazioni*

Abstract

For many years, the orchestration of network resources and services has been addressed by considering homogeneous communication infrastructures and simple Service Level Agreements (SLAs), generally defined through a list of traditional Key Performance Indicators (KPIs). Unfortunately, state-of-the-art solutions risk being quite ineffective for future telecommunication systems. Beyond 5G networks, for instance, are emerging as complex and heterogeneous ecosystems where resources belonging to diverse network domains with evolving capabilities can be dynamically exposed to support much more complex and cross-domain services and applications. At the same time, SLAs will be defined by also considering novel performance demands, including security, economic, and environmental needs. Based on these premises, this work proposes a novel orchestration strategy designed to fulfill service requirements expressed through Key Value Indicators (KVIs), while combining the potentials of both Network Digital Twins and Intent-Based Networking. Leveraging insights from Network Digital Twins, multiple service orchestration options are explored to optimize resource utilization. Simultaneously, Intent-Based Networking is adopted to streamline network management via intents, specifying Beyond 5G requirements through KPIs and KVIs. An optimal orchestration scheme has been conceived through a multi-criteria decision-making algorithm and a many-to-many matching game between domains and service requests mapped into intents, aiming to minimize SLA violations over time. The performance of the conceived solution has been investigated through computer simulations in realistic scenarios. The obtained results clearly highlight its effectiveness and demonstrate that it is able to reduce SLA violations (related to latency, throughput, costs, and cyber risk requirements) by up to 22.44% compared to other baseline techniques.

Keywords: Services orchestration, Intent-Based Network, Network Digital Twin, Service Level Agreements

1. Introduction

As the evolution of Beyond 5G (B5G) networks progresses, complex infrastructures are increasingly integrated with cutting-edge technologies, services, and stakeholders, thereby imposing a wide range of diverse and rigorous performance requirements [1]. Considering this aspect, the orchestration of services, denoting the coordination and administration of network resources to deliver tailored services to end-users or clients, necessitates the adoption of prospective methodologies. Such methodologies are essential for judiciously allocating resources and fulfilling specified demands efficiently.

By leveraging insights derived from Network Digital Twins (NDTs), service orchestration platforms can dynamically enhance the efficiency of network resources and service delivery processes. Serving as digital representations of physical networks, NDTs have the capability to incorporate real-time data to faithfully replicate their behavior and attributes [2]. This functionality enables the analysis, prediction, and optimization of network performance, while also facilitating experimentation with diverse scenarios and configurations [3].

A plethora of solutions addressing service orchestration in next-generation networks are extensively documented in the scientific literature. Specifically, existing methodologies endeavor to allocate tasks to appropriate network domains, each characterized by stringent performance requirements [4–9]. Furthermore, significant contributions capitalize on the opportunities afforded by NDTs to discern effective strategies for orchestrating services and identifying network resources suitable for service provisioning [10–19].

On the other hand, beyond performance requirements traditionally expressed through Key Performance Indicators (KPIs) and readily managed via NDT, the complexity of service orchestration in B5G networks also encompasses a focus on security, economic, and environmental considerations articulated through Key Value Indicators (KVIs) [20, 21]. These are value metrics aimed at quantifying significant ethical principles such as sustainability and trust, aligning with the United Nations (UN) Sustainable Development Goals (SDGs) [22, 23]. In this context, while the evaluation of KPIs focuses on short-term, tangible outcomes, KVIs are designed to capture the overall value and long-term impact of desired outcomes. This approach ensures alignment with stakeholder objectives and delivers sustained value over time. To concurrently address these diverse requirements within a resource-constrained ecosystem, consumers and service providers establish Service Level Agreements (SLAs) as an explicit declaration of the expected service, its associated performance metrics, and penalties in the event of non-compliance [24, 25].

However, formulating a service orchestration strategy to effectively fulfill demanding services while adhering to SLAs poses several challenges. Network resources exhibit heterogeneity, and their capabilities evolve over time, leading to divergent satisfaction levels for identical services depending on the employed resources and deployment locations. Furthermore, exclusive reliance on representation via NDT may not consistently offer a comprehensive understanding of the expectations, security considerations, and economic responsibilities of all stakeholders involved.

To address this issue, exploring Intent-Based Networking (IBN) can prove beneficial in streamlining network management and operation by prioritizing the desired outcomes [24]. Through the adoption of this networking paradigm, users can articulate technology-agnostic demands via intents, specifying performance requirements through KPIs. Moreover, going one step further, they can also communicate emerging B5G demands through KVIs, thereby enabling the configuration of novel network policies to foster collaboration across diverse environments for efficient service delivery [26]. An intent-based network then transparently translates these high-level objectives into measurable metrics and actions, bridging the gap between user intentions and technical execution [27].

To the best of the authors' knowledge, current literature does not fully explore the potential offered by the combination of NDTs and IBN in service orchestration frameworks. By abstracting service deployments using NDT, multiple service orchestration options can be explored with the aim of meeting established SLAs, while also fostering collaborations among networks to handle service requests mapped through intents. Additionally, current contributions in this direction address only a limited number of KPIs and fail to comprehensively model important KVIs such as network security and cyber risks, thus restricting the effectiveness and reliability of service orchestration. Facing these open issues and aiming to extend and enhance the existing scientific literature, the main contributions of this work are summarized as follows:

- A model for orchestrating B5G services is proposed with the goal of adhering to established SLAs. It investigates and determines the appropriate resources of network domains to handle service requests. Specifically, by leveraging the IBN structure, it enables informed decision-making considering both KPIs and KVIs. This approach jointly minimizes the violations characterized by the number of requests not aligned with latency, throughput, costs, and security requirements.

- The service provisioning orchestration problem is modeled as a many-to-many matching game between service requests, which are mapped into intents, and network domains, whose resources and capabilities are available through their NDTs representation. Players' preferences are generated using the TOPSIS decision-making algorithm, which utilizes the Entropy-Weighting Method (EWM) to compute weights.

- To evaluate the efficiency of the proposed model, a simulation campaign is conducted, comparing it against three different baseline approaches. The obtained results demonstrate reductions in SLA violations related to latency, throughput, costs, and security requirements, by up to 22.44% compared to baseline techniques.

The remainder of this paper is organized as follows: Section 2 discusses related works. Section 3 illustrates the reference scenario and formulates the system model. Section 4 presents the formulated optimization problem. The proposed service orchestration solution is described in Section 5, and the evaluations, including comparisons against baseline schemes, are presented in Section 6. Finally, conclusions are drawn in Section 7.

2. Related Works

Network orchestration and service provisioning in B5G networks involve the coordinated management and deployment of network resources and services. This process necessitates a sophisticated orchestration framework capable of integrating advanced technologies, diverse services, and a broad range of stakeholders. The framework must efficiently manage the complexities associated with discovering, allocating, and controlling disaggregated and distributed resources to ensure the delivery of customized, high-quality services to end users. Additionally, it must meet stringent performance requirements, typically quantified by KPIs, while also addressing broader considerations such as security, economic impact, and environmental sustainability, which are evaluated through KVIs. Current scientific research is actively exploring solutions for service orchestration and network management in next-generation networks to achieve these objectives.
Initially, researchers concentrated on investigating KPIs in service orchestration solutions by formulating complex optimization problems related to deployment costs [4–7] or reducing delays [8]. Other recent advancements in service orchestration frameworks have been driven by the integration of NDTs, which enable flexible testing of configurations prior to deployment, as explored in [10–19].

In [11], the authors propose a mechanism that leverages NDTs to simulate the aggregation of network resources that fulfill specific service requirements. This problem is modeled as a Boolean satisfiability problem, which is addressed using a heuristic algorithm due to its inherent complexity. In contrast, the authors of [12] formulate two optimization problems for deploying Virtual Network Functions (VNFs), with their reliability predicted through NDTs. The objective is to minimize service costs while maximizing the number of admitted service requests. To solve these problems, they propose an approximation strategy and implement an online algorithm.

Instead of leveraging NDTs for simulation and prediction of parameters, the authors of [14, 15, 18] utilize the virtual counterpart of the physical network to represent the estimated processing rates and computational capacities of resources. They formulate optimization problems aimed at minimizing average processing delay, reducing end-to-end offloading latency, and maximizing utility across different computation modes, respectively.

In alignment with the current state of the art, our previous research work [19] employed NDTs to identify the most suitable network segments for task processing, with a particular focus on ensuring survivability during disaster events. We introduced an orchestration algorithm designed to select domains capable of providing services based on their reliability, storage availability, and computational capability. Notably, we associated the reliability parameter with network segments to assess their trustworthiness.

In addition to the use of NDT technology, service orchestration has been significantly enhanced by adopting the IBN paradigm. IBN enables the automatic translation of high-level objectives into specific network configurations, as explored in [28–32]. On the one hand, IBN is utilized to automate network slicing orchestration, as demonstrated in [28, 31], where reinforcement learning and AI techniques are applied to prevent Quality of Service (QoS) violations and proactively manage resources. On the other hand, in [29, 30, 32], the IBN paradigm effectively facilitates the translation of user intents into service function chains and policy trees.

Nevertheless, while the current state of the art strategies offer valuable orchestration solutions, they are predominantly focused on optimizing a narrow subset of KPIs. Although they align with the requester's goal of deploying services promptly and economically, they may not guarantee the comprehensive utilization of constrained resources and the achievement of broader societal goals and security considerations. The formulation of SLAs and the potential to incorporate KVIs, particularly security, into service orchestration strategies are often overlooked. To address this gap, this study proposes a service orchestration scheme that integrates NDTs and IBN. By leveraging intents, heterogeneous requests can be articulated to encompass both KPIs and KVIs, enabling accurate and efficient service deployments that proactively minimize SLA violations.

3. The Reference Architecture and System Model Definition

This section introduces the reference architecture and delineates the system model.

3.1. The reference system architecture

Fig. 1 illustrates the layered reference architecture inspired by Network 2030 [24], which capitalizes on the IBN paradigm. The depicted environment comprises consumers, who are users or clients accessing services, and providers, which are network domains delivering those services or content. In the upper layer, consumers articulate service requests through applications, encompassing demands for specific network services or functionalities. The second layer, termed the intent layer, assumes a pivotal role in enabling intent-driven network management within IBN, crucial in facilitating network automation. Instead of managing low-level configurations, users can articulate their desired outcomes as high-level objectives in a language that the network can interpret, which is then seamlessly translated into specific configurations and policies. By providing such a mechanism, it maps service requests into intents, thereby enhancing agility and ensuring alignment with specific objectives [33]. Through intent mapping, both the consumer and the service provider can specify the level of service expected from the network via SLAs, delineating the quality, availability, and reliability of services provided by the network to ensure they fulfill the needs and expectations of the users. In particular, conventional performance requirements can be stipulated through KPIs, while sustainability and security demands can be articulated through KVIs. Once high-level service requests are mapped into structured and quantifiable intents, they are transmitted to the Orchestration layer. Within this layer, the Network Orchestration module, in conjunction with the Service Orchestration module, collaborates to determine how and where to deploy these intent demands. Specifically, the former generates NDTs of domains by exploring the characteristics and data of the physical Network Infrastructure. Furthermore, they are not just digital replicas, but as dynamic and aggregated data from network domains, they engage with their physical counterpart for predictive maintenance and operation. NDTs are leveraged to simulate possible service implementations, from which it is possible to identify resource configurations to be instantiated on-demand based on the availability of the resources. Conversely, the latter is tasked with identifying the appropriate network domain resources capable of meeting SLA requirements. The collaboration between these modules thus enables service providers to simulate various resource configurations, leveraging the abstraction provided by NDT, and evaluate their impact on network performance and behavior before actual deployment in the network infrastructure. Finally, the suitable resources of the selected domain, compliant with established

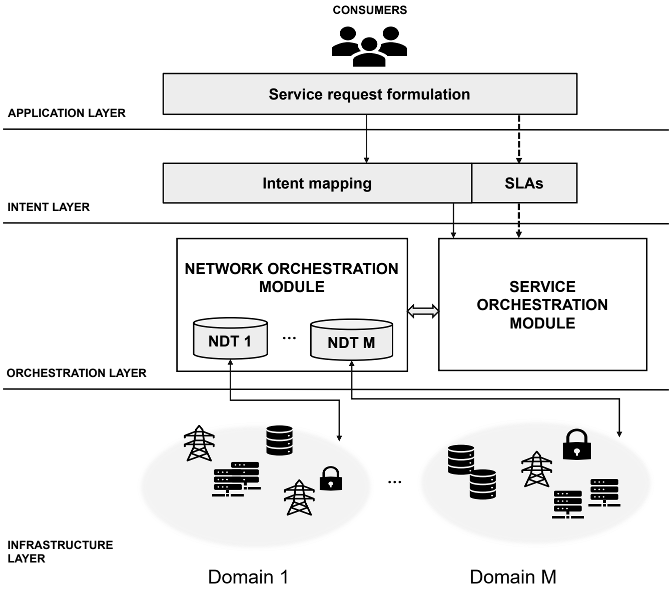


Figure 1: The proposed architecture inspired by the framework of IBN for Network 2030 [24].

SLAs, are allocated in the Infrastructure layer to fulfill the service request. The details of the entire orchestration procedure will be provided in the next sections.

3.2. System Model

As outlined in the previous section, consumers articulate service requests that encompass specific demands for network services. The explored IBN paradigm, facilitated by proper translation [33], enables the mapping of these service requests into a category of intent, representing the definition of service requirements. For the sake of clarity, let $\mathbb{I}$ denote the set of N intent categories, such that $\mathbb{I} = \{I_1, I_2, \dots, I_n, \dots, I_N\}$. It is noteworthy that several service requests can be mapped to the same n-th intent category. With this in mind, we can formally define $\mathbb{R}^{(n)}$ as the set of all consumer requests mapped into the n-th intent category, expressed as $\mathbb{R}^{(n)} = R_1^{(n)}, R_2^{(n)}, \dots, R_s^{(n)}, \dots, R_S^{(n)}$. Consequently, the entire set of service requests in the system can be defined as $\mathbb{R} = \bigcup_{n=1}^{N} \mathbb{R}^{(n)}$. Overall, when a consumer sends a service request $R_s^{(n)} \in \mathbb{R}^{(n)}$, it specifies the following set of parameters: $\left\langle T^{(n)}, \gamma^{(n)}, \tau^{(n)}, \Delta^{(n)}, \theta^{(n)}, \beta^{(n)}, B^{(n)} \right\rangle$. Specifically,

- $\mathbb{T}^{(n)}$ represents the set of K tasks into which the requested service is decomposed. Specifically, $\mathbb{T}^{(n)} = \left\{T_1^{(n)}, T_2^{(n)}, \dots, T_k^{(n)}, \dots, T_K^{(n)}\right\}$. This decomposition of the service into simpler tasks facilitates a more systematic and manageable approach to service delivery, enabling the network domain to enhance efficiency, reliability, and performance across the entire service delivery process.

- $\gamma^{(n)}$ indicates the deadline of the $R_s^{(n)}$-th service request, expressed in [s]. Thus, completing a service within its requested timeframe is a traditional performance requirement.

- $\tau^{(n)}$ represents the minimum throughput requested by the consumer for the $R_s^{(n)}$-th service request.

- $\Delta^{(n)}$ is the impact of a security attack and reflects the effects or consequences caused by an attack resulting from the provision of malicious service. It is a crucial value closely related to the SDGs outlined by the UN, particularly concerning privacy, confidentiality, and trustworthiness [22].

- $\theta^{(n)}$ represents the acceptable risk associated with the requested service, often referred to as risk appetite. Essentially, it signifies the consumer's willingness to assume risk.

- $\beta^{(n)}$ is the budget representing the available funds allocated by the consumer for the utilization and storage of providers' resources. It embodies a key value linked to the UN SDGs of economic growth and sustainability [23].

- $B^{(n)}$ represents the input size of the data to be processed by a service provider, measured in [bits].

On the other hand, let $\mathbb{D}$ be the set of M network domains, such that $\mathbb{D} = \{D_1, D_2, \dots, D_m, \dots, D_M\}$. Each domain $D_m \in \mathbb{D}$ exposes storage and computational resources and is characterized by the following set of parameters: $\left\langle \Omega^{(m)}, q^{(m)}, L^{(m)}, \sigma^{(m)}, Bw^{(m)} \right\rangle$, detailed as follows:

- $\Omega^{(m)}$ represents the processing capability of the m-th domain, measured in [cycles/s].

- $q^{(m)}$ represents the maximum quota of tasks that the m-th domain can handle simultaneously. This parameter accounts for the number of computational cores in a domain, each with a capability denoted by $\Omega^{(m)}$.

- $L^{(m)}$ symbolizes the likelihood of attacks against the security of the m-th domain. It represents the probability that a threat will exploit a vulnerability, thereby causing damage, in the m-th domain.

- $\sigma^{(m)}$ is described by a tuple of values $(C_u^{(m)}, C_{lf}^{(m)}, C_{sec}^{(m)})$. Specifically, $C_u^{(m)}$ represents the parameter for usage costs that consumers must pay to utilize the resources provided by the m-th network domain, measured in [\$/s]. $C_{lf}^{(m)}$ and $C_{sec}^{(m)}$ represent the life cycle costs and security costs, respectively, measured in [\$], that a domain must support to maintain the activity of its resources and security measures.

- $Bw^{(m)}$ is the minimum guaranteed value of available bandwidth of a given domain for a service request.

Table 1 summarizes the key symbols used to describe the environment along with their meanings.

3.3. Parameters Computation

- Delay Computation

Considering the overall end-to-end delay as the time elapsed from the instant the service request is made to the moment the service is provided to the consumer, it is assumed to be the linear combination of two contributions. The former, indexed by $t^{(tran)}_{(R_s^{(n)}, D_m)}$, represents the transmission delay and refers to the amount of time spent ensuring the arrival of the $R_s^{(n)}$-th service request in the m-th network domain. To calculate the transmission delay across the communication channel, it is necessary to determine the transmission rate between the consumer and the provider domain. The aforementioned rate, referring to the $R_s^{(n)}$-th request towards the m-th network domain, denoted by $U_{(R_s^{(n)}, D_m)}$, is calculated using the Shannon formula [34] and reported as follows:

$$U_{(R_s^{(n)}, D_m)} = Bw^{(m)} \cdot \log_2 \left(1 + \frac{P_{tx} \cdot H^{(m)}}{P_{noise}} \right), \qquad (1)$$

where P_{tx} is the transmission power of the consumer requesting services, $H^{(m)}$ is the transmission channel gain, and P_{noise} is the noise power of the transmission channel. Considering the

Table 1: Main Symbols Description.

Symbol	Meaning
$\mathbb{I}$	Set of intents
$\mathbb{D}$	Set of network domains
N	number of intent categories
M	number of network domains
$\mathbb{R}^{(n)}$	Set of consumer requests mapped into n-th intent category
$\mathbb{T}^{(n)}$	Set of tasks composing a service request
K	Number of tasks composing a service
$\tau^{(n)}$	Minimum requested throughput for $R_s^{(n)}$-th service request
$\gamma^{(n)}$	Deadline of the $R_s^{(n)}$-th service request
$\Delta^{(n)}$	Impact of a security attack
$\beta^{(n)}$	Consumer budget for the $R_s^{(n)}$-th request
$B^{(n)}$	Input size of the $R_s^{(n)}$-th service request
$\theta^{(n)}$	Tolerable risk associated with the $R_s^{(n)}$-th service request
$\Omega^{(m)}$	Processing capability of the m-th domain
$L^{(m)}$	Likelihood of attacks in the m-th domain
$\sigma^{(m)}$	Cost to pay for resources provided by the m-th domain
$C_{lf}^{(m)}$	Life cycle cost in the m-th domain
$C_{u}^{(m)}$	Usage cost in the m-th domain
$C_{sec}^{(m)}$	Security cost in the m-th domain
$q^{(m)}$	Maximum quota of tasks that the m-th domain can handle
$Bw^{(m)}$	Minimum guaranteed bandwidth for a request by the m-th domain
$t^{(tran)}_{(R_s^{(n)},D_m)}$	Transmission delay for the $R_s^{(n)}$-th request in the m-th domain
$t^{(proc)}_{(R_s^{(n)},D_m)}$	Processing delay for the $R_s^{(n)}$-th request in the m-th domain
$U_{(R_s^{(n)},D_m)}$	Transmission rate between consumer and the m-th domain
$\rho_{(R_s^{(n)},D_m)}$	Cyber risk associated to the $R_s^{(n)}$-th request in the m-th domain
$Rev^{(m)}$	Total revenue of the m-th domain
$x_{(T_k(R_s^{(n)}),D_m)}$	Binary indicator variable indicating if a task T_k of $\mathcal{R}s^{(n)}$ is allocated to D_m
$\mathcal{F}$	Set of tasks with unmet deadline requirements
$\mathcal{E}$	Set of tasks with unmet throughput requirements
C	Set of tasks overshooting the user's budget
$\mathcal{S}$	Set of tasks violating cyber risk requirements

$B^{(n)}$ input size of the data to be processed for the requested service and the previously calculated transmission rate, the transmission delay can be computed as reported in eq.2.

$$t^{(tran)}_{(R_s^{(n)},D_m)} = \frac{B^{(n)}}{U_{(R_s^{(n)},D_m)}}. \tag{2}$$

The latter contribution, instead, indexed by $t^{(proc)}_{(R_s^{(n)},D_m)}$, represents the processing delay and refers to the amount of time needed by the m-th provider domain to process and complete the $R_s^{(n)}$-th requested service. It can be calculated as reported in eq.3.

$$t^{(proc)}_{(R_s^{(n)},D_m)} = \frac{\Phi \cdot B^{(n)}}{\Omega^{(m)}}, \tag{3}$$

where Φ is the number of CPU cycles needed to process a single bit, and according to [35], it has been set equal to 1000 cycles per bit.

Finally, considering the previous contributions, the overall end-to-end delay for the provision of the $R_s^{(n)}$-th requested service by the m-th domain is calculated as shown in eq. 4.

$$t_{(R_s^{(n)},D_m)} = t^{(tran)}_{(R_s^{(n)},D_m)} + t^{(proc)}_{(R_s^{(n)},D_m)}. \tag{4}$$

- Cyber Risk Computation

Designed to connect vast numbers of devices and incorporate technologies like Software Defined Networking (SDN), network slicing, and edge computing, B5G networks support critical infrastructure and services, including healthcare (e.g., remote surgery), transportation (e.g., autonomous vehicles), energy grids, and public safety. Additionally, they promise ultra-fast speeds and low latency, which are essential for real-time applications. The integration of multiple technologies increases the number of potential entry points for cyber attackers, creating new vulnerabilities and making network security more challenging. Each component (e.g., virtualized network functions, edge devices) could serve as a potential target for cyber threats. A successful cyber-attack can severely impact network performance, with catastrophic consequences for critical infrastructure and services, including loss of life, large-scale economic disruption, and threats to national security.

In this context, service providers must also account for non-negligible risk factors when taking on service requests. These risks can be caused by natural disasters or man-made events, such as human failure, cybercrime (e.g., extortion, fraud), cyberwar, or cyberterrorism [36]. Generally, cyber risk compromises the confidentiality, availability, or integrity of data or services. Overall, it can be considered a function of three parameters: (i) vulnerability, (ii) threat, and (iii) impact [37]. A vulnerability is a flaw or weakness in an asset's design, implementation or operation and management, thus it depends on the security level of network domains. A threat is a potential exploitation of a vulnerability. Typically, these two factors are grouped together and indexed as L, representing the likelihood of a successful attack [38]. Then, the impact of cyber risk refers to the consequences or effects that security threats and vulnerabilities can have on individuals, organizations, and systems as a whole. It represents the reputational, economic, and legal damages that consumers, who formulate service requests, will suffer if they become victims of a threat.

Estimating the security parameter affecting network domains is a challenging task. A variety of heterogeneous risk assessment methods are available in the literature for determining it. Without loss of generality, this work adopts the model based on

Attack Countermeasure Trees (ACT) [39] as the reference one for likelihood modeling. This model enables the determination of attack paths or vulnerabilities and the selection of appropriate countermeasures to detect and mitigate them. The domain's security posture can thus be assessed by evaluating the probability of attack success. This is modeled as a "k-out-of-n" probability, where A_S represents the successful attack events out of the total A_{DM} attacks detected and mitigated. Then, the corresponding expression for the likelihood of attack success for the m-th network domain is provided in eq.5.

$$L^{(m)} = \sum_{j=A_S}^{A_{DM}} \left(\begin{array}{c} A_S \\ j \end{array} \right) Pr_{Atk}^{j}(1 - Pr_{Atk})^{A_S - j} \tag{5}$$

where Pr_{Atk} represents the probability of an attack event. Therefore, considering the $R_s^{(n)}$-th service request handled by the m-th domain, the associated Cyber Risk $\rho_{(R_s^{(n)},D_m)}$ can be defined by the following equation:

$$\rho_{(R_s^{(n)},D_m)} = \Delta^{(n)} \cdot L^{(m)}. \tag{6}$$

The proposed system model seeks to enhance orchestration strategies by surpassing traditional KPI-based service provider selection. It integrates the UN' SDGs, monitored via KVIs, to meet next-generation network objectives. Although B5G technologies facilitate these goals, they also introduce new vulnerabilities, positioning cyber risk as a crucial KVI for maintaining B5G security.

- Cost Computation

The service provisioning procedure also incurs costs. Specifically, domain providers must manage the entire life cycle of resources and invest in security countermeasures. This work assumes $C_{lf}^{(m)}$ as the cost associated with the life cycle of managed resources, expressed as $C_{lf}^{(m)} = C_{act}^{(m)} + (C_{depl}^{(m)} \cdot q^{(m)})$, where $C_{act}^{(m)}$ represents the cost of service activation, and $C_{depl}^{(m)}$ denotes the deployment and maintenance cost of the physical network. $C_{sec}^{(m)}$, on the other hand, denotes the cost associated with auditing, vulnerability assessment, threat mitigation, implementation of security tools, as well as regular updates and patches within the m-th network domain [40]. These costs ensure that the domain remains secure and compliant with necessary regulations, reducing vulnerabilities to cyber attacks. Meanwhile, consumers are charged based on the duration they utilize resources from the domains. $C_{ut}^{(m)}$ represents the cost associated with the usage of processing and storing resources. This cost is often determined by the time resources are allocated to the consumer's service requests, ensuring that consumers pay proportionately for the resources they consume. In summary, the total cost structure for service provisioning includes the life cycle costs ($C_{lf}^{(m)}$), security costs ($C_{sec}^{(m)}$), and usage costs ($C_{ut}^{(m)}$). These costs are essential for maintaining the infrastructure and ensuring secure, reliable service delivery to consumers. Considering the incurred costs to maintain network resources related to their life cycle and the security of the infrastructure, along with the profits generated from consumers, the total revenue of the m-th domain, indexed by $Rev^{(m)}$, can be computed as shown in Eq. 7.

$$Rev^{(m)} = C_{ut}^{(m)} \cdot t^{(proc)}_{(R_s^{(n)},D_m)} - (C_{lf}^{(m)} + C_{sec}^{(m)} \cdot L^{(m)}), \tag{7}$$

where the cost for resource employment and the security costs are weighted by the time of resource usage and the likelihood of cyber attacks succeeding in the m-th network domain, respectively. This ensures that the revenue calculation accurately reflects the economic impact of both operational costs and potential security risks.

4. Problem Formulation

The primary objective of this study is to develop a service orchestration strategy for selecting available network domains that can fulfill consumer service requests while minimizing inconsistencies caused by SLA violations. Specifically, the proposed methodology aims to identify resource configurations that simultaneously achieve the following goals:

- Minimizing the number of tasks whose completion time exceeds their deadline.
- Minimizing the number of tasks that fail to meet their throughput threshold.
- Minimizing the costs incurred by task deployment exceeding the consumer's budget.
- Minimizing the task assignments whose associated cyber risk exceeds the consumer's risk appetite.

Given the inherent heterogeneity of consumers and their related service requests, intent mapping serves as a crucial starting point for organizing and orchestrating service requests effectively. Depending on the nature of the business or service, intent mapping can automate the categorization process by analyzing the content of consumer inquiries. This innovative approach facilitates the identification of the optimal network domain for processing and executing tasks to meet service requirements efficiently. Moreover, the proposed architecture facilitates the decomposition of requested services into smaller, easily manageable tasks that can be executed across different network domains. Each task ideally represents a distinct function or step in the service delivery process. This approach enhances agility and improves the overall quality of service delivery for organizations. In line with this, the proposed service orchestration method deploys tasks from various services, identifying a suitable and distributed combination of network domains. This combination collectively addresses the required completion time, throughput, cost, and cyber risk requirements.

Let $x_{(T_k(R_s^{(n)}),D_m)}$ denote a binary indicator variable that indicates whether a task T_k assigned to fulfill the service request $\mathcal{R}s^{(n)}$ is allocated to domain D_m:

$$x_{(T_k(R_s^{(n)}),D_m)} = \begin{cases} 1 & \text{if } T_k \text{ for } R_s^{(n)} \text{is assigned to } D_m, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

with $n \in [1, N]$ representing the n-th requested intent category, and $m \in [1, M]$ representing the m-th network domain.

Considering that each demanded service mapped into an intent category has a deadline, it's essential to define the set $\mathcal{F}$ of tasks with unmet completion time requirements. Assuming that each task composing the $R_s^{(n)}$-th service request can be executed simultaneously by the selected network domain, the resulting time $t_{(R_s^{(n)},D_m)}$ spent to serve the consumer corresponds to the maximum among the end-to-end delays of all provider domains to fulfill their assigned tasks. According to this, $\mathcal{F}$ can be expressed as follows:

$$\begin{aligned}\mathcal{F} = \{&T_k^{(n)} \in \mathbb{T}^{(n)} \mid x_{(T_k(R_s^{(n)}),D_m)} \cdot t_{(R_s^{(n)},D_m)} > \gamma^{(n)}; \\ &\forall R_s^{(n)} \in \mathbb{R}^{(n)}, k \in [1, K], n \in [1, N], m \in [1, M]\}.\end{aligned} \tag{9}$$

Throughput represents another SLA metric for a demandable service, referring to the rate at which tasks are processed or completed within a given timeframe. In this context, let $\mathcal{E}$ be the set of tasks whose completion does not guarantee the fulfillment of the throughput requirements. Its expression is reported as follows:

$$\begin{aligned}\mathcal{E} = \{&T_k^{(n)} \in \mathbb{T}^{(n)} \mid x_{(T_k(R_s^{(n)}),D_m)} \cdot U_{(R_s^{(n)},D_m)} < \tau^{(n)}; \\ &\forall R_s^{(n)} \in \mathbb{R}^{(n)}, k \in [1, K], n \in [1, N], m \in [1, M]\}.\end{aligned} \tag{10}$$

Let also $\mathcal{C}$ be the set of service tasks entailing costs that fit within the user's budget. Considering the possible task division among domains, the processing time $t^{(proc)}_{(R_s^{(n)},D_m)}$ must account for the maximum among the processing times of all provider domains to process their assigned tasks. Its expression is as follows:

$$\begin{aligned}\mathcal{C} = \{&T_k^{(n)} \in \mathbb{T}^{(n)} \mid x_{(T_k(R_s^{(n)}),D_m)} \cdot C_{ut}^{(m)} \cdot t^{(proc)}_{(R_s^{(n)},D_m)} > \beta^{(n)}; \\ &\forall R_s^{(n)} \in \mathbb{R}^{(n)}, k \in [1, K], n \in [1, N], m \in [1, M]\}.\end{aligned} \tag{11}$$

Furthermore, if the cyber risk $\rho_{(R_s^{(n)},D_m)}$ associated with the n-th service request, denoted as $R_s^{(n)}$, handled by the m-th domain exceeds the acceptable levels of risk appetite $\theta^{(n)}$, it suggests a potential misalignment between the risk tolerance levels of the two parties involved. Consequently, cybersecurity concerns are not appropriately addressed. Considering this, let $\mathcal{S}$ be the set of tasks of a service assigned to domains that violate their cyber risk requirements, as reported below:

$$\begin{aligned}\mathcal{S} = \{&T_k^{(n)} \in \mathbb{T}^{(n)} \mid x_{(T_k(R_s^{(n)}),D_m)} \cdot \rho_{(R_s^{(n)},D_m)} > \theta^{(n)}; \\ &\forall R_s^{(n)} \in \mathbb{R}^{(n)}, k \in [1, K], n \in [1, N], m \in [1, M]\}.\end{aligned} \tag{12}$$

Therefore, to identify the optimal decision in task assignment for the service orchestration procedure, addressing violations of the SLAs becomes pivotal. This can be achieved by jointly minimizing the cardinality of the previously defined sets $\mathcal{F}, \mathcal{E}, \mathcal{C}, \mathcal{S}$, thereby reducing the number of tasks that fail to adhere to service requirements. Accordingly, the following multi-objective optimization problem is formulated and reported in eq.13:

$$\min_{x_{(T_k(R_s^{(n)}),D_m)}} \quad |\mathcal{F}| + |\mathcal{E}| + |\mathcal{C}| + |\mathcal{S}| \tag{13}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{k=1}^{K} x_{(T_k(R_s^{(n)}),D_m)} \leq q^{(m)}, \forall m \in [1, M] \tag{14}$$

$$\sum_{m=1}^{M} \sum_{k=1}^{K} x_{(T_k(R_s^{(n)}),D_m)} = K, \tag{15}$$

$$\forall R_s^{(n)} \in \mathbb{R}^{(n)}, \forall n \in [1, N]$$

$$\sum_{m=1}^{M} \sum_{s=1}^{S} \sum_{k=1}^{K} x_{(T_k(R_s^{(n)}),D_m)} \leq \sum_{m=1}^{M} q^{(m)} \tag{16}$$

$$x_{(T_k(R_s^{(n)}),D_m)} \in \{0, 1\}. \tag{17}$$

The proposed multi-objective problem is subject to constraints reported in eq.(14)-(16). Constraint (14) expresses that the m-th network domain can handle at most $q^{(m)}$ tasks simultaneously. Constraint (15) indicates that all K tasks of a requested service must be assigned to a network domain. Finally, constraint (16) formalizes that the number of tasks referred to the overall set of requested services does not exceed the total capacity of all network domains. Therefore, all requested services mapped into intents must be assigned and processed by at most one network domain.

The formulation expressed in eq.(13)-(17) constitutes a combinatorial problem, proven to be NP-hard [41]. Indeed, the search spaces for solutions of such problems tend to grow exponentially as the size of the input increases, making it infeasible to find an optimal solution in polynomial time.

5. The Proposed Solution

In this section, a service orchestration procedure is developed utilizing a multi-criteria decision-making algorithm within a stable matching game framework. This approach employs a heuristic method to address the NP-hard problem outlined in Section 4.

5.1. Service Orchestration Procedure Overview

The conceived novel procedure leverages matching theory, a field that enables the development of flexible solutions, particularly matching games, for combinatorial problems [42–44]. Typically, a matching game involves interactions between elements of two distinct sets, referred to as players, and produces a matching function that establishes valuable relationships among them. These relationships are based on the preferences expressed by each player of one set toward players of the opposing set, indicating the level of satisfaction with the match, which is computed according to appropriate criteria [44]. Considering these assumptions, the players in the proposed work are represented by sets of network domains and service requests

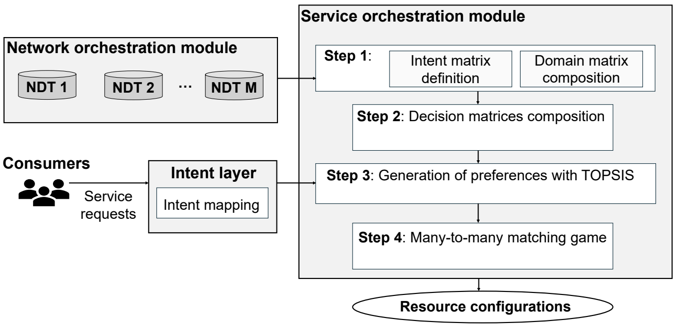

Service orchestration module
Network orchestration module
NDT 1
NDT 2
···
NDT M
Step 1:
Intent matrix definition
Domain matrix composition
Step 2: Decision matrices composition
Consumers
Service requests
Intent layer
Intent mapping
Step 3: Generation of preferences with TOPSIS
Step 4: Many-to-many matching game
Resource configurations

Figure 2: Overall orchestration procedure.

mapped into categories of intents. Their preferences are evaluated using the TOPSIS multi-criteria decision-making algorithm [45]. By defining players and computing preferences in this manner, it is possible to match intents to domains capable of processing them in a timely, efficient, convenient, and safe manner. Conversely, domains are matched with intents that entail higher revenues.

The rest of this section illustrates the methodology followed by the conceived procedure, schematically shown in Fig. 2 and fully deployable through the modules of the reference architecture described in Section 3. Specifically, starting from the NDT representations and the intent category definitions, the conceived Service Orchestration module builds the domain and intent matrices. Utilizing this collected data, a decision matrix or vector is composed for each player in the matching game (e.g., consumers' intents and domains). The values within the decision matrices, representing the outcomes of NDTs, effectively capture the resource configurations a domain must expose to fulfill a specified intent. This mathematical modeling through matrices expresses the capacity of NDTs to self-simulate their capabilities, enabling analysis, prediction, and offering decision-making support for the Service Orchestration module. After computing the weights, the TOPSIS multi-criteria decision-making algorithm derives preferences from the decision matrices for each player. The evaluation of these preferences paves the way for establishing the associations in the many-to-many matching game, leading to the extraction of beneficial network resource configurations.

Step 1 *- Domains and intent matrix composition*

By providing an abstraction of network domains, NDTs enable effective monitoring and analysis of complex network systems. This approach facilitates the construction of detailed behavioral models, capturing performance metrics such as latency and throughput. The proposed work leverages the outcomes of NDTs as an initial step in the service orchestration process, presenting the results using a matrix representation. Specifically, $\mathbf{D}$ represents the domain matrix, with dimensions $[M \times V]$, where M denotes the number of network domains considered, while V represents the set of features captured by the NDTs abstraction. These features include processing capability, the quota of assignable tasks, the likelihood of security-compromising attacks, costs, and the guaranteed bandwidth of each domain, as detailed in Section 3. Here, the domain matrix entries $d_{m,v}$, where $m \in [1, M]$ and $v \in [1, V]$, represent the value associated with the v-th feature for the m-th domain. As the parameters captured by the NDTs fluctuate throughout the day, the domain matrix values are updated accordingly. The matrix is represented as follows:

$$\mathbf{D} = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,V} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ d_{M,1} & d_{M,2} & \cdots & d_{M,V} \end{bmatrix}_{M \times V} \tag{18}$$

Moreover, the proposed procedure entails the definition of intent categories based on the service requirements levels that a consumer can articulate. These are encompassed in an intent matrix, $\mathbf{I}$, with dimensions $[N \times Z]$ and entries $i_{n,z}$, where $n \in [1, N]$ and $z \in [1, Z]$. Here, N refers to the considered intent categories, while Z represents the set of parameters specified with each service request, as detailed in Section 3. Specifically, these parameters include the set of tasks, deadline, minimum throughput, impact, risk appetite, budget, and input size of the data to be processed. The defined intent matrix is represented as follows:

$$\mathbf{I} = \begin{bmatrix} i_{1,1} & i_{1,2} & \cdots & i_{1,Z} \\ i_{2,1} & i_{2,2} & \cdots & i_{2,Z} \\ \vdots & \vdots & \ddots & \vdots \\ i_{N,1} & i_{N,2} & \cdots & i_{N,Z} \end{bmatrix}_{N \times Z} \tag{19}$$

Step 2 *- Decision matrices formulation*

The second step involves the construction of decision matrices, which serve as the input for the subsequent multi-criteria decision-making algorithm that characterizes preferences. Specifically, starting from the values expressed by the matrix $\mathbf{D}$, a domain decision matrix is constructed for each of the N intent categories. The goal of this approach is to capture the domain state when handling a service request mapped to the n-th intent category. Therefore, a given domain decision matrix is denoted by $\mathbf{D}^{(n)}$, where $n \in [1, N]$, and its dimensions are $[M \times f]$. Here, M represents the number of network domains, and f corresponds to the system parameters evaluated in the optimization problem formulated in Section 4 (e.g., completion time, throughput, cost, and cyber risk). A domain decision matrix is reported as follows:

$$\mathbf{D}^{(n)} = \begin{bmatrix} {d^{(n)}}_{1,1} & {d^{(n)}}_{1,2} & \cdots & {d^{(n)}}_{1,4} \\ {d^{(n)}}_{2,1} & {d^{(n)}}_{2,2} & \cdots & {d^{(n)}}_{2,4} \\ \vdots & \vdots & \ddots & \vdots \\ {d^{(n)}}_{M,1} & {d^{(n)}}_{M,2} & \cdots & {d^{(n)}}_{M,4} \end{bmatrix}_{M \times 4} \tag{20}$$

Similarly, beginning with the values expressed by the matrix $\mathbf{I}$, an intent decision vector is generated for each of the M domains. This vector is denoted by $\mathbf{I}^{(m)}$, where $m \in [1, M]$, and possesses a size of $[N \times 1]$, where $n \in [1, N]$. These vectors signify the revenues derived from deploying, utilizing, and storing domains' resources for every requested intent. In contrast to the domain decision matrices, for a specific domain, the decision to handle a request hinges exclusively on the potential revenues from allocating and processing the consumer's request. This value, elucidated in Section 3, delineates the costs and also mirrors the potential security risks associated with deploying services. An intent decision vector is reported as follows:

$$\mathbf{I}^{(m)} = \begin{bmatrix} i^{(m)}{}_{1,1} \\ i^{(m)}{}_{2,1} \\ \vdots \\ i^{(m)}{}_{N,1} \end{bmatrix}_{N\times 1} \tag{21}$$

Step 3 *- Preference lists generation*

The third step of the procedure entails calculating the preferences, which serve as input for the matching game. These preferences quantify the satisfaction and suitability of assigning a domain to an intent, and vice versa. In this regard, the devised strategy employs the well-known TOPSIS multi-criteria decision-making algorithm to compute the lists of preferences, leveraging the features presented by the domain decision matrices and the intent decision vectors as the criteria for decisions. Here, the designed solution categorizes all criteria presented by the domain decision matrices as costs, with the exception of throughput which is classified as a benefit. This classification aligns with the revenue criterion outlined by the intent decision vector that also falls under the benefits.

According to the TOPSIS algorithm, each criterion of every decision matrix must be associated with a weight that quantifies its importance in deriving preferences. This work employs the EWM [46] to evaluate the relevance of the criteria. Specifically, EWM computes the entropy of each criterion for each column of the decision matrices, assigning higher weights to criteria with greater variability as a consequence of lower entropy. For each domain decision matrix, the resulting weight vector is denoted by $\mathbf{W}^{(n)}$:

$$\mathbf{W}^{(n)} = [w_1^{(n)}, w_2^{(n)}, w_3^{(n)}, w_4^{(n)}]. \tag{22}$$

where $\sum_{f=1}^{4} w_f^{(n)} = 1$ and $0 \le w_f^{(n)} \le 1$.

Once a weight vector is associated with every decision matrix, TOPSIS starts the procedure to provide the preference lists. Here, the adopted multi-criteria decision-making algorithm normalizes the decision matrices and multiplies their entries by the respective weights assigned through the EWM procedure, as follows:

$$\hat{d^{(n)}}_{m,f} = w_f^{(n)} \cdot \frac{d^{(n)}{}_{m,f}}{\max\limits_{m\in M}\left(d^{(n)}{}_{m,f}\right)} \tag{23}$$

Then, TOPSIS identifies the Positive Ideal Solution (PIS), encompassing the maximum value for benefit criteria and the minimum value for cost criteria. Similarly, it evaluates the Negative Ideal Solution (NIS), comprising the minimum value for benefit criteria and the maximum value for cost criteria. These are represented as follows:

$$\mathrm{PIS}^{+} = \left\{\hat{d}_{m,1}^{(n)+}, \ldots, \hat{d}_{m,4}^{(n)+}\right\} \tag{24}$$

$$\mathrm{NIS}^{-} = \left\{\hat{d}_{m,1}^{(n)-}, \ldots, \hat{d}_{m,4}^{(n)-}\right\}. \tag{25}$$

At this point, the procedure computes the Euclidean distance from each entry in the normalized domain decision matrices to the values in the PIS and NIS sets. For clarity, the computed Euclidean distances for each row of the decision matrices are denoted as d_m^+ and d_m^-, respectively, and are outlined as follows:

$$S_m^+ = \sqrt{\sum_{f=1}^{4}\left(\hat{d}_{m,f}^{(n)} - \hat{d}_{m,f}^{(n)+}\right)^2}, \quad m = 1, \ldots, M, \tag{26}$$

$$S_m^- = \sqrt{\sum_{f=1}^{4}\left(\hat{d}_{m,f}^{(n)} - \hat{d}_{m,f}^{(n)-}\right)^2}, \quad m = 1, \ldots, M. \tag{27}$$

By leveraging these distances and considering the selected criteria, TOPSIS can now compute how close each domain is to the ideal solutions when handling a service request mapped to a specified intent category. This closeness parameter, representing the preference score, is defined as the Relative Closeness (RC) and is computed for each domain in the domain decision matrix as follows:

$$RC_m = \frac{S_m^-}{S_m^+ + S_m^-}, \quad m = 1, \ldots, M \tag{28}$$

The preferences for all domains are then ranked in decreasing order of their RC.

This procedure is iteratively applied to each domain decision matrix, where the rows correspond to the behavior of each domain in response to a service request within the n-th intent category. Consequently, the designed orchestration process produces a prioritized list of provider domains for each specified request $R_s^{(n)}$, as illustrated below:

$$R_s^{(n)}{}_{\mathrm{pref}} = (D_1, D_2, \ldots, D_M). \tag{29}$$

The ranking presented in eq.29 suggests that the $R_s^{(n)}$-th request would be more effectively managed by domain D_1 compared to domain D_2. This preference relation is denoted as $D_1 \succ_{R_s^{(n)}} D_2$.

Similarly, the orchestration procedure is performed for each intent decision vector, generating a preference list for every domain D_m. This can be expressed in the following form:

$$D_{m\mathrm{pref}} = (I_1, I_2, \ldots, I_N). \tag{30}$$

This ordering implies that domain D_m has a stronger preference for handling the I_1-th intent over the I_2-th intent. Formally, this preference is represented by the relation $I_1 \succ_{D_m} I_2$.

Both domain and intent preference lists then serve as inputs to the subsequent matching game step.

***Step 4** - Many-to-many matching game*

The final step involves executing the many-to-many matching game to establish associations between consumer service requests and domains providing services, enabling optimal resource configurations. Specifically, considering the set of network domains $\mathbb{D}$ and service requests $\mathbb{R}$ as disjoint and finite sets of players, the associations among them are determined based on their individual preferences defined in the previous step. The game produces a function λ that is defined as follows:

$$\lambda : \mathbb{D} \cup \mathbb{R} \rightarrow \mathbb{D} \cup \mathbb{R}. \tag{31}$$

This function represents a match if it satisfies the following conditions:

- *Condition 1*: each domain $D_m \in \mathbb{D}$ is matched with a subset of service requests mapped to the n-th intent category, i.e., $\lambda(D_m) \subseteq \mathbb{R}^{(n)} \in \mathbb{R}$.
- *Condition 2*: each service request $R_s^{(n)} \in \mathbb{R}^{(n)}$ is matched with a subset of domains, i.e., $\lambda(R_s^{(n)}) \subseteq \mathbb{D}$.
- *Condition 3*: each domain $D_m \in \mathbb{D}$ can process at most $q^{(m)}$ tasks composing service requests, i.e., $\mid \lambda(D_m) \mid \leq q^{(m)}$.
- *Condition 4*: a domain $D_m \in \mathbb{D}$ is matched with a service request $R_s^{(n)} \in \mathbb{R}$ if and only if the service request $R_s^{(n)}$ is matched to the domain D_m,
  i.e., $R_s^{(n)} \in \lambda(D_m) \iff D_m \in \lambda(R_s^{(n)})$.

The individual players' preferences generated through TOPSIS are complete, meaning every player ranks all players in the other set. Dealing with complete preferences is necessary because the constraint (16) of the formulated optimization problem requires that all tasks composing service requests be assigned to at most one network domain, even if individual domains may fail to meet the service requirements. Additionally, preferences are strict, indicating that players exhibit a strict order of preference towards other players.

To enable optimal resource configurations, the matching game must result in a stable outcome. To achieve this, λ must be individually rational, meaning that no player would prefer being unmatched over being matched with another player. Additionally, λ must be pairwise stable, meaning that there are no blocking pairs in the game.

Accordingly, the proposed stable many-to-many matching algorithm is based on the well-known Deferred Acceptance Algorithm (DAA) [47]. The details are provided in the pseudocode presented in *Algorithm 1*. The ultimate goal of the designed game is not only to provide a stable matching between service requests and provider domains, but also to investigate the possibility of distributing tasks of the same request to different providers in order to achieve better performance. In *steps 1-2*, each unassigned service request sends a proposal to its most preferred domain. Every domain that receives at least one proposal performs the following operations:

Algorithm 1 Many-to-many matching algorithm

```
    Input: $R_s^{(n)}{}_{\mathrm{pref}} \forall R_s^{(n)} \in \mathbb{R}^{(n)}, D_{m\mathrm{pref}} \forall D_m \in \mathbb{D}$
    Output: $\lambda : \mathbb{D} \cup \mathbb{R} \rightarrow \mathbb{D} \cup \mathbb{R}$
 1: while $\exists R_s^{(n)} \in \mathbb{R}^{(n)}$ not completely assigned do
 2:     Send proposal to its most preferred $D_m$
 3:     if $D_m$ busy cores < $q^{(m)}$ then
 4:         if $D_m$ can accept the entire request then
 5:             Assign $R_s^{(n)}$ to $D_m$
 6:         else
 7:             Assign $T_k^{(n)}$ of $R_s^{(n)}$ to $D_m$
 8:         end if
 9:         Update $D_m$ busy cores and $R_s^{(n)}$ to be assigned
10:     else
11:         Find $R_{s'}^{(n)}$ in $\lambda(D_m) \mid R_s^{(n)} \succ_{D_m} R_{s'}^{(n)}$
12:         if $\exists R_{s'}^{(n)}$ then
13:             Remove $R_{s'}^{(n)}$ from $\lambda(D_m)$
14:             Remove $D_m$ from $\lambda(R_{s'}^{(n)})$
15:             if $D_m$ can accept the entire request then
16:                 Assign $R_s^{(n)}$ to $D_m$
17:             else
18:                 Assign $T_k^{(n)}$ of $R_s^{(n)}$ to $D_m$
19:             end if
20:             Update $D_m$ busy cores and $R_s^{(n)}$ to be assigned
21:         else
22:             Not assign $R_s^{(n)}$ to $D_m$
23:         end if
24:     end if
25:     Delete $D_m$ from $R_s^{(n)}{}_{\mathrm{pref}}$
26: end while
```

- If all of its computational cores are not currently utilized, the domain evaluates whether it can process the entire $R_s^{(n)}$-th service request or if it has the capacity to process only some tasks, and then accepts the request accordingly (*steps 3-9*).
- If it reaches its quota and has previously accepted a request less preferred than the incoming one (e.g., $R_{s'}^{(n)}$), the game replaces the less preferred request with the most preferred one accordingly (*steps 11-22*). Otherwise, $R_s^{(n)}$ is not assigned to its D_m-th preferred provider domain.

After completing the matching procedure (*step 1-24*), the first request's preference is removed from the $R_s^{(n)}{}_{\mathrm{pref}}$ list. Then, the algorithm iterates until all service requests have been completely matched.

Designing such a strategy ensures that service requests are matched to domains according to their ordered preference lists, ensuring they are matched with domains beneficial to them and guaranteeing an individually rational matching. Similarly, domains accept beneficial proposals based on their preferences.

Moreover, the matching λ prevents the formation of blocking pairs. Furthermore, given that all preferences are strict, this algorithm yields a matching beneficial for the players initiating the procedure and sending out proposals first. Therefore, the whole procedure can be considered service request-optimal and indicates that the matching is the best possible outcome for all service requests, given the constraints of stability. Consequently, the outcome of the matching game identifies the combination of domains that ensures the optimal resource configuration. This is achieved by leveraging preferences that capture the service requirements specified by consumers.

6. Numerical Results

This section outlines the environmental setup and reviews the results from the simulation campaigns obtained using a Matlab script. These campaigns evaluate the effectiveness of the proposed service orchestration model in deploying B5G service requests mapped to intents, with the aim of minimizing SLAs violations.

6.1. Environmental setup

The investigated environment considers 10 network domains that exhibit varying computing capabilities, likelihood of security attacks, countermeasure investments, resource management costs, and available bandwidth, as detailed in Section 3. These parameters are uniformly sampled from the specified ranges reported in Table 3. Specifically, the processing capabilities of domains are expressed in megacycles per second [megacycles/s] and range from 40 to 2000. The maximum quota of tasks that a domain can handle varies from 750 to 2800. The likelihood of attack success is assumed to be in the range [0, 1]. The costs are defined by assessing the prices for accessing and using cloud resources. In this regard, the usage cost $C_{ut}^{(m)}$ is expressed in dollars per second [\$/s] and ranges from 1 to 2. Meanwhile, the lifecycle cost components $C_{lf}^{(m)}$ and the security countermeasure costs $C_{sec}^{(m)}$ are expressed in dollars [\$] and vary in the ranges [0.001, 0.002] and [0.003, 0.005], respectively. Finally, the transmission rate $U_{(R_s^{(n)},D_m)}$ between consumers and the provider domain is expressed in megabits per second [Mbps] and is set in the range [100, 1000]. The parameters of the aforementioned network domains are captured by their NDTs and updated every 30 minutes.

The environment also includes an indefinite number of consumers who demand service requests mapped through the designed architecture into categories of intents. The considered number of service requests ranges from 750 to 2500 (i.e., 750, 1000, 1500, 2000, 2500) each 30 minutes. Concurrently, five classes of B5G services have been investigated to define the intent categories. These classes include E-Health, Massive Twinning, Mixed-Reality, Industrial, and Smart Cities use cases and services [48], as shown in Table 2. These can be mapped to the three common scenarios defined by ITU-T IMT for 2020 and beyond [49], including enhanced Mobile BroadBand (eMBB), massive Machine Type Communications (mMTC), and Ultra-Reliable and Low Latency Communications (URLLC). Specifically, the E-Health class aligns with the URLLC usage scenario due to its stringent latency requirements, which prioritize real-time communications. Additionally, it demands a high level of reliability, security, and privacy due to the sensitive nature of the critical data being transmitted. The Massive Twinning class also corresponds to the URLLC usage scenario, given its requirements for extremely high reliability, low latency, and precise synchronization in industrial control, decision-making, and operations adaptation. These characteristics allow the managing of critical situations, real-time simulations, and dynamic reconfigurations in industrial environments. The Mixed-Reality class is associated with delivering high-quality, immersive experiences that demand high bandwidth, low latency, and high data rates, key characteristics of eMBB. The fully immersive experience and real-time context awareness necessitate the high-capacity data transmission that eMBB aims to provide. This class primarily involves outdoor environments with a high density of users, with safety considerations mainly related to the movement of people in urban scenarios. The Industrial class aligns with both the URLLC and eMBB usage scenarios due to the required precise positioning, synchronization, and the need for reliable, low-latency, and high-data-rate communications. Finally, the Smart Cities class pertains to both the eMBB and mMTC scenarios, requiring high data rates to support the collection, transmission, and processing of data from sensors, alongside robust and resilient communication solutions capable of handling a large number of connections. In general, network security and trustworthiness are crucial due to the volume and nature of the collected data. To quantify the parameters and metrics of the system model, the parameters listed in Table 2 have been considered, with values defined based on the requirements of B5G services and the characteristics described above in terms of the time deadline, expressed in seconds [s], throughput, expressed in [Mbps], budget, impact, and risk appetite. Besides, the described B5G service classes consider the number of tasks and the size of the request that can be mapped to each category.

In this context, the consumer budget, expressed in dollars [\$], varies in the range $[10, 120]$. Impact and risk appetite are both represented with values in the range $[1, 3]$. Generally, higher impact indicates more severe consequences caused by an attack, while higher risk appetite reflects a consumer's willingness to accept greater risk. The number of tasks ranges from 5 to 10, and the input size, $B^{(n)}$, expressed in megabits [Mb], has been sampled from the range $[0.6, 1.2]$ for each category. All these specified ranges have been set based on the type of deployment and devices involved in the specific service, as noted in [48].

Therefore, 48 test runs are considered for every simulation, spanning from 00:00 to 23:30 at runtime.

To improve the evaluation of the conceived methodology, four different evaluation scenarios have been considered, namely S1, S2, S3, and S4. According to the most common B5G needs, these correspond to specific use cases encompassing the majority of service requests. These needs typically entail stricter deadlines, higher throughput, lower budgets, and higher cyber risk requirements, respectively. For each scenario, the results are obtained by considering 50 different seeds, accounting for varying distributions of service requests These results are averaged over a 2.5-hour time window, sliding by half an hour, within a 80% confidence interval.

Scenario S1 involves 80 % of articulated service requests with stricter deadlines, associated with E-Health and Mixed-Reality B5G services. Conversely, scenario S2 involves 60 % of consumer requests requiring high throughput. Scenario S3 considers 80 % of generated service requests mapped to intents associated with lower budgets. Lastly, scenario S4 entails 80 % of service requests mapped to risky intent categories, associated with higher impacts and lower risk appetite. These refer to the E-Health, Industrial, and Smart Cities services.

6.2. *Baseline approaches*

The performance of the proposed solution has been evaluated against three baseline approaches, each distinguished by their respective methods of distributing service requests across provider domains and their ability to decompose services into individual tasks.

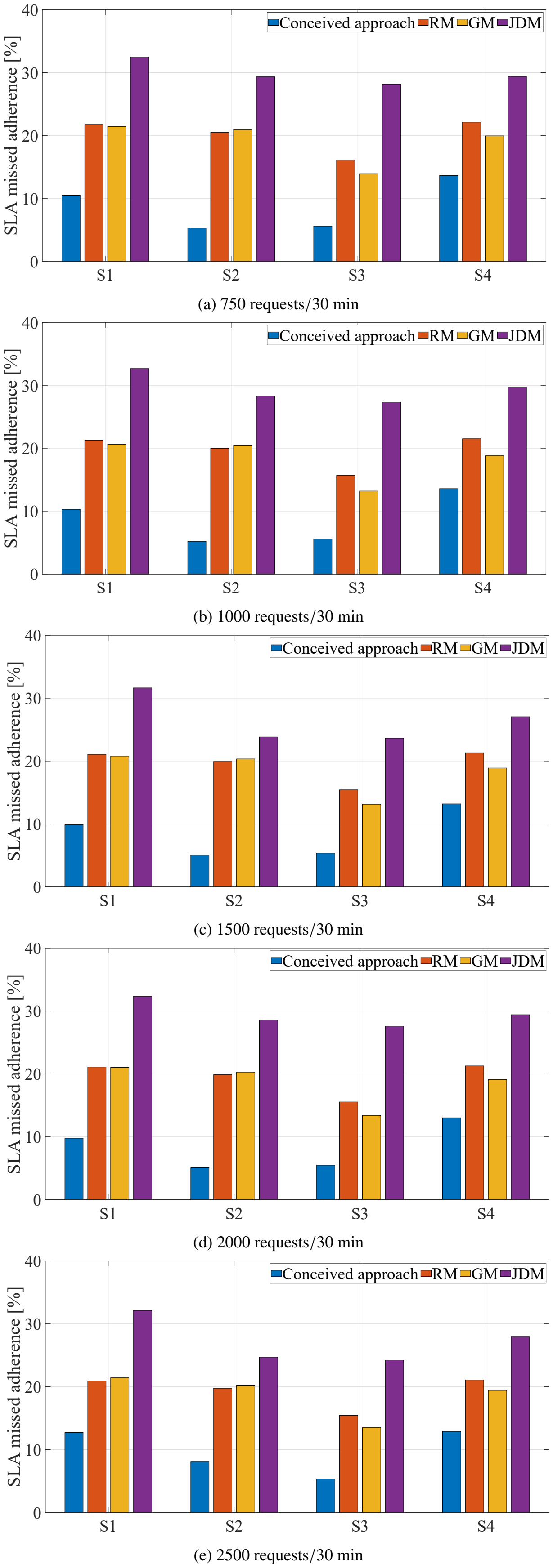
The considered baseline approaches are described below:

- Random Matching (RM): associations between service requests and available domains are executed randomly, without taking into account any specific preferences.
- Greedy Matching (GM): service requests are greedily assigned to domain providers by iteratively selecting the available and most favorable option based on a given criterion. Specifically, service requests are assigned a priori to the domain that guarantees the lowest end-to-end delay, highest throughput, lowest costs, and lowest cyber risk in the S1-th, S2-th, S3-th, and S4-th scenarios, respectively.
- Joint Decision Making (JDM): service requests are assigned to domains using the algorithm proposed in our previous work [19]. This approach does not include the decomposition of service requests into tasks. Instead, it computes service request preferences through the TOPSIS algorithm and assigns them to the highest-ranked available provider domain without performing a matching procedure.

6.3. *Service Level Agreements adherence Analysis*

Firstly, the performance of the proposed approach is evaluated against baseline solutions in terms of SLA missed adherence for each scenario (e.g., S1, S2, S3, S4) and for varying numbers of service requests generated per day (e.g., 750, 1000, 1500, 2000, 2500 requests every 30 minutes). For each generated service request, the average number of inconsistencies related to deadline, throughput, budget, and risk appetite with respect to the established SLAs is calculated. This value is assessed and averaged for all service requests generated every 30 minutes and expressed as a percentage.

The percentages in Fig. 3 illustrate the effectiveness of the proposed model, showcasing a significant improvement in SLA



(a) 750 requests/30 min

(b) 1000 requests/30 min

(c) 1500 requests/30 min

(d) 2000 requests/30 min

(e) 2500 requests/30 min

Figure 3: SLA missed adherence.

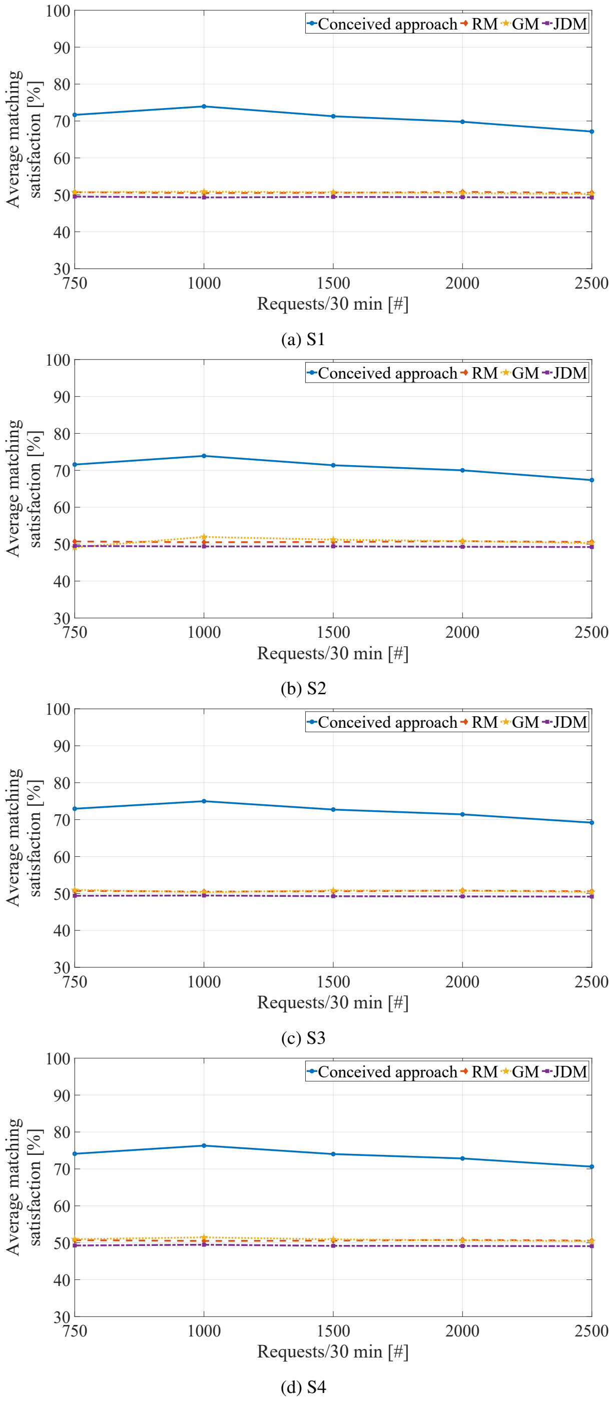
Table 2: Intent categories definition.

B5G service	Intent Category	Task [#]	Deadline [s]	Throughput [Mbps]	Impact	Risk appetite	Budget [$]
E-Health	1	5	0.200	0.1	3	1	100
Massive Twinning	2	8	2	1000	1.5	2.5	90
Mixed-Reality	3	8	0.100	100	1	3	30
Industrial	4	10	0.200	0.1	2	2.5	60
Smart Cities	5	6	1	10	1.8	2.8	10

Table 3: Simulation settings.

Parameter	Setting
M	10
N	5
Reqs/30 min	[750, 1000, 1500, 2000, 2500]
$\Omega^{(m)}$	[40, 2000] Megacycles/s
$q^{(m)}$	$[750, 2800]$
$L^{(m)}$	$[0, 1]$
$C_u^{(m)}$	$[1, 2]$ $/s
$C_{lf}^{(m)}$	$[0.001, 0.002]$$
$C_{sec}^{(m)}$	$[0.003, 0.005]$$
$U_{(R_s^{(n)}, D_m)}$	[100, 1000] Mbps
$B^{(n)}$	[0.6, 1.2] Mb

adherence compared to the worst-performing outcomes among the baseline approaches. Low percentages of SLA missed adherence indicate that the proposed approach can effectively identify suitable and available network domain resources to meet all declared service requirements. It consistently achieves a failed compliance rate below 14 % across all scenarios and settings, with a minimum of 5.04 % in the S2-th scenario with 1500 service requests. In contrast, the baseline approaches frequently report a failed compliance rate exceeding 30 %. The JDM approach, in particular, results in the highest failure rate of 32.41 % in the S1-th scenario with 750 service requests. Overall, the baseline approaches perform poorly due to their inability to thoroughly evaluate all service requirements articulated by the consumer simultaneously. Specifically, when compared to the worst-performing JDM approach, the proposed approach shows an improvement of up to 22 % for 750 service requests every 30 minutes, as reported in Fig. 3a. The proposed solution enhances SLA adherence by up to 22.31 % with the generation of 1000 service requests every 30 minutes, as illustrated in Fig. 3b. With 1500 service requests, the approach achieves an improvement of up to 22.36 %, as depicted in Fig. 3c. Furthermore, the approach demonstrates improvements of up to 22.44 % and 21.14 % with 2000 and 2500 service requests every 30 minutes, respectively, as shown in Figures 3d and 3e. These results underscore the approach's ability to identify the service provider that minimizes SLA violations for all intent categories across all reference scenarios.



(a) S1

(b) S2

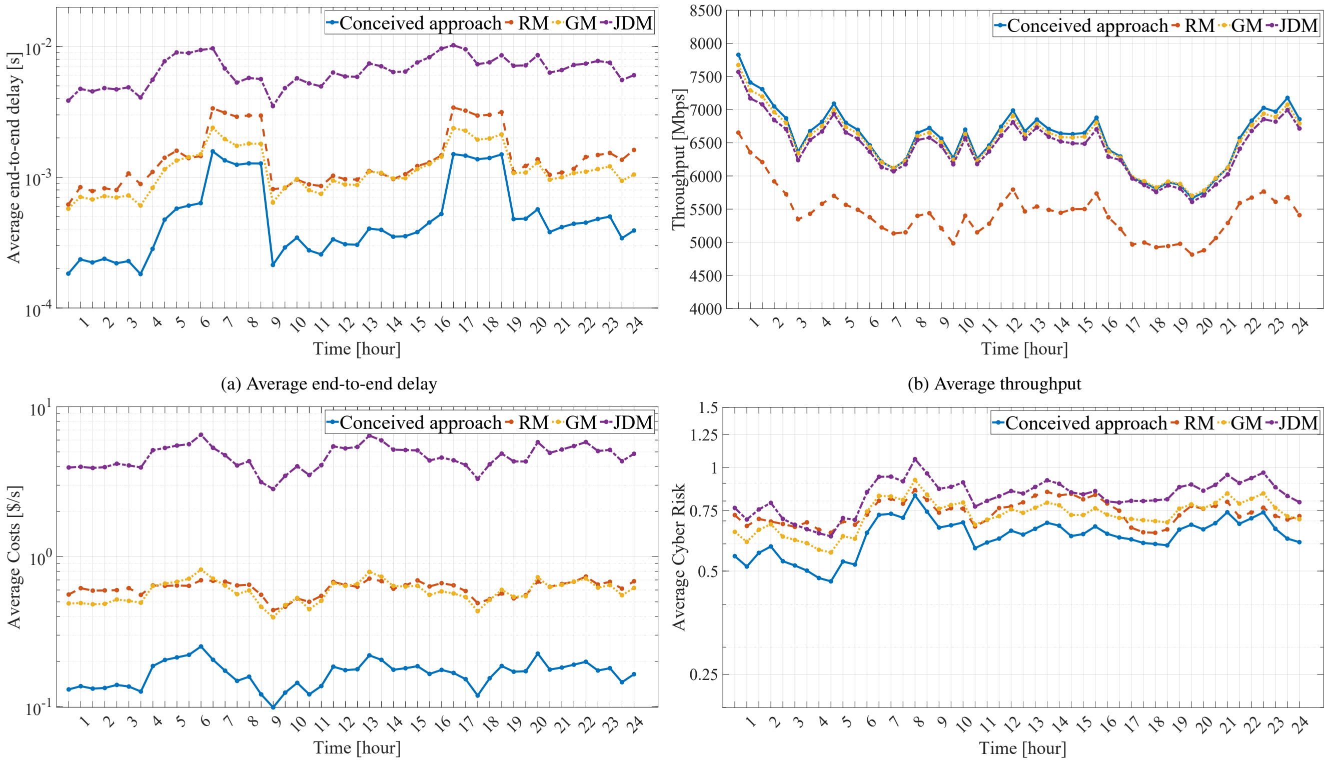
(c) S3

(d) S4

Figure 4: Average matching satisfaction per request.

6.4. Matching Satisfaction Analysis

Fig. 4 presents the average matching satisfaction for each reference scenario. It depicts the mean RC assigned to all do-



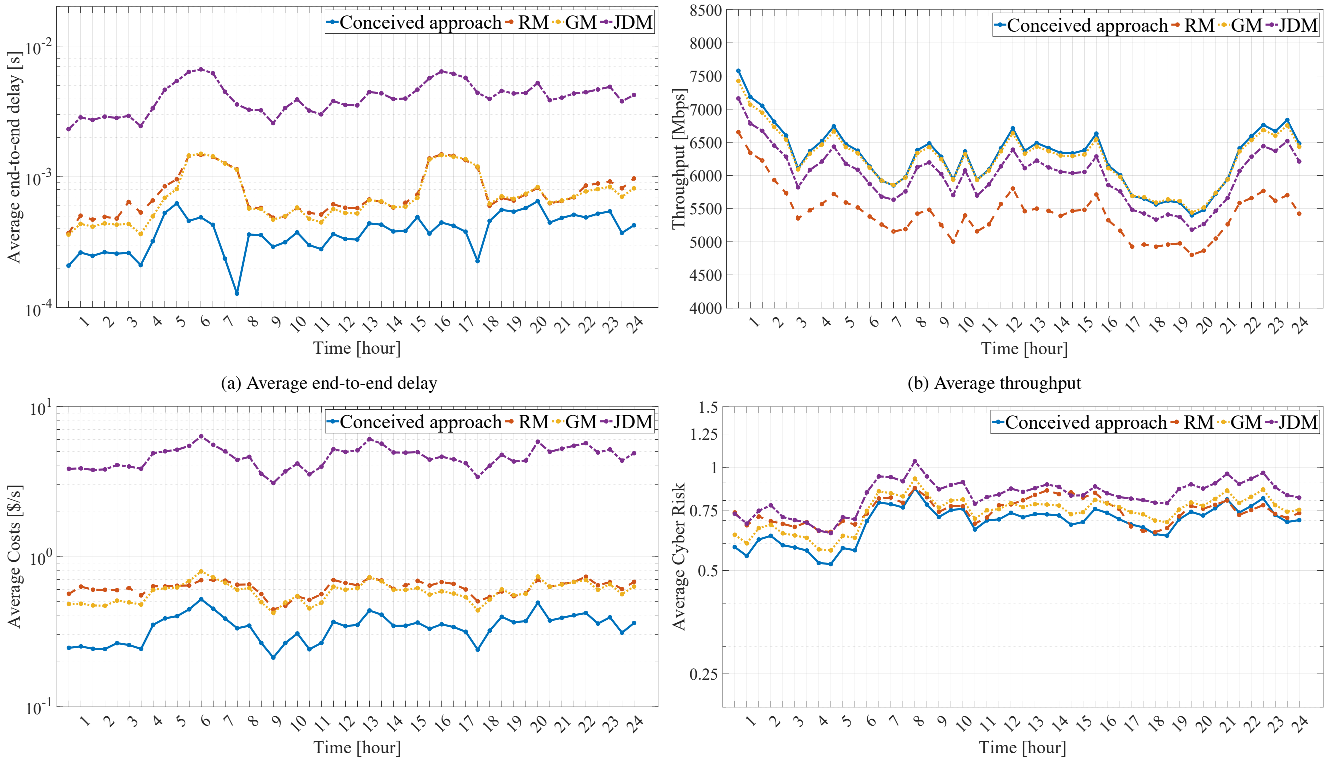
(a) Average end-to-end delay

(b) Average throughput

(c) Average costs

(d) Average cyber risk

Figure 5: KPI and KVI analysis assessed in S2 with 1500 requests/30 min.



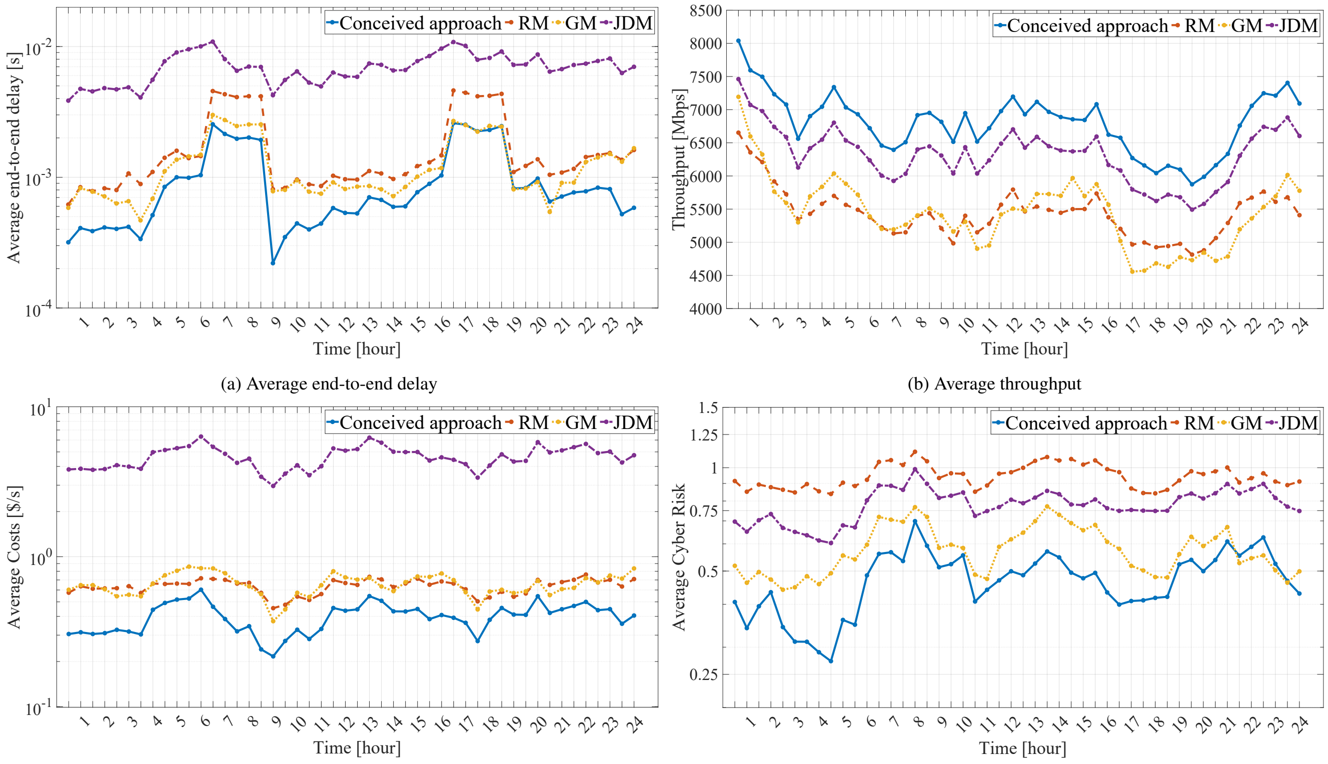
(a) Average end-to-end delay

(b) Average throughput

(c) Average costs

(d) Average cyber risk

Figure 6: KPI and KVI analysis assessed in S2 with 2500 requests/30 min.



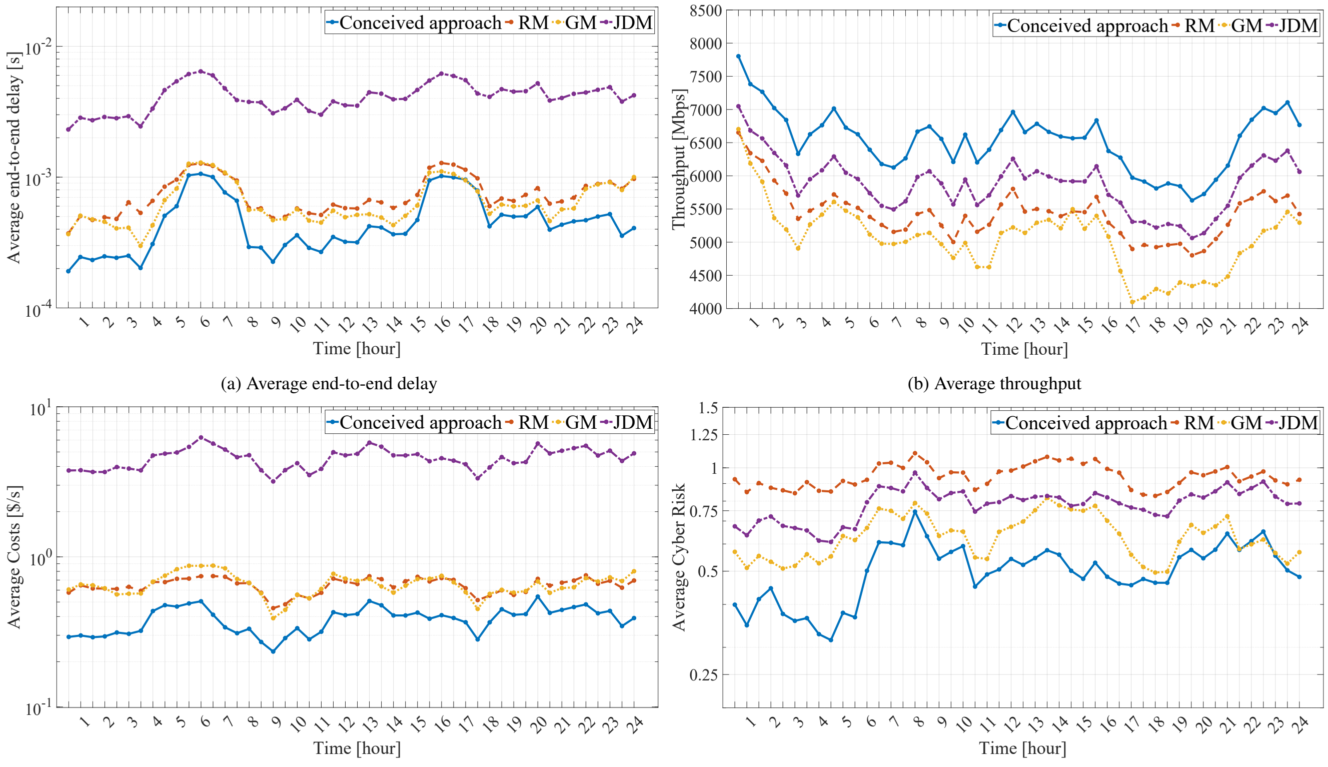
(a) Average end-to-end delay

(b) Average throughput

(c) Average costs

(d) Average cyber risk

Figure 7: KPI and KVI analysis assessed in S4 with 1500 requests/30 min.



(a) Average end-to-end delay

(b) Average throughput

(c) Average costs

(d) Average cyber risk

Figure 8: KPI and KVI analysis assessed in S4 with 2500 requests/30 min.

mains selected through the orchestration procedure to fulfill the tasks of the requested services.

Here, a higher average matching satisfaction indicates the selection of the highest-ranked domain in the computed preference list, which better meets the requirements specified by consumers. This evaluation metric is expressed as a percentage and is reported for an increasing number of generated service requests. Overall, the proposed approach demonstrates superior performance compared to baseline approaches, consistently achieving the highest average matching satisfaction, typically within the 70-80 % range across all setups. The baseline approaches, on the other hand, report an average matching satisfaction hovering around 50 % (50 % for RM and GM, 49 % for JDM). Specifically, the GM yields substantially lower results compared to the proposed approach. This is because the domain provider selection is based solely on one of the four criteria considered for deriving RC values, without accounting for any trade-off among the evaluation of other requirements. When considering the JDM, despite its initial selection of the highest-ranked domain based on relative closeness, the absence of task subdivision means that the most appropriate and available provider choice is not assured. This deficiency arises because JDM does not incorporate any matching game among preferences. Consequently, this approach ranks the lowest in terms of matching satisfaction.

6.5. Analysis of Kpi and Kvi impacting preferences

To provide further insights into the effectiveness of the proposed approach, this section presents an analysis of the evaluation criteria for provider selection, traditionally defined as KPIs. Specifically, the average end-to-end delay, expressed in seconds [s], measures the total elapsed time, encompassing both transmission and processing delays, while throughput, expressed in [Mbps], represents the transmission rate between the consumer and the selected network provider. Additionally, novel business needs and security requirements are articulated through KVIs. These include the average cost per service request, expressed in [$/s], and the average cyber risk per service request. The latter metric quantifies the potential risk a consumer faces when their service request is processed by a matched domain, weighted by the associated risk appetite.

Figures 5 and 6 illustrate the analysis in the S2-th scenario, characterized by a majority of service requests requiring higher throughput. Specifically, Fig. 5 compares the performance of the proposed approach against baseline solutions considering 1500 service requests generated every 30 minutes.

Regarding the average end-to-end delay, the novel strategy proposed in this work consistently achieves the lowest values, remaining under 0.01 seconds. In contrast, all the baseline approaches report significantly higher delays. Specifically, the average end-to-end delay with the proposed strategy is reduced by an order of magnitude compared to RM and GM, and by two orders of magnitude compared to JDM.

In terms of average throughput, the proposed approach achieves peak performance of 7800 Mbps, which is 15% higher compared to RM, which does not consider any service requirements for provider domain selection. The other baseline approaches, instead, report performance similar to the proposed approach. In this scenario, GM consistently favors provider domains that guarantee higher transmission rates, resulting in a throughput outcome very close to the optimal. Similarly, JDM achieves an average throughput that is 5% lower than the proposed approach, as the JDM methodology assigns higher weights to this KPI in its multi-criteria decision-making process, thereby favoring domains that guarantee higher transmission rates.

Also considering the average costs per service request, the proposed approach registers the most favorable results, consistently remaining under 1 $/s, whereas the RM and GM report costs that are an order of magnitude higher.

The JDM approach incurs the highest costs, nearly two orders of magnitude greater than those of the proposed solution. This is attributed to the intensive use of the same network resources by the selected domain, which could be the most expensive.

Furthermore, considering the average cyber risk, the proposed approach consistently yields results below the ones reported by the other approaches, demonstrating the capacity to select the safest domain provider for consumers' requests. Specifically, the baseline approaches report results that are 10-30 % higher if compared to the proposed approach. In particular, both the RM and GM entirely neglect the security requirement. Although the JDM approach employs the same multi-criteria decision-making algorithm, it does not perform any matching procedure. As a result, it replaces the highest-ranked but unavailable domains with lower-ranked and potentially riskier domains.

Fig. 6 confirms the efficiency of the proposed approach even with an increasing number of generated service requests. Specifically, when considering 2500 service requests per 30 minutes, the registered average end-to-end delay remains under 0.001 seconds. This is 20% lower than the delays reported by both RM and GM, and 90% lower than the delay reported by JDM.

This analysis confirms the validity of the proposed methodology compared to the baselines, demonstrating its ability to identify network providers that concurrently satisfy both KPIs and KVIs, even with an increasing number of generated service requests.

To investigate the performance of the proposed methodology under different conditions, the S4-th scenario was evaluated. In this scenario, the majority of service requests are characterized by a lower risk appetite for the demand.

In detail, Fig. 7 compares the performance of the proposed solution against baseline approaches with 1500 service requests generated every 30 minutes.

In terms of average end-to-end delay, the proposed approach consistently reports values generally under 0.001 seconds, with peaks reaching 0.003 seconds during certain hours. These fluctuations can be attributed to the varying resources exposed by domains through NDTs, which change throughout the day. Both RM and GM exhibit results up to 50% higher than those obtained with the proposed approach. The JDM once again registers the highest average end-to-end delays, doubling the results achieved with the proposed strategy.

Similar to the S2-th scenario, the proposed approach guarantees the highest average throughput, peaking at 8000 Mbps. It effectively distributes requests among domain providers to ensure high performance for all service requests. In contrast, all baseline approaches report significantly lower performance. Specifically, in this case, GM neglects the throughput requirement of the generated service requests entirely, resulting in an average throughput that is 10% lower than that achieved by the proposed strategy. RM consistently fails to identify provider domains guaranteeing high throughput by not addressing any articulated service requirements. Meanwhile, JDM reports results varying between 7500 and 5500 Mbps, often identifying providers that do not fully meet the requirements specified by consumers.

Concerning the average costs, the reported KVI for the proposed approach is consistently lower than those reported by the baseline approaches, as it is closely related to the transmission delays contributing to the overall end-to-end delay. Consequently, the average costs remain under 0.6 $/s, whereas other approaches (e.g., RM, GM) reach values 50% higher. The JDM approach, by not performing a matching procedure, selects high-ranked and available domains that may, however, utilize the most expensive resources.

Regarding the registered cyber risk, the proposed strategy exhibits provider selections similar to those reported by the GM. Indeed, this baseline approach overlooks all service requirements except for security in this scenario, thus favoring network domains that guarantee a safer processing of service requests. Conversely, the other baseline techniques identify riskier provider domains by either neglecting the security requirement (e.g., RM) or assigning requests to available but riskier domains (e.g., JDM).

The efficiency and scalability of the conceived solution are evaluated in Fig. 8, which reports the performance in the S4-th scenario with an increasing number of generated service requests. In particular, when considering 2500 service requests per 30 minutes, the approach demonstrates even better performance in terms of average end-to-end delay, with peaks reaching 0.001 seconds.

Overall, this analysis underscores both the effectiveness and scalability of the proposed approach in deploying B5G service requests mapped to different intent categories. This results in the joint minimization of SLAs violations related to latency, throughput, costs, and cyber risk requirements.

7. Conclusions

The work proposed a model for the orchestration of B5G services to minimize the number of tasks, composing service requests, not complying with the established SLAs. The conceived approach identifies the suitable resources exposed by network domains through NDTs to handle service requests appropriately mapped through intents.

Mapping service requests to intents allows consumers and providers to define and satisfy, respectively, the requested performance, business, and security demands. By jointly considering KPIs and KVIs, the service provisioning problem is modeled as a many-to-many matching game between service requests mapped to intents and domains. Preferences of the players have been computed via TOPSIS. Computer simulations have testified the validity of the conceived approach against baseline solutions in terms of missed SLA-adherence, matching satisfaction and the joint evaluation of average end-to-end delay, throughput, costs, and cyber risk. In the future, the research efforts will investigate the effectiveness of the proposed approach with an experimental testbed and in more complex scenarios entailing a larger array of KPIs and KVIs, centered around environmental sustainability, specifically addressing energy consumption and the use of energy sourced from renewable resources, and trustworthiness, which involves quantifying the reliability of a network domain. Moreover, an intent mapping scheme will be implemented to enrich the orchestrating model.

Acknowledgements

This work was supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, in the context of partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART", CUP: D93C22000910001), national center on "Sustainable Mobility" (CN00000023 - program "MOST", CUP: D93C22000410001), and partnership on "Cybersecurity" (PE00000007 - program "SERICS", CUP: D33C22001300002, project ISP5G+). It was also supported by the PRIN 2022 projects INSPIRE (grant no. 2022BEXMXN_01) and HORUS (grant no. 2022P44KA8) funded by the Italian MUR, by the HORIZON MSCA project BRIDGITISE (grant no. 101119554) and the HORIZON JU SNS project 6G-GOALS (grant no. 101139232), and by "The house of emerging technologies of Matera (CTEMT)" project funded by the Italian MIMIT.

References

[1] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, L. Hanzo, On the road to 6g: Visions, requirements, key technologies, and testbeds, IEEE Communications Surveys & Tutorials 25 (2) (2023) 905–974.

[2] C. Zhou, H. Yang, X. Duan, D. Lopez, A. Pastor, Q. Wu, M. Boucadair, C. Jacquenet, Digital Twin Network: Concepts and Reference Architecture, Internet-Draft draft-irtf-nmrg-network-digital-twin-arch-04, Internet Engineering Task Force (Oct. 2023).

[3] ITU, Digital twin network - Requirements and architecture, Recommendation ITU-T Y Y.3090, International Telecommunication Union Std (Feb. 2022).

[4] P. K. Thiruvasagam, A. Chakraborty, A. Mathew, C. S. R. Murthy, Reliable placement of service function chains and virtual monitoring functions with minimal cost in softwarized 5g networks, IEEE Transactions on Network and Service Management 18 (2) (2021) 1491–1507. doi:10.1109/TNSM.2021.3056917.

[5] L. Ji, S. He, W. Wu, C. Gu, J. Bi, Z. Shi, Dynamic network slicing orchestration for remote adaptation and configuration in industrial iot, IEEE Transactions on Industrial Informatics 18 (6) (2022) 4297–4307. doi:10.1109/TII.2021.3131355.

[6] J.-S. Tsai, I.-H. Chuang, J.-J. Liu, Y.-H. Kuo, W. Liao, Qos-aware fog service orchestration for industrial internet of things, IEEE Transactions on Services Computing 15 (3) (2022) 1265–1279. doi:10.1109/TSC.2020.2978472.

[7] J. Sahni, D. P. Vidyarthi, A cost-effective deadline-constrained dynamic scheduling algorithm for scientific workflows in a cloud environment, IEEE Transactions on Cloud Computing 6 (1) (2018) 2–18. doi:10.1109/TCC.2015.2451649.

[8] L. Qu, C. Assi, K. Shaban, Delay-aware scheduling and resource optimization with network function virtualization, IEEE Transactions on Communications 64 (9) (2016) 3746–3758. doi:10.1109/TCOMM.2016.2580150.

[9] C. Swain, M. N. Sahoo, A. Satpathy, K. Muhammad, S. Bakshi, J. J. P. C. Rodrigues, V. H. C. de Albuquerque, Meto: Matching-theory-based efficient task offloading in iot-fog interconnection networks, IEEE Internet of Things Journal 8 (16) (2021) 12705–12715. doi:10.1109/JIOT.2020.3025631.

[10] Y. Tao, J. Wu, X. Lin, W. Yang, Drl-driven digital twin function virtualization for adaptive service response in 6g networks, IEEE Networking Letters 5 (2) (2023) 125–129. doi:10.1109/LNET.2023.3269766.

[11] Y. Ma, F. Song, G. Pau, I. You, H. Zhang, Adaptive service provisioning for dynamic resource allocation in network digital twin, IEEE Network (2023) 1–1doi:10.1109/MNET.2023.3337245.

[12] J. Li, S. Guo, W. Liang, Q. Chen, Z. Xu, W. Xu, A. Y. Zomaya, Digital twin-assisted, sfc-enabled service provisioning in mobile edge computing, IEEE Transactions on Mobile Computing 23 (1) (2024) 393–408. doi:10.1109/TMC.2022.3227248.

[13] R. Zhang, Z. Xie, D. Yu, W. Liang, X. Cheng, Digital twin-assisted federated learning service provisioning over mobile edge networks, IEEE Transactions on Computers 73 (02) (2024) 586–598. doi:10.1109/TC.2023.3337317.

[14] N. A. Mitsiou, V. K. Papanikolaou, P. D. Diamantoulakis, T. Q. Duong, G. K. Karagiannidis, Digital twin-aided orchestration of mobile edge computing with grant-free access, IEEE Open Journal of the Communications Society 4 (2023) 841–853. doi:10.1109/OJCOMS.2023.3260165.

[15] D. Van Huynh, V.-D. Nguyen, S. R. Khosravirad, V. Sharma, O. A. Dobre, H. Shin, T. Q. Duong, Urllc edge networks with joint optimal user association, task offloading and resource allocation: A digital twin approach, IEEE Transactions on Communications 70 (11) (2022) 7669–7682. doi:10.1109/TCOMM.2022.3205692.

[16] T. Liu, L. Tang, W. Wang, X. He, Q. Chen, X. Zeng, H. Jiang, Resource allocation in dt-assisted internet of vehicles via edge intelligent cooperation, IEEE Internet of Things Journal 9 (18) (2022) 17608–17626. doi:10.1109/JIOT.2022.3156100.

[17] Y. Gong, Y. Wei, Z. Feng, F. R. Yu, Y. Zhang, Resource allocation for integrated sensing and communication in digital twin enabled internet of vehicles, IEEE Transactions on Vehicular Technology 72 (4) (2023) 4510–4524. doi:10.1109/TVT.2022.3228583.

[18] B. Hazarika, K. Singh, C.-P. Li, A. Schmeink, K. F. Tsang, Radit: Resource allocation in digital twin-driven uav-aided internet of vehicle networks, IEEE Journal on Selected Areas in Communications 41 (11) (2023) 3369–3385. doi:10.1109/JSAC.2023.3310048.

[19] F. de Trizio, G. Sciddurlo, I. Cianci, D. Striccoli, G. Piro, G. Boggia, Surviving disaster events via dynamic in-network processing assisted by network digital twins, in: 2023 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), 2023, pp. 1–6. doi:10.1109/ICT-DM58371.2023.10286937.

[20] M. A. Uusitalo, P. Rugeland, M. R. Boldi, E. C. Strinati, P. Demestichas, M. Ericson, G. P. Fettweis, M. C. Filippou, A. Gati, M.-H. Hamon, M. Hoffmann, M. Latva-Aho, A. Pärssinen, B. Richerzhagen, H. Schotten, T. Svensson, G. Wikström, H. Wymeersch, V. Ziegler, Y. Zou, 6g vision, value, use cases and technologies from european 6g flagship project hexa-x, IEEE Access 9 (2021) 160004–160020. doi:10.1109/ACCESS.2021.3130030.

[21] B. Aazhang, P. Ahokangas, H. Alves, M.-S. Alouini, J. Beek, H. Benn, M. Bennis, J. Belfiore, E. Strinati, F. Chen, K. Chang, F. Clazzer, S. Dizit, D. Kwon, M. Giordiani, W. Haselmayr, J. Haapola, E. Hardouin, E. Harjula, P. Zhu, Key drivers and research challenges for 6G ubiquitous wireless intelligence (white paper), University of Oulu, 2019.

[22] G. I. Association, et al., What societal values will 6g address? 6g infrastructure association vision and societal challenges working group, Tech. rep., 6G SNS IA (2022).

[23] H. Wymeersch, H. Chen, H. Guo, M. Furkan Keskin, B. M. Khorsandi, M. H. Moghaddam, A. Ramirez, K. Schindhelm, A. Stavridis, T. Svensson, V. Yajnanarayana, 6G Positioning and Sensing Through the Lens of Sustainability, Inclusiveness, and Trustworthiness, arXiv e-prints (2023). doi:10.48550/arXiv.2309.13602.

[24] ITU-T, Network 2030 - Architecture Framework, Technical report (tr), ITU-T Telecommunication Standardization Sector of ITU, ITU-T Recommendation FG-NET2030-Arch(2020) (2020).

[25] K. Mehmood, K. Kralevska, D. Palma, Intent-driven autonomous network and service management in future cellular networks: A structured literature review, Computer Networks 220 (2023) 109477. doi:https://doi.org/10.1016/j.comnet.2022.109477.
URL https://www.sciencedirect.com/science/article/pii/S1389128622005114

[26] A. Leivadeas, M. Falkner, A survey on intent-based networking, IEEE Communications Surveys and Tutorials 25 (1) (2023) 625–655. doi:10.1109/COMST.2022.3215919.

[27] A. Clemm, L. Ciavaglia, L. Z. Granville, J. Tantsura, Intent-Based Networking - Concepts and Definitions, RFC 9315 (Oct. 2022). doi:10.17487/RFC9315.
URL https://www.rfc-editor.org/info/rfc9315

[28] C. V. Nahum, V. H. L. Lopes, R. M. Dreifuerst, P. Batista, I. Correa, K. V. Cardoso, A. Klautau, R. W. Heath, Intent-aware radio resource scheduling in a ran slicing scenario using reinforcement learning, IEEE Transactions on Wireless Communications 23 (3) (2024) 2253–2267. doi:10.1109/TWC.2023.3297014.

[29] A. Leivadeas, M. Falkner, Vnf placement problem: A multi-tenant intent-based networking approach, in: 2021 24th Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), 2021, pp. 143–150. doi:10.1109/ICIN51074.2021.9385553.

[30] K. Dzeparoska, A. Tizghadam, A. Leon-Garcia, Emergence: An intent fulfillment system, IEEE Communications Magazine 62 (6) (2024) 36–41. doi:10.1109/MCOM.001.2300270.

[31] K. Abbas, A. Nauman, M. Bilal, J.-H. Yoo, J. W.-K. Hong, W.-C. Song, Ai-driven data analytics and intent-based networking for orchestration and control of b5g consumer electronics services, IEEE Transactions on Consumer Electronics 70 (1) (2024) 2155–2169. doi:10.1109/TCE.2023.3324010.

[32] J. Zhang, C. Yang, R. Dong, Y. Wang, A. Anpalagan, Q. Ni, M. Guizani, Intent-driven closed-loop control and management framework for 6g open ran, IEEE Internet of Things Journal 11 (4) (2024) 6314–6327. doi:10.1109/JIOT.2023.3312795.

[33] E. ZSM, Zero-touch network and service management (zsm); intent-driven autonomous networks; generic aspects, Tech. rep., ETSI, ISG ZSM, ETSI GR ZSM 011 V1.1.1 (2023-02) (2023).

[34] A. Goldsmith, Wireless Communications, Cambridge University Press, 2005.

[35] G. Zhang, F. Shen, Y. Zhang, R. Yang, Y. Yang, E. A. Jorswieck, Delay minimized task scheduling in fog-enabled iot networks, in: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), 2018, pp. 1–6. doi:10.1109/WCSP.2018.8555532.

[36] G. Strupczewski, Defining cyber risk, Safety Science 135 (2021) 105143. doi:https://doi.org/10.1016/j.ssci.2020.105143.
URL https://www.sciencedirect.com/science/article/pii/S0925753520305397

[37] C. Biener, M. Eling, J. H. Wirfs, Insurability of cyber risk: An empirical analysis, The Geneva Papers on Risk and Insurance-Issues and Practice 40 (2015) 131–158.

[38] G. Stergiopoulos, D. Gritzalis, V. Kouktzoglou, Using formal distributions for threat likelihood estimation in cloud-enabled it risk assessment, Computer Networks 134 (2018) 23–45. doi:https://doi.org/10.1016/j.comnet.2018.01.033.
URL https://www.sciencedirect.com/science/article/pii/S1389128618300446

[39] A. Roy, D. S. Kim, K. S. Trivedi, Cyber security analysis using attack countermeasure trees, in: Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research, 2010, pp. 1–4.

[40] A. Ofori-Yeboah, R. Addo-Quaye, W. Oseni, P. Amorin, C. Agangmikre, Cyber supply chain security: A cost benefit analysis using net present

value, in: 2021 International Conference on Cyber Security and Internet of Things (ICSIoT), 2021, pp. 49–54. doi:10.1109/ICSIoT55070.2021.00018.

[41] J. Kleinberg, E. Tardos, Algorithm Design, Addison-Wesley Longman Publishing Co., Inc., USA, 2005.

[42] Y. Gu, W. Saad, M. Bennis, M. Debbah, Z. Han, Matching theory for future wireless networks: fundamentals and applications, IEEE Communications Magazine 53 (5) (2015) 52–59. doi:10.1109/MCOM.2015.7105641.

[43] J. Ren, F. Xia, X. Chen, J. Liu, M. Hou, A. Shehzad, N. Sultanova, X. Kong, Matching algorithms: Fundamentals, applications and challenges, IEEE Transactions on Emerging Topics in Computational Intelligence 5 (3) (2021) 332–350.

[44] S. Bayat, Y. Li, L. Song, Z. Han, Matching theory: Applications in wireless communications, IEEE Signal Processing Magazine 33 (6) (2016) 103–122. doi:10.1109/MSP.2016.2598848.

[45] E. Roszkowska, Multi-criteria decision making models by applying the topsis method to crisp and interval data, Multiple Criteria Decision Making/University of Economics in Katowice (2011).
URL https://api.semanticscholar.org/CorpusID:2363622

[46] A. Jahan, F. Mustapha, S. Sapuan, M. Ismail, M. Bahraminasab, A framework for weighting of criteria in ranking stage of material selection process, International Journal of Advanced Manufacturing Technology 58 (2012) 411–420. doi:10.1007/s00170-011-3366-7.

[47] A. Roth, Deferred acceptance algorithms: History, theory, practice, and open questions, International Journal of Game Theory 36 (2008) 537–569. doi:10.1007/s00182-008-0117-6.

[48] B. Masood Khorsandi, et al., Targets and requirements for 6g - initial e2e architecture, Deliverable 1.3, A flagship for B5G6G vision and intelligent fabric of technology enablers connecting human, physical, and digital worlds (2022).
URL https://hexa-x.eu/deliverables/

[49] ITU, IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond, Recommendation ITU-R M M.2083-0, International Telecommunication Union Std (Sep. 2015).