

Service Function Chain Placement in Distributed Scenarios: a Systematic Review

Guto Leoni Santos

Universidade Federal de Pernambuco

Diego de Freitas Bezerra

Universidade Federal de Pernambuco

Élisson da Silva Rocha

Universidade Federal de Pernambuco

Leylane Ferreira

Universidade Federal de Pernambuco

André Moreira

Universidade Federal de Pernambuco

Glauco Gonçalves

Universidade Federal Rural de Pernambuco

Maria Valéria Marquezini

Ericsson Research

Ákos Recse

Eotvos Lorand University: Eotvos Lorand Tudományegyetem

Amardeep Mehta

Ericsson Research

Judith Kelner

Universidade Federal de Pernambuco

Djamel Sadok

Universidade Federal de Pernambuco

Patricia Takako Endo (✉ patriciaendo@gmail.com)

Universidade de Pernambuco <https://orcid.org/0000-0002-9163-5583>

Research Article

Keywords: Network Function Virtualization, Service Function Chain, Distributed Scenarios, Systematic Review

Posted Date: July 31st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-563622/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Network and Systems Management on September 14th, 2021. See the published version at <https://doi.org/10.1007/s10922-021-09626-4>.

Service Function Chain Placement in Distributed Scenarios: a Systematic Review

Guto Leoni Santos · Diego de Freitas
Bezerra · Elisson da Silva Rocha ·
Leylane Ferreira · André Moreira ·
Glauco Gonçalves · Maria Valéria
Marquezini · Ákos Recse · Amardeep
Mehta · Judith Kelner · Djamel Sadok ·
Patricia Takako Endo

Received: date / Accepted: date

Abstract The network function virtualization (NFV) paradigm is an emerging technology that provides network flexibility by allowing the allocation of network functions over commodity hardware, like legacy servers in an IT infrastructure. In comparison with traditional network functions, implemented by dedicated hardware, the use of NFV reduces the operating and capital expenses and improves service deployment. In some scenarios, a complete network service can be composed of several functions, following a specific order, known as a service function chain (SFC). SFC placement is a complex task, already proved to be NP-hard. *Moreover, in highly distributed scenarios, the network performance can also be impacted by other factors, such as traffic oscillations and high delays.* Therefore, a given SFC placement strategy must be carefully developed to meet the network operator service constraints. In this paper, we present a systematic review of SFC placement advances in dis-

Guto Leoni Santos, Diego de Freitas Bezerra, Elisson da Silva Rocha, Leylane Ferreira, André Moreira, Judith Kelner, and Djamel Sadok
Universidade Federal de Pernambuco, Recife, Brazil
E-mail: gls4@cin.ufpe.br, {diego.bezerra, elisson.rocha, leylane.silva, andre.jk, jamel}@gprt.ufpe.br

Glauco Gonçalves at Universidade Federal Rural de Pernambuco, Recife, Brazil
E-mail: glaucogoncalves@ufrpe.br · Maria Valéria Marquezini
Ericsson Research, Indaiatuba, Brazil
E-mail: maria.marquezini@ericsson.com

Ákos Recse
Eötvös Loránd University, Budapest, Hungary
E-mail: akosrecse@inf.elte.hu

Amardeep Mehta
Ericsson Research, Kista, Sweden
E-mail: amardeep.mehta@ericsson.com

Patricia Takako Endo
Universidade de Pernambuco, Recife, Brazil
E-mail: patricia.endo@upe.br

1 tributed scenarios. Differently from the current literature, we examine works
2 over the last 10 years which addressed this problem while focusing on dis-
3 tributed scenarios. We then discuss the main scenarios where SFC placement
4 has been deployed, as well as the several techniques used to create the place-
5 ment strategies. We also present the main goals considered to create SFC
6 placement strategies and highlight the metrics used to evaluate them.
7

8 **Keywords** Network Function Virtualization · Service Function Chain ·
9 Distributed Scenarios · Systematic Review
10

11 1 Introduction

12 With the rise of 5G and Internet of Things (IoT) technologies, it is expected
13 that the network traffic and the number of active devices increase dramati-
14 cally over the next few coming years [3]. In this context, different applications
15 with different network requirements will demand customized network configu-
16 rations [102]. Additionally, guaranteeing Service Level Agreements (SLA) via
17 automation in a highly dynamic environment – as in the 5G and IoT scenarios
18 where the number of connected users, services and data traffic vary over time
19 – represents a considerable challenge.
20

21 Traditionally, network functions (e.g. proxies, network address translation,
22 and firewalls) are offered as proprietary closed hardware-driven implemen-
23 tations or dedicated middle-boxes [84]. Each vendor provides its own man-
24 agement software and firmware, resulting in significant capital investment to
25 create and launch a new service [97]. This lack of flexibility in traditional
26 networks turns their integration and deployment of new service cumbersome
27 while compromising network management since middle-boxes have expensive
28 and slow provisioning cycles besides facing waste of resources in low traffic due
29 to difficult re-purposed and must be dimensioned at peak loads [100, 86].
30

31 NFV promises virtual services deployment in a more cost-efficient and flex-
32 ible manner [100]. Under such a paradigm, network functions are decoupled
33 from the physical network and deployed on low-cost commodity hardware us-
34 ing virtualization (running on virtual machines or simpler containers) [70].
35 In addition, the Software Defined Networking (SDN) paradigm suggests the
36 separation of control and data planes when managing network flows. Despite
37 their separate development paths, the integration between NFV and SDN can
38 increase flexibility in terms of network abstraction, infrastructure sharing, and
39 reconfigurability [11]. For instance, SDN can dynamically be used to control
40 network flows and routes based on criteria that could involve, e.g. network
41 traffic condition [100] and control load.
42

43 When creating a new service based on NFVs, there are some cases in which
44 the network manager needs to specify the service functions and their respective
45 execution order (or part of them). This way, a Service Function Chain (SFC) is
46 created to formally represent such service. A SFC is composed of a set of VNFs
47 and virtual links that connect them, where the VNFs have an execution order
48 [28]. It is also worth that different SFCs from different tenants may share the
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 same hardware components (physical machines and network links) to improve
2 resource utilization and hence operating cost.
3

4 SFC orchestration consists of determining “a feasible path on the physical
5 network where the vertices and edges of the path can satisfy both the comput-
6 ing resource requests of VNFs and the bandwidth demands of virtual links”
7 [102], but there are also cases such as multi-domain SFC networks, where the
8 VNFs need to be deployed in different domains through non-linear paths [121].
9 However, SFC orchestration can be a complex task. For example, considering
10 a large data center hosting hundreds of racks with thousands of servers and
11 dozens of network components [7], allocating the VNFs and maximizing re-
12 source usage in this kind of complex scenario is quite challenging. As demon-
13 strated by Kuo et al. [56], SFC mapping is an NP-hard problem, in which
14 obtaining an optimal solution is often difficult and time-consuming.
15

16 Furthermore, additional challenges emerge when a distributed scenario
17 is considered. Different from centralized scenarios, when for instance, a dis-
18 tributed cloud architecture is considered, the unpredictable communication
19 delay can become a high risk factor for the time sensitive applications [74].
20 Indeed, considering a multi-domain scenario, the communication over the In-
21 ternet can make it difficult to meet delay constraints [124], and the SFC place-
22 ment solutions need to consider the high delay of links. In addition, the links
23 that connect different domains and/or data centers may have different costs,
24 and particularity need to be considered in the SFC placement solution in or-
25 der to find the optimal placement strategy [20]. Liu et al. [63] also highlighted
26 that the SFC placement in distributed scenarios in comparison to centralized
27 scenarios brings additional challenges such as confidentiality of intra-domain
28 information, domain’s local autonomy, and hardware variety. These aspects
29 must be considered in SFC placement solutions in order to ensure the network
30 manager goals meeting its constraints.
31
32

33 Another challenge faced in SFC context regards to availability in these
34 networks. As noted by Fan et. al [28], even when the availability of each vir-
35 tual function is high, the availability of the whole SFC may not be acceptable.
36 For instance, considering a linear SFC composed of three VNFs, where each
37 one has an availability equals to 95%, the availability of the whole SFC is
38 85% approximately (the multiplication of the availability of all VNFs). This
39 can become a problem for operators as different network and link events may
40 impact SFC availability, including network downtime, software failures, mis-
41 configurations, and infrastructure failures [97].
42
43

44 This paper presents a systematic review of SFC placement in distributed
45 scenarios, discussing the main work contributions in the last decade. We evalu-
46 ated the SFC placement goals, the solutions proposed for the SFC placement,
47 and the scenarios considered in the papers. We also present which metrics are
48 considered in the papers to evaluate their solutions. Finally, we discuss the
49 results found in this systematic review.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

2 Related work

Bhamare et al. [9] presented a survey of architectural approaches focused on the implementation and automation of SFC in multi-cloud environments. Authors classified the works into (a) architectural models and implementation, and (b) optimization models. The authors highlighted that despite the existence of a significant amount of research on SFC, major issues remain open for consideration, including the need for the definition of a stable and reliable SFC architecture.

A recent work from Bonfim et al. [11] presented a systematic literature review with particular emphasis on integrated NFV and SDN architectures that support SFC allocation in distributed cloud scenarios. The authors investigated the state of the art of NFV and SDN architectures and identify research topics for further improvements. Authors argued that current architectures for Mobile Edge Computing (MEC) are still immature, and better approaches are necessary to offer an efficient distributed SFC management.

The survey presented by Rotsos et al. [82] considers existing standardization efforts for the orchestration (that includes automation, coordination, and management processes) of SFCs from an operator's perspective. Authors discussed some industry challenges for the standardization; they argued that some protocol solutions are still in their infancy and require further investigation and development before they can be successfully deployed in complex and large-scale environments.

In [39], Hantouti et al. presented a survey of traffic steering for SFC, analyzing its mechanisms on the efficiency of service chaining. Authors classified the studies into three approaches: (a) based on programmable switches, it means SDN switches (b) based on switches with compact packet tags and (c) based on packet header. Based on these classifications, they analyzed a list of SDN-based SFC methods used to traffic steering. Furthermore, authors presented a qualitative evaluation of SDN-based SFC approaches, analyzing the configuration used in these studies.

In [96], Souza et al. systematically reviewed the integration of NFV in data centers. They highlighted some research problems related to performance, dependability, resource allocation, cost, management, and resource interoperability. The authors summarized the main lessons learned from the literature, and also discuss some open research challenges. They stated that "*in the case of SFCs, new challenges appear because they have to deal with the implementation of multi-vendor service functions in geographically distributed data centers*".

In [87], Schardong et al. carried out a systematic review and proposed a taxonomy for NFV resource allocation. They focused specifically on the VNF Forwarding Graph Embedding (VNF-FGE) problem. The studies were classified based on the proposed taxonomies, consisting of: Behavior (online or offline), Allocation task (Placement, Chaining or both), Objective (reduce network resources or host resources or economic cost, or trade-off between network and host resources), Factors (storage, memory, link capacity or latency, among others) and technique style (exact, heuristic or meta-heuristic). The

authors also identified challenges and deficiencies in the development of (sub-)optimal solutions for the problem in question, such as life cycle management of VNFs and comparison of proposed algorithms.

Our systematic review complements the works done by [96, 87], highlighting the existing solutions for SFC placement in distributed scenarios proposed over the last decade. Furthermore, in comparison with other surveys such as [9, 11], our systematic review provides a more comprehensive view, presenting SFC placement goals and solutions as well as the scenarios in which the solutions are applied. Finally, our work does not focus on specific metrics as presented in [39, 82].

3 SFC Definitions

As mentioned by Mirjalily et al. [72], there are four stages needed to deliver a SFC. The following stages are identified for handling possible issues related to automatic SFC deployment in network infrastructures: description, composition, placement, and scheduling, as illustrated in Figure 1. A generic SFC example is considered in Figure 1 composed of the three functions VNF1, VNF2, and VNF3.

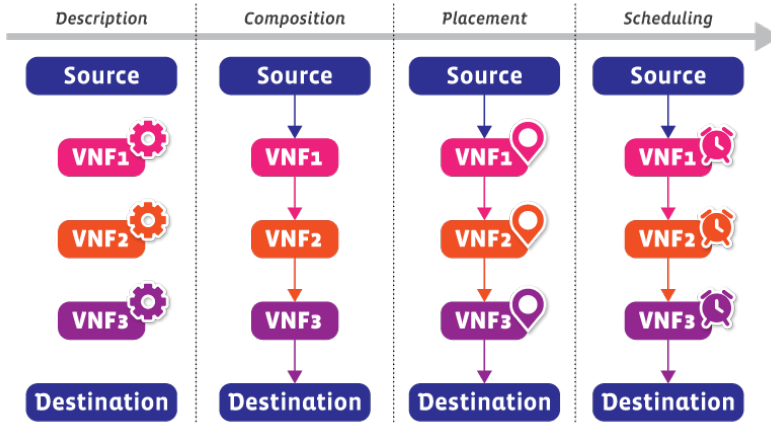


Fig. 1 Stages of SFC orchestration

The first stage is the description which is responsible for describing each VNF and detailing its functional (e.g., type of service provided) and nonfunctional (e.g., information about disk, memory, number of CPUs) properties. This stage is important for network providers to specify the functionalities and requirements of each VNF needed to make the SFC completely operational [72]. An example of SFC definition is shown in Figure 2. There are two physical nodes represented by source and destination blue boxes. From the

source, there are three chained VNFs (magenta, red, and purple boxes) with different requirements in terms of CPU, memory, and storage, to achieve the destination. The source, VNFs and destination entities are connected through links with specific bandwidth requirements. Although, we use simple cloud resource requirements for the sake of this example, in real scenarios there is a much broader set of possible requirements (e.g. affinity, anti-affinity, hardware acceleration).

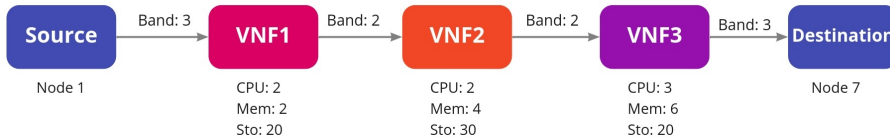


Fig. 2 Example of a SFC definition

Several alternatives can be applied for the VNF description. Both the Web Service Description Language (WSDL) [51] and the Unified Service Description Language (USDL) [16, 104] are efforts for web services description, while Cloud Service Declarative Definition Language (CSDL) and its variations [33] focus on the description of cloud services. As highlighted by Mehraghdam and Karl [69], the YANG data model can be used for the VNF description. An additional data model that can be used is the Topology and Orchestration Specification for Cloud Applications (TOSCA), which provides a data model and templates to orchestrate and manage application services with NFV in cloud environments [52].

The second stage in SFC deployment is the composition. This step defines how the VNFs will be ordered in an SFC. The second stage in SFC deployment is the composition. This step defines how the VNFs will be placed. According to Yang et al. [116], an SFC may be classified in two different types: totally ordered, in which there is a total dependency order on the VNF set, and partially ordered, which there is a dependency among a subset of VNFs, justifying the reason of the composition step. Although these functions are usually independent, their order (or part of them) must be defined carefully and is determined by the service they seek to offer. For example, considering an IoT scenario, where one can have large data to transmit, the first VNF of a SFC can perform data compression, in order to reduce the bandwidth consumption in the network [80].

The third stage is the SFC placement. It establishes where each VNF from a given chain will be hosted. The main goal of this phase is to ensure an efficient allocation, taking into account several constraints (from the infrastructure or customers) [72]. However, some authors present other additional concerns for this stage. A common approach is to divide the SFC placement into two sub steps: VNF placement and SFC chaining [55]. While placement

consists in assigning VNFs to the physical infrastructure, chaining is defined as interconnecting VNFs in order to deploy a SFC. In other words, the VNF placement defines the servers where the VNFs are deployed, and the SFC chaining defines the physical links that will connect these servers. The chaining step can be done in different ways. Kiji et al. [54] considered that there are predefined paths, calculated through Dijkstra algorithm. On the other hand, Jiao et al. [48] presented a joint optimization of VNF selection and traffic steering. A placement example of the SFC considered in Figure 1 is shown in Figure 3. The VNFs are allocated across different distributed data centers and connected through virtual links.

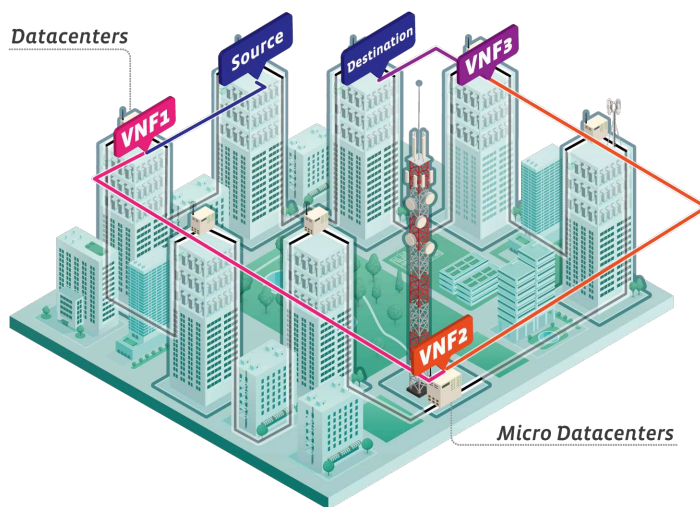


Fig. 3 SFC placement example

The fourth step is the SFC scheduling. After composing and defining which servers the VNFs will be allocated to, the scheduling step determines when that allocation will happen. The main focus of this step is to minimize the total execution time of the SFCs allocated [72]. As highlighted by Alameddine et al. [1], the scheduling step is still a field under investigation, and received little attention by researchers. Riera et al. [81] formulated the scheduling problem as a flexible job-shop problem. Mijumbi et al. [71] proposed a Tabu-search-based algorithm to solve the scheduling problem. The work presented in [79], formulated the scheduling problem as an MILP and solved it using a genetic algorithm.

The placement step is the main focus of our systematic review. We present the main solutions used for SFC placement in distributed scenarios.

4 Methodology

The methodology used in this work to guide the systematic review is based on [25], described in Figure 4. We used this methodology to find the works that addressed SFC placement in distributed scenarios. The methodology steps are described in the following subsections.

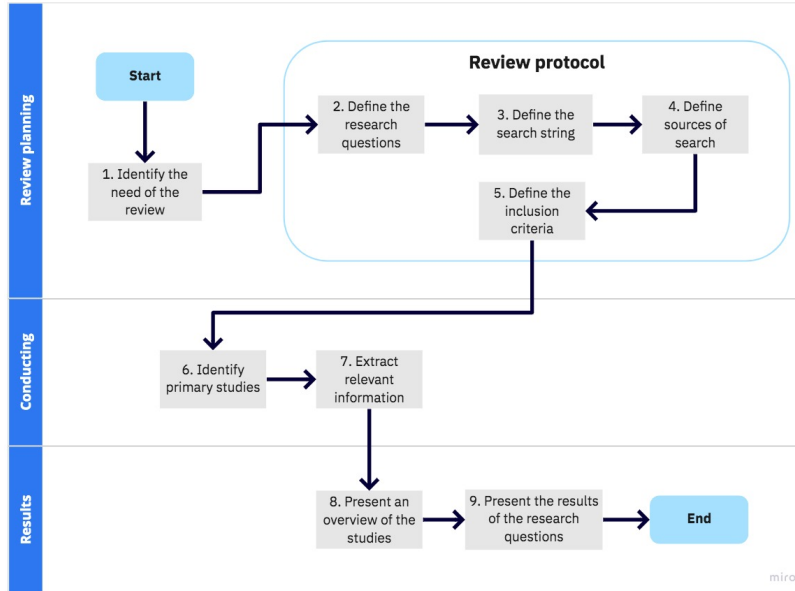


Fig. 4 Systematic review methodology

4.1 Activity 1: Identify the need for the review

As discussed previously, the use of the NFV paradigm to create virtual network functions promises to reduce the network costs while increasing the network flexibility [9]. In addition, connecting VNFs in a logical order to compose a SFC is a big challenge. Considering distributed scenarios, some aspects increase the placement complexity, such as capacity and delay constraints. Although the VNF paradigm is not a new concept and past works have presented reviews and surveys about this technology, this paradigm continues to evolve and deserves renewed attention and evaluation. Technologies, such as IoT, fog computing, CDN, e-health, stand to benefit from deploying SFCs to create customized services that meet customers requirements. In addition, technologies such as 5G brought up fundamental recent architectural changes which further raises the need for our new survey.

1 In the systematic review presented in this paper, we focus on SFC place-
2 ment in distributed scenarios, evaluating their possible applications and goals.
3 We identify the goals and constraints considered by the primary papers for
4 the placement of SFCs, as well as the most common scenarios where the SFC
5 paradigm has been applied. In addition, we identify the most common tech-
6 niques applied to create placement strategies.
7

8 9 10 4.2 Activity 2: Define research questions

11 The main goal of this work is to answer the following research questions:

- 12 – **RQ.1:** What are the main goals considered during the SFC placement?
- 13 – **RQ.2:** What are the most common techniques used to define SFC place-
14 ment strategies in distributed scenarios?
- 15 – **RQ.3:** In which scenarios the SFC placement has been commonly used?
- 16 – **RQ.4:** What are the most used metrics to evaluate SFC placement strate-
17 gies?
18
19

20 21 22 4.3 Activity 3: Define search string

23 The search string used in this work to identify relevant papers was: (*“service
24 function chain” OR SFC*) AND placement. Note that we did not include ‘dis-
25 tributed’ in the search string, because it would have restricted the search a
26 great deal, returning only few articles. Despite having other nomenclatures for
27 VNF concepts, such as the European Telecommunication Standards Institute
28 (ETSI) Network Service (NS) [32] definition, this systematic review assumes
29 the Internet Engineering Task Force (IETF) definitions.
30
31

32 33 4.4 Activity 4: Define sources of research

34 We considered the following databases as the main sources for our research:
35 IEEE Xplore¹, Science Direct², and ACM Digital Library³.
36
37
38

39 4.5 Activity 5: Define criteria for inclusion and exclusion

40 In order to limit our scope, we considered only works published in journals and
41 conferences in the last 10 years (between 2010 and 2020). A selected paper must
42 focus on SFC placement in distributed scenarios. It is important to highlight
43 that we removed works that consider SFC placement in centralized scenarios
44 and works which do not clearly define the scenario as being distributed.
45
46

47 ¹ <http://ieeexplore.ieee.org/Xplore/home.jsp>

48 ² <http://www.sciencedirect.com/>

49 ³ <http://dl.acm.org/>
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

4.6 Activity 6: Identify primary studies

In June 2020, the search returned 117, 411, and 213 works (741 in total) from IEEE Xplore, Science Direct, and ACM Digital Library, respectively. After removing duplicate and short papers, we read all their abstracts applying the criteria for inclusion and exclusion. This activity is made manually, i.e., without any support tool. We finally selected 58 works for the ultimate data extraction and evaluation.

4.7 Activity 7: Extract relevant information

After reading the 58 papers selected in Activity 6, we extracted relevant information that answers the research questions previously established in the Activity 2. It is important to highlight that this activity is made without any support tool.

4.8 Activity 8: Present an overview of the studies

An overview of all works selected in this systematic review was elaborated in this activity, which is presented in Section 5, with the purpose of classifying the papers according to the research questions formulated in Activity 2.

4.9 Activity 9: Present the results of the research questions

Finally, in this activity we discuss our findings and address our research questions regarding the works that present SFC placement in distributed scenarios.

4.10 Threats to validity

In this section, we present the limitations regarding this systematic review. We have identified the following threats to validity:

- **Research question:** we formulated our search string taking into account the terminology adopted by the IETF. Despite being a widely used term, since 741 articles were raised in our searches in the search bases, some articles may not have been included in our systematic review. However, other terminologies have been considered in recent reviews, such as [87].
- **Data extraction:** as mentioned in section 4.5, we considered only works that presented SFC placement solutions in distributed scenarios. The aim was to verify proposed solutions that considered relevant characteristics of these scenarios. However, other works that did not enter our research may present relevant solutions to the problem of SFC placement.

5 Results and discussions

5.1 Descriptive analysis

Figure 5 shows the number of articles published per year, from 2010 to 2020. The articles are very recent, mostly appearing after 2015, and up to 85% were published in the last three years: 16 in 2018, 24 in 2019, and 11 in 2020. This is a clear indication that SFC placement remains a relevant research field.

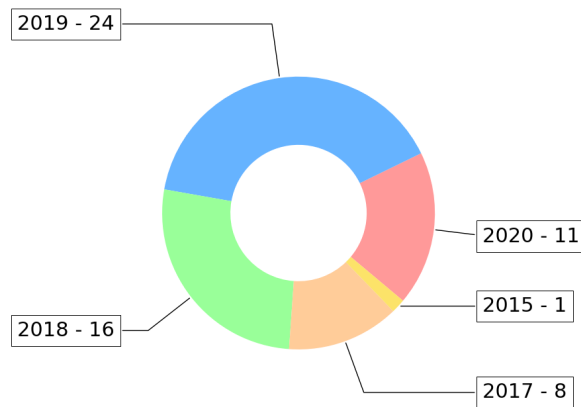


Fig. 5 Articles per year

Figure 6 shows the number of works per source. One can see that most of the articles were published in the IEEE Xplore, followed by Science Direct and ACM Digital Library with 38, 20, and 2 articles, respectively.

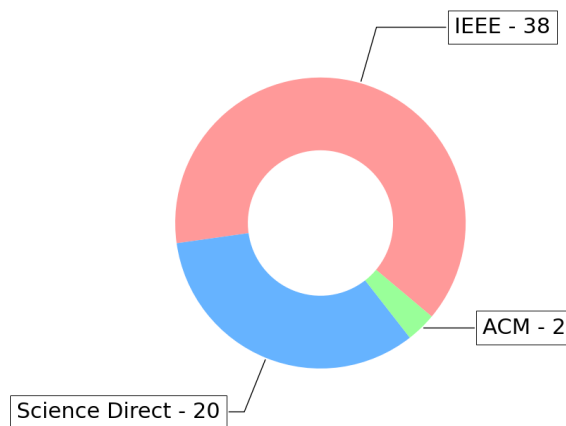


Fig. 6 Articles per source

5.2 Goals and solutions

In this section we describe the goals considered by the papers reviewed in this systematic review. We also present the solutions proposed by these articles to achieve their goals.

Most of the articles found in this systematic review proposed mono-objective solutions, i.e., only one objective function is optimized during the SFC placement process. Figure 7 shows the main placement goals found in the primary studies. The most common goal is the minimization of operational cost (nineteen papers), followed by minimization of delay and maximization of resource utilization with ten and eight articles, respectively. Some papers also addressed other specific problems, which are not related to the previous ones, and we present them in a separated subsection. Finally, we present the papers that proposed multi-objective solutions to achieve their goals, i.e. more than one objective function is considered in the SFC placement process. A more detailed analysis of each paper for the placement goal is presented in the following sections.

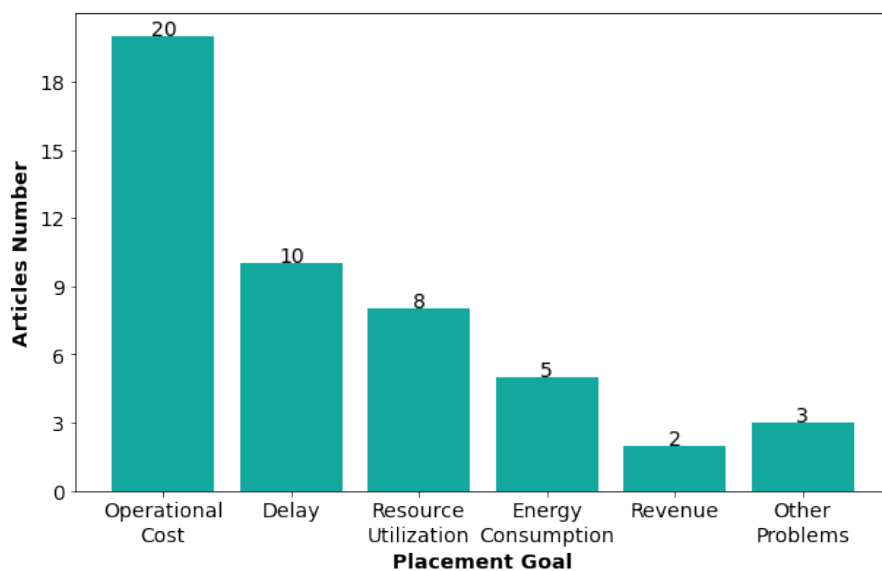


Fig. 7 Articles per placement goal for mono objective solutions

5.2.1 Operational cost

Most of the primary studies addressed the SFC operational cost since one of the biggest advantages in adopting NFV is to achieve a flexible utilization of computational and network resources. Due to migration of network functions

1 from dedicated hardware to general purpose servers both network operators'
2 operational expenditures (OPEX) and capital expenditures (CAPEX) are re-
3 duced [50]. The operational costs of a SFC can be divided in two types: costs
4 about the resources and costs about the VNF placement [78]. The former is
5 related to the remaining resources on servers and links from the physical in-
6 frastructure. The latter is about the costs related to the placement, such as
7 energy consumption, license fees, among others [12][6]. Nonetheless, obtain-
8 ing the optimal SFC placement cost requires careful planning. For instance,
9 if a SFC placement strategy takes into account only the costs regarding the
10 network bandwidth, the costs related to the node resource consumption can
11 be high, or the placement solution can result in a high VNF migration cost
12 [62]. In addition, minimizing the operational costs of SFCs can conflict with
13 other metrics, such as performance and resiliency [77].

14
15 Five articles found in this systematic review addressed cost optimization
16 for cloud computing scenarios. Chen et al. [20] considered SFC placement in
17 cloud computing scenarios, more precisely public cloud services. The main
18 goal was to minimize the overall cost taking as constraints QoS aspects, such
19 as delay. The authors formulated the optimization problem as an integer lin-
20 ear programming (ILP) and proposed a heuristic algorithm based on Hidden
21 Markov Model (HMM) to solve the problem for large scenarios.

22
23 Pei et al. [78] solved SFC placement problem with dynamic VNF place-
24 ment for a geo-distributed cloud, considering that the placement cost includes
25 resource cost and VNF placement cost. Unlike [20], the proposed solution fo-
26 cused on online SFC placement, adapting the placement strategy according to
27 dynamic network load. Authors formulated the SFC-EP with dynamic VNF
28 placement as a binary integer programming (BIP) model and proposed two
29 algorithms: SFC embedding approach and VNF Dynamic Release Algorithm.
30 The former obtains the selection, placement, and concatenation solutions of
31 VNF instances by running the shortest path algorithm (e.g., Dijkstra) in a
32 multi-layer graph; while the latter is responsible for the optimization of placed
33 VNF instances while seeking to reduce their running time.

34
35 The work proposed by Feng et al [30] also addressed online SFC placement.
36 It considered flow processing, routing decisions and at the same time the asso-
37 ciated allocation of cloud and network resources. They focused on stabilizing
38 the service input rate of cloud services with minimum cost. The proposed algo-
39 rithms are based on Lyapunov drift-plus-penalty control methodology, which
40 is applicable to the dynamic control of queue network [43, 75]. Tang et al.
41 [107] also proposed a Lyapunov based algorithm for the cost optimization for
42 SFC placement. They considered a 5G access network scenario and formu-
43 lated a model for minimizing cost considering network stability. Afterwards,
44 they transformed this problem into the minimization of the upper bound of
45 drift-penalty function by applying the Lyapunov optimization technique. A
46 genetic algorithm based heuristic was proposed for VNF scheduling and map-
47 ping.

48
49 In [37], Gupta et al. formulated a cost optimized latency aware placement
50 (COLAP) framework for optimum SFC placement in a distributed cloud. CO-
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 LAP implements a randomized selection of clouds and a heuristic for placement
2 based on the latency prediction through support vector regression. It takes as
3 input both tenant's initial requests and online requests during operation with
4 data for latency threshold and cost budgets and tariffs.
5

6 Two other articles proposed online solutions for SFC placement with concern
7 for cost minimization but did not specifically consider cloud computing
8 scenarios. Yao et al. [117] studied the cost minimization of dynamic SFC placement
9 and scaling in distributed data center networks. The authors proposed
10 an online scaling algorithm, which is composed of two parts. The first part is
11 a Fourier-series-based forecasting approach to minimize the placement cost by
12 avoiding frequent changes in network topology. The second part is an online
13 deployment algorithm to properly deploy VNF instances based on online learning.
14 An SFC placement strategy for content delivery networks was proposed by
15 Jahromi et al. [47]. They considered that the SFC is composed of Value Added
16 Services, that are functions for video delivery applications. The main goal was
17 to minimize the SFC costs while QoS of all service requests were jointly satisfied,
18 taking into account the eventual reuse and migration of already deployed
19 VNFs from previous SFC requests. The optimization problem was formulated
20 as an ILP and solved using the CPLEX optimization tool for a small scale scenarios.
21 The work proposed by Hejja and Hesselbach [44] presented an offline
22 algorithm for SFC placement with the purpose of minimizing the operational
23 cost about the SFC migration and the energy consumption cost. A heuristic
24 base algorithm is proposed to solve the placement in a problem in a reasonable
25 time. It is important to highlight that the authors also proposed an online version
26 of the algorithm, but focused on minimize the power consumption, and
27 it will be described in the Subsection 5.2.4.
28

29 The concept of reuse of already deployed VNFs within instantiated SFCs
30 was considered in two other articles in order to reduce the cost. Mohamad and
31 Hassanein [73] proposed a sharing-based placement where underutilized VNFs
32 deployed in current SFC requests should be given priority for reuse instead
33 of instantiating a new VNF. Since the authors considered an edge computing
34 scenario, sharing is important due to the limited computing resources of the
35 network devices. They formulated their problem as an ILP model in order to
36 minimize the overall deployment cost, seeking to reduce resource utilization
37 while satisfying customers' QoS requirements. The proposed ILP model was
38 solved using the Gurobi solver [38]. The main objective of work proposed by
39 Guo et al. [34] was to reduce deployment costs by sharing VNFs between different
40 SFCs in order to avoid multiple VNF instantiation across the network.
41 The authors formulated the SFC placement problem as a mixed ILP (MILP)
42 and presented an algorithm for solving it. This algorithm has two main steps:
43 an initial phase based on a centrality measure that searches servers as candidates
44 to place shared VNFs and an adjustment phase based on a Markov Decision
45 Process that can adjust the shared VNFs and allocates remaining
46 VNFs.
47

48 Two articles used reinforcement learning algorithms to optimize the cost of
49 SFC placement: [65] and [35]. In [65], Luo et al. proposed a deep reinforcement
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 learning framework that combines a recurrent neural network with a reinforcement
2 learning agent to deploy in geo-distributed data centers. The recurrent
3 neural network model was used to predict the traffic flow while the deep re-
4 inforcement learning agent was used to produce the SFC placement decisions.
5 The main goal was to reduce the overall cost of SFC deployment in the dis-
6 tributed data centers. The work presented in [35] proposed a cloud-edge SFC
7 orchestration architecture based on blockchain and deep reinforcement learn-
8 ing for IoT applications. Deep reinforcement learning was used to support
9 automatic SFC orchestration and dynamic adjustment. For the placement,
10 the SFC cost (including SFC orchestration and SFC) and constraints about
11 the CPU, and memory resources and link utilization were considered. The use
12 of reinforcement learning algorithms by these articles was carried out to han-
13 dle the current condition of the environment, and to decide the most suitable
14 SFC placement response: in [65] considering the pattern of traffic variation,
15 and in [35] taking into account a high-mobility IoT network.

17 Similarly to [35], Nguyen et al. [76] also addressed the SFC placement prob-
18 lem in an NFV-based edge cloud system considering an IoT network topology.
19 The authors proposed an optimization problem for VNF placement and chain-
20 ing taking into account aggregated traffic from IoT gateways. The main goal
21 was to find the placement solution which gives the minimum cost over the
22 allocated components. They formulated the problem as a non-convex Integer
23 Programming problem. Two algorithms were proposed to solve the problem:
24 a Markov approximation approach and a node ranking-based heuristic.

26 Jin et al. [50] addressed SFC placement over devices located at the edge
27 of the network. The authors proposed a SFC allocation solution for online
28 gaming in a fog computing scenario. The resource allocation problem was
29 modeled to optimize the total cost of SFC placement, considering CPU/GPU
30 cost, VNF instantiation cost and transmission cost, with delay, resources and
31 link capacity as constraints. Then, a heuristic algorithm named Probabilistic
32 SFC Embedding based on Cost Optimization (PSECO) was proposed to solve
33 the optimization problem. PSECO can be divided in two steps, where the
34 first one maps the VNFs into physical network node, and the latter maps the
35 virtual links among instances to the physical links based on the shortest path.
36

37 The work by Fang et al. [29] sought to guarantee SFC reliability and op-
38 timize the costs of network resources in the implementation of the VNFs. To
39 this end, the article proposed two reliability protection mechanisms: an all-
40 nodes protection mechanism and the single-node protection mechanism. For
41 each protection mechanism, the authors formulated the problem as an ILP
42 model and proposed a heuristic algorithm based on dynamic programming
43 and Lagrangian relaxation.
44

45 In [110], the optimization task was to minimize the total deployment cost
46 satisfying all the SFC requirements of the flows. SFC placement was modeled
47 as a set cover problem and two algorithms were proposed with logarithmic
48 approximation factor. The first algorithm is based on Linear programming
49 rounding, and the second one is based on the greedy approach.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 The work presented in [62] tried to minimize the cost of activating a function
2 node (a node that can host VNF). This cost is associated with energy
3 and maintenance costs needed to deploy the VNF in a node. A BIP model
4 and an algorithm were provided. Note that the model is an offline version of
5 the problem, whereas the algorithm copes with a dynamic online version.
6

7 The work [54] presented SFC placement and routing that specifies the path
8 which passes through all required VNFs while meeting service bandwidth re-
9 quirements. The main goal was to minimize total cost associated with VNF
10 placement and link usage. The authors then proposed an ILP problem for-
11 mulation, considering reduction of the number of VNF instances, reduction of
12 link usage and cost minimization. Then, a heuristic was proposed to handle
13 cases where the ILP is intractable.

14 Tastevin et al. [109] considered a problem of allocating SFCs over an in-
15 frastructure composed of commodity servers, called points of presence (PoPs),
16 with VNF hosting capabilities. PoP can be defined as an access point in an
17 IP network owned by a telecommunication carrier [19]. A PoP usually is com-
18 posed of interconnected co-located backbone routers but may also represent
19 other equipment, such as servers and switches [18]. They proposed a cost-
20 driven ILP formulation for this problem. To solve it, a graph-based heuristic
21 that combines graph centrality and multi-stage graphs was proposed.
22

23 Tashtarian et al. [108] put forward a solution for SFC placement with dif-
24 ferent strategies: NFV parallelism, NFV distribution, optimal resource alloca-
25 tion, and optimal data rate. NFV parallelism and distribution allow to execute
26 multiple replicas of a single VNF and distribute the traffic among them. The
27 authors defined only one function that jointly covers four cost aspects: the to-
28 tal cost of deployment, energy costs, distribution costs, and parallelism costs.
29 This function is considered in the placement problem, which was modeled as
30 an MILP and solved using the greedy algorithm.

31 Evaluating the works that addressed the optimization of the SFC opera-
32 tional cost as the main goal, all of them considered the allocation cost of nodes
33 (server, switches, etc.) and links. The costs about the nodes are usually re-
34 lated to the cost of allocating a VNF into the nodes according to the available
35 computational resources (CPU, memory, and storage), where more powerful
36 nodes increase the placement cost. The cost of the links are usually related to
37 the network bandwidth. However, the work presented in [110] does not provide
38 details about the operational cost. The authors only define the cost of placing
39 a VNF into a physical node.
40

41 In addition to the costs about the nodes and links, some works considered
42 other aspects of SFC operational costs. The work presented in [50], which is
43 focused on SFC for online mobile games, considered the costs of CPU and
44 GPU allocated to run the VNFs. In [30, 117] the authors considered the pro-
45 cessing cost, which is related to the additional cost of deploy/execute a VNF
46 in a node. Some works are more specific and define the VNF deployment cost
47 as a “renting cost” [65, 62, 107]. Luo et al., for instance, defined the VNF
48 deployment cost as “for renting a virtual machine or container with required
49 resource configuration for running the instance” [65]. Liu et al. [62] and Tang
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 et al. [107] also considered the renting cost of deploying a VNF in a server,
2 but such cost is inversely proportional to the amount of available resources in
3 the server. The works presented in [65] and [108] considered the operational
4 cost of migrating the VNF's image to another server. In [65], the authors con-
5 sidered the cost of copying the VNF's image to the data center and launching
6 a VM/container with the image, while in [108] the placement cost is propor-
7 tional to the cost of transferring the image and booting that VNF on a physical
8 machine.

9
10 Regarding the operational cost of links, the works presented in [20] and
11 [50] considered more detailed aspects of cost related to bandwidth. In [20],
12 a multi cloud scenario is considered and the authors considered that links
13 inter and intra clouds have different costs. In [50], where a mobile scenario
14 is considered, the cost about the transmission also takes into account the
15 wireless transmission cost between users and small base stations, and wired
16 transmission cost among the small base stations.

17 Finally, two works also considered the costs of SFC migration. In [47], the
18 migration cost is defined as the total costs for migrating the already-deployed
19 VNFs from one server to another. In [35], SFC migration cost is composed
20 of the data transfer rate, the link resource usage and the data transfer delay.
21 After the migration, the authors also consider the cost of network configuration
22 that takes into account the network reconfiguration overhead on a server.

23 5.2.2 Delay

24
25 The network delay is a metric that deserves attention in SFC placement, spe-
26 cially in distributed environments [108]. In the SFC context, the delay is the
27 time that the flow takes to get from the source to the destination, passing
28 through all the VNFs from an SFC, considering the link and processing delay
29 [2][14]. Therefore, the overall delay can be considered as the sum of the pro-
30 cessing delay by each VNF that composes the SFC and the communication
31 delay of links. The VNF processing delay, which is higher than the dedicated
32 hardware due to the virtualization, may vary depending on the virtualization
33 type (traditional virtual machines or containers) and the hardware used [64].
34 The communication delay is associated with the different delays when the traf-
35 fic is routed through, for example, different data centers, since different links
36 may have different associated delays and resources [65].

37
38 A typical example of VNF distribution is that of a cloud computing sce-
39 nario, where the VNFs of a SFC are deployed in different racks or even in
40 different geographically distributed data centers. This geographical distribu-
41 tion combined with unpredictable delay variations due to sharing of physical
42 infrastructure among different services, can result in violation of delay require-
43 ments, resulting in penalty costs [91]. Therefore, minimizing SFC delay is an
44 important concern for network operators.

45
46 All works presented in this subsection considered the SFC delay as the opti-
47 mization goal. However, other links aspects are also considered as constraints,
48 such as the bandwidth [113, 67, 119, 99].
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Five articles have considered SFC placement problem in geographically dis-
2 tributed cloud or multi domains while taking delay as their main constraint.
3 The optimization of end-users delay with optimal placement of SFCs was ad-
4 dressed by Bhamare et al. [10]. The multi-cloud scenario considers some con-
5 straints like total deployment cost and SLA requirements. An ILP formulation
6 was proposed and a polynomial heuristic “affinity-based allocation” was pre-
7 sented to cope with the complexity of ILP in larger scenarios. Bhamare et
8 al. also proposed a solution in [8] for the placement of SFCs in multi-cloud
9 environment for CRANs, where the goal was to minimize the overall response
10 time in a multi-cloud RAN. [The placement problem was modeled as a combi-
11 natorial optimization problem and two different approaches were proposed for
12 large networks: branch-and-bound and simulated annealing. In addition, some
13 enhancements to the standard branch-and-bound algorithm were presented in
14 the solution proposed by the authors.](#)

15
16 Martín-Pérez et al. [67] solved the SFC placement problem in a multi do-
17 main scenario, considering federated clouds. They proposed six algorithms to
18 minimize the end-to-end mapping delay. All algorithms are variations of a
19 greedy approach and search the closest servers to deploy the VNFs meeting
20 computing constraints. Depth first search and breadth first search algorithms
21 were used to boost up the mapping performance in terms of running time, while
22 using Tabu search was used to minimize the delay. In [113], Xu et al. addressed
23 the low latency multi-domain security service chain embedding problem, but
24 for mobile edge computing in 5G networks. The problem of cross-domain ser-
25 vice chain embedding is based on two sub-processes: service chain partition
26 and service subchain mapping. The authors formulated this problem as two
27 ILP problems in order to obtain an optimal solution in small networks. Then,
28 a heuristic solution was proposed to improve the classical Viterbi algorithm
29 [31] to obtain a near-optimal solution for large networks.

30
31 Still considering 5G networks, in [99], Subramanya et al. proposed a ma-
32 chine learning based solution to predict the number of VNF instances needed
33 to meet the traffic demand in 5G networks. Similar to [113], an ILP formu-
34 lation was proposed for the latency-optimal SFC placement problem, but the
35 adopted solution was different: a neural network was used to predict the num-
36 ber of VNFs taking as input measurements from the network.

37
38 Sun et al. [101] addressed the SFC placement in edge networks, similarly
39 to [113]. The authors proposed a solution to solve the SFC request orches-
40 tration considering a SFC request as a workflow based service request while
41 seeking to reduce the overall delay in edge computing. Thus, they sought to
42 find a network service placement schema with minimal delay. Two algorithms
43 were presented to solve the problem: a heuristic to map workflow-like request
44 minimizing response time; and another algorithm, similar to the previous one,
45 but it prohibits the VNF in the same network layer from being deployed on
46 the same substrate node.

47
48 Two articles addressed the SFC delay optimization taking into account
49 fog computing networks. The work proposed in [85] presented a controller to
50 allocate container-based SFCs in fog computing environments and focused on
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 a prototype solution. The controller was implemented as an extension of the
2 Kubernetes platform⁴. The default container scheduler of Kubernetes calls the
3 extension proposed when a scheduling decision is needed in order to minimize
4 the “latency established by the calculation of the shortest paths or based on
5 the target location for the pod deployment” [85]. If the latency-aware schema
6 is selected, the SFC controller best candidate selects a node based on the
7 calculation of Dijkstra’s shortest path algorithm. When the location-aware
8 scheme is selected, the selected node is chosen based on latency minimization
9 depending on the target location defined by the network manager.
10

11 Zamani et al. [119] proposed a solution for SFC placement considering
12 the fog-to-cloud scenario for use in IoT networks. The main objective was to
13 minimize the overall latency taking bandwidth consumption of IoT devices as
14 constraints. The SFC placement was formulated as an ILP problem but the
15 authors failed to state which algorithm they would use to solve the formulated
16 model. According to their approach, the functions can be deployed in either
17 fog or cloud nodes. The SFC mapping was divided into two steps: firstly, the
18 possible configurations to deploy the functions are selected satisfying various
19 constraints on nodes and links; secondly the route from the source to destina-
20 tion is defined satisfying various links’ constraints.
21

22 Ren et al. [80] also considered the SFC placement problem in an IoT sce-
23 nario and the main goal was to process IoT data as early as possible, then,
24 the VFN should be deployed in nodes as close to the data source as possi-
25 ble, reducing the overall network delay. Therefore, the authors formulated an
26 MILP problem to minimize the actual network node chain (ANNC), which is
27 the distance from the VNFs of a SFC to its data source, taking into account
28 the service requirements and resources limitations. A composition mechanism
29 was proposed for SFCs placement in IoT scenarios, and a variation of a genetic
30 algorithm was used to minimize the ANNC.
31

32 Cai et al. in [14] proposed an algorithm for parallel placement of SFCs in
33 distributed networks. The objective was to minimize the average SFC delay
34 over all of the parallelized SFCs taking into account servers and link con-
35 straints. An algorithm was proposed to convert a serial SFC deployment into
36 a parallel one, since some VNFs could run in parallel. Next, another algo-
37 rithm was proposed to allocate the VNFs (both parallelized and serial VNFs)
38 into servers, considering the total delay, VNF resource and bandwidth require-
39 ments.
40

41 Considering the works that addressed the delay minimization, the common
42 approach is to consider the delay about the links chosen to place the SFCs,
43 i.e., to forward the traffic among the VNFs. However, there are different de-
44 lays that can be considered in the end-to-end communication of the SFC. For
45 instance, as it was considered in [101], [10], and [8], the overall SFC delay
46 is divided into two categories: transmission delays associated with links and
47 computational delays associated with the clouds. The transmission delay of the
48 scenarios considered in the works can be very high and potentially have a big
49

4 <https://kubernetes.io/>

50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 impact on the end-to-end SFC delay: [10] considered a multi-cloud scenario,
2 [101] considered an IoT scenario, and [8] considered a CRAN architecture for
3 mobile networks. The transmission delay in centralized scenarios is too slow
4 and has a smaller impact than distributed scenarios. The consideration of the
5 computational delay is also important because the virtualization can increase
6 the VNF delay in comparison to the dedicated hardware. However, the VNF
7 processing delay is still neglected by most of the works that tried to minimize
8 the SFC delay.
9

10 Similar to [10] and [8], the work presented in [99] also considered the processing
11 time as required for all VNFs in an SFC to apply a specific network
12 operation on the arriving packets. However, they considered several delay types
13 in the scenario, such as the user equipment processing delay, over-the-air transmission
14 delay, hybrid automatic repeat request (HARQ) retransmission delay, and the transmission
15 delay between the base stations and the MEC layers. Therefore, it is important to consider
16 the particular delays that can emerge depending on the scenario evaluated, since they can
17 impact the overall SFC delay.
18

19 5.2.3 Resource utilization

20
21
22 Optimal resource utilization is another concern for network operators when
23 making SFC placement decisions. The optimal resource utilization can be
24 translated into minimum cost and maximum profit by efficiently utilizing computing
25 and network resources such as CPU, memory, and link bandwidth, at
26 the same time improving the network performance, availability and reliability
27 [114].
28

29 During the course of this systematic review, we discovered two articles that
30 addressed resource utilization in edge computing scenarios. Yang et al. [116]
31 studied how to place SFCs in edge and public cloud environments considering a
32 delay-aware VNF placement and routing problem. The main goal was to minimize
33 the maximum link load and at the same time satisfying user's requested
34 delay. The problem was formulated as an integer nonlinear programming and
35 transformed in an ILP. The authors developed a polynomial-time approximation
36 algorithm called Randomized Rounding VNF placement and routing
37 algorithm to solve this problem. Li et al. [58] also considered the SFC placement
38 in edge computing networks taking into account the link resource usage
39 minimization. In the context of their work, a VNF can be placed across the
40 network through duplicated copies in order to reduce bandwidth consumption.
41 An ILP formulation was developed. It considered the mapping of a series of
42 SFC requests to data centers and edge infrastructures. Next, a priority based
43 greedy heuristic was proposed to map SFC requests to data centers with an
44 adjustment process to merge the VNF requests demanding for the same type
45 of VNF.
46

47 Two articles found in this systematic review considered the resource utilization
48 minimization in PoP. The model proposed by Luzielli et al. [64] aimed
49 to minimize the number of virtual network functions instances mapped on
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 a PoP infrastructure. They considered three phases: placement, assignment,
2 and chaining. Then, an ILP model was proposed considering virtual network
3 functions to be placed in PoPs and SFC requests. To solve the modeled prob-
4 lem, the authors developed a heuristic that performs a binary search to adjust
5 the number of network functions to be instantiated dynamically (reducing the
6 resource consumption). The work presented in [88] also considered PoP in-
7 frastructures, but focused on SFC availability. The authors highlighted that
8 the link congestion and failures can impact SFC availability. Therefore, they
9 proposed an optimization model to minimize maximum load on links while
10 deploying SFCs. Khebbache et al. [53] did not address SFC placement in PoP
11 infrastructures specifically, but they considered the SFC placement in large
12 scale network infrastructures. It is a relevant problem due to the complexity
13 to find good solutions in an acceptable time. Thus, the authors presented a
14 VNF-forwarding graph chain placement problem that finds subsets of cycles in
15 the physical graph that can host the SFCs and meet their flow requirements.
16 Due to the nature of the problem, the authors associated it to the well known
17 perfect 2-matching problem which consists in finding cycles in the substrate
18 graph with a minimum amount of resources, verifying the resource limitations
19 and constraints.

21 We found two articles that addressed resource minimization in SFC place-
22 ment considering telecom infrastructures. In [17], Carpio et al. proposed a
23 solution for SFC placement in an Evolved Packet Core (EPC) network, which
24 is the core of 4G networks. The main goal was to minimize the link and
25 server utilization considering load balancing with constraints on the available
26 resources in mobile core networks. An ILP formulation was proposed by the
27 paper but no algorithm was suggested to solve it. The work proposed by Li
28 et al. [59] presented a solution to optimize either CPU usage or bandwidth
29 resource utilization in 5G networks. They also proposed an ILP formulation
30 for the optimization problem, but used reinforcement learning to solve it. The
31 ILP formulation can be adapted according to the objective of the network
32 operator: either to minimize bandwidth consumption or to minimize the CPU
33 consumption. In order to mitigate the ILP complexity, an adaptive deep Q-
34 learning based SFC mapping approach was proposed. The reinforcement learn-
35 ing agent learns how to make decisions based on two low-complexity heuristic
36 SFC mapping algorithms.

38 Similar to the work presented in [59], Song et al. [94] presented a solution to
39 minimize the number of CPU cores and the communication resources, in terms
40 of link bandwidth; but in generic distributed networks. An ILP formulation was
41 proposed to minimize the objective function considering several constraints in
42 terms of computing resources capacity, link capacity, and order of VNFs within
43 a SFC. A heuristic algorithm based on Hidden Markov Model was proposed
44 for SFC placement.

46 Most works that addressed the resources utilization optimization focused
47 on links resources, but with different concerning. The work presented by Shang
48 et al. [88] focused on to minimize the maximum link load, while other focused
49 on to minimize the bandwidth consumption [58, 53, 17, 59, 94]. Although the
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 objective is similar (minimizing the resource consumption of the links), these
2 different approaches result in different mathematical formulations. The work
3 proposed by Yang et al. [116] presents a different approach to minimize the
4 resource consumption of links: they considered the link load ratio. The main
5 goal is to avoid situations where one or more links are highly loaded, while
6 many links are less loaded.
7

8 Besides the concerning about the links resources consumption, the server
9 resources consumption minimization is also considered for some works. For
10 instance, the main focus of the work present by Li et al. [58] was to mitigate
11 the CPU consumption and the memory consumption of the servers where the
12 VNFs are placed. Li et al. [59] tried to minimize the CPU remaining rate
13 of the servers. These works also minimized the bandwidth consumption as
14 mentioned previously. Different of previous works, the works presented in [64]
15 and [17] tried to minimize the computational resources in a different way. In
16 [17], Carpio et al. tried to minimize the number of nodes where the VNFs will
17 be placed, in order to reduce the number of nodes with computing resources
18 consumed by the VNFs. In [64], a VNF replica can be placed in more than
19 one server, consuming more resources from the server, then the goal is to
20 minimize the number of replicas allocated in the infrastructure (reducing the
21 computational resources consumption).
22

23 24 *5.2.4 Energy Consumption*

25
26 Energy consumption reduction is becoming increasingly important due to the
27 explosive growth of data center traffic and power consumption that result
28 in increasing costs and carbon footprint [95]. Some studies indicate that the
29 adoption of the NFV paradigm can save up to 50% in energy consumption
30 as opposed to traditional network infrastructures [46]. However, to achieve a
31 considerable energy consumption reduction, the placement algorithms need to
32 consider this relevant metric in the SFC placement decisions [95].
33

34 The reduction of power consumption in multi-domain scenarios was ad-
35 dressed in two articles. In [55], Kouah et al. formulated an offline optimization
36 problem as a Mixed Integer Program (MIP) to minimize the SFC energy con-
37 sumption in IoT scenarios. The authors considered the SFC placement not
38 only for IoT equipment as well as in conjunction of a cloud infrastructure.
39 The problem was solved using a common solver (Gurobi [38]). A Genetic Al-
40 gorithm for reducing solution time was also proposed. The work presented in
41 [103] also addressed the energy consumption optimization for SFC placement
42 in multiple domains, more specifically, it considered a multiple cloud scenario.
43 Different from [55], the Sun et al. [103] formulated the problem as an ILP and
44 a low-complexity heuristic algorithm for SFC orchestration across multiple
45 domains sufficient for obtaining a near-optimal solution.
46

47 In [115] and [122], the authors focused on energy consumption optimization
48 of SFC placement across telecommunication networks. An ILP formulation was
49 proposed to minimize the SFC energy consumption while considering as con-
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 straints the links and servers capacity. A Markov-based algorithm was adopted
2 to find the near-optimal solution in a polynomial time.
3

4
5 The work proposed by Hejja and Hesselbach [44] presented an online algo-
6 rithm for SFC placement with the purpose of minimizing energy consumption.
7 They proposed an online algorithm that places the arriving SFC. To minimize
8 the energy consumption, the algorithm proposed monitors the servers in the
9 infrastructure, and after remove an SFC (due to expired lifetime) it turns off
10 all servers that do not host a VNF and do not receive traffic to process.
11

12
13 Evaluating the works that proposed solutions to optimize the energy con-
14 sumption, one can note that most of them considered the energy consumption
15 about the computing and communication. For instance, Sun et al. [103] con-
16 sidered that the computing energy consumption is generated by the servers
17 where the VNFs are hosted. A server consumes energy if it is on, even with-
18 out VNFs placed. If a server hosts VNFs, its energy consumption increases
19 according to the workload. On the other hand, the forwarding energy con-
20 sumption is related to the energy consumed by a server to forward data to
21 another servers. The forwarding consumption depends on the number of ports
22 used in the communication with other servers. The works presented by Xu et
23 al. [115] and Zhang et al. [122] defined that the energy consumption of SFCs in
24 a Telecom network scenario is composed of two main parts: energy consump-
25 tion of servers hosting VNF instances and the energy consumption of physical
26 links transmitting traffic. The server energy consumption consists of two parts:
27 the energy consumption to keep the server on, and the energy consumption
28 produced to process the VNF demands of SFC requests, which is related to
29 the CPU utilization. The energy consumption about the links “depends on its
30 on/off state and bandwidth utilization [115].”
31

32
33 However, two other works considered different energy consumption models.
34 The work presented by Hejja and Hesselbach [44] considered that the energy
35 consumption only about the servers that is composed of the sum server’s idle
36 power and the power consumption due to the traffic load, i.e., the utilization
37 of that server. Kouah et al. [55] considered an IoT scenario for the SFC place-
38 ment, and the energy consumption can be defined as a sum of three energy
39 consumption of three different aspects: communication, processing, and sens-
40 ing. The communication energy consumption takes into account the energy
41 consumption for transmit bits between two IoT nodes, which varies depending
42 on the distance between these nodes and the communication technology used
43 (WiFi, Bluetooth, Zigbee, Sigfox, LTE, and so on). The processing consump-
44 tion is about the energy consumed to processing a flow in a specific node,
45 and depends on the amount of the computational resource required by the
46 VNF. Finally, the sensing energy consumption is the consumed energy when
47 the sensing function of the IoT device is activated. The sensing function is
48 activated according to the application demand.
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

5.2.5 Revenue

Maximizing revenue is a common objective of network operators, but is hardly a trivial task. The revenue maximization is not an exclusive goal of network managers, which provides SFC services to their customers. Several works tried to optimize the revenue in cloud environments, which also is based on virtualization, i.e., launching virtual machines to host applications [83, 90, 5, 4]. However, the way in which revenue will be maximized may differ in an SFC scenario. One possibility may be to find the right trade-off between request acceptance and rejection rate of SFC requests [68].

A big challenge in SFC placement is to find the trade-off between admitting as many requests as possible while reducing the operational costs. For example, data centers in different locations have different energy consumption costs that impact the network operator revenue [66], thus choosing the most adequate data center for a SFC can dramatically impact revenue.

Two articles found in this systematic review addressed the revenue optimization in SFC placement: [112] and [61]. Both articles formulated the optimization problem as an ILP, but adopted different algorithms reflecting their different scenarios. Xie et al. [112] studied the problem of placing SFCs in a dynamic scenario, i.e., one with time-variant service flow sizes and VNF sequencing. The main goal was to maximize the revenue of the telecommunication service providers. Two algorithms were proposed to solve the optimization problem: the first one called DynAmic vnF placemenT (DAFT), was designed to solve the formulated problem by adopting a primal-dual technique. The second one called Feasible DAFT (FDAFT) complies to capacity constraints when creating an auxiliary graph for solving the general problem. Li et al. [61] considered the SFC placement across multiple vendors, which tend to be very different due to vendor dissimilarities. Due to the variability of scenarios, the authors proposed a reinforcement learning-based algorithm for the revenue maximization taking into account service benefit, resource cost, and service quality of candidate SFC paths.

Evaluating these two works, one can see that the revenue is calculated based on the the gains of meet an SFC request and possible penalties for not meet the SFC request or for not adequately meeting. For Xie et al. [112], the total revenue is calculated based on the revenue of serving an SFC request and the penalty of not serving an SFC request. The goal is to maximize the total revenue of a Telecom service provider considering several SFC requests over the time. On the other hand, for Li et al. [61] the revenue is calculated based on the service benefit, the service cost, and the service quality of the deployed SFC. These parameters are calculated based on the path (set of links) selected to place the SFC.

5.2.6 Other problems

In this subsection, we present articles that addressed different problems from those illustrated in the previous subsections. We identified three articles that fell in this category.

The work presented in [60] proposed a SFC placement solution for multi-domain scenarios. The proposed solution offers a heuristic based algorithm for SFC placement in multi-domain scenarios taking into account context information. This algorithm aims to mitigate the bandwidth over-provisioning problem and to optimize the revenue-cost ratio (from CPUs and bandwidth). A greedy based algorithm is proposed to allocate the SFC, considering the computational resources and link constraints.

In [15], a solution to maximize the acceptance rate of SFC requests was proposed considering as constraints the satisfaction of subscribers preferences. An ILP formulation for the optimization problem was developed by the authors as well as an algorithm to solve it in an offline and online modes.

The problem addressed in [21] considered users roaming over a cellular network (from cell to cell) and a SFC (for an undefined service) attending each respective user. The VNFs of these SFCs are allocated on MECs directly linked to the cellular network and linked between them. The main goal was to minimize the service interruption (downtime) for users. The SFC placement problem was formulated as an ILP and a heuristic was proposed to solve it. The heuristic can be divided into two parts: placement and migration. In the placement phase, the SFCs were deployed in servers considering delay constraints. In SFC migration, upon a user's handoff, the VNFs of a SFC were migrated among the servers in a way that avoids service interruption.

5.2.7 Multi-objective

Other articles found in this systematic review proposed problem formulations with two or more optimization objectives. These multi-objective optimization problems are more complex than the mono-objective ones because in most of the cases there conflicts among such objectives and a balance must be drawn [26].

Bouten et al. [13] proposed a solution for SFC placement considering affinity and anti-affinity constraints. Since the service providers are free to define their SFC constraints, it is possible that conflicting constraints are introduced. Then, the SFC requests are evaluated using a semantic approach in order to check their validity. Multiple optimization objectives are presented: maximizing revenue and the acceptance ratio; minimizing the number of substrate nodes; and minimizing the total bandwidth consumption. A heuristic based approach was proposed to solve the optimization problem.

Gupta et al. in [36] considered a problem to dynamically provide placement of the complete SFCs belonging to a virtual network service in multi-cloud systems. They proposed the P-ART framework which is an innovative predictive dynamic placement algorithm that considers changes in the state of the cloud

1 environment to ensure placement validity. The framework takes into account
2 cost and delay requirements from customers for the SFC placement.

3
4 A multi-domain network SFC placement problem was addressed in [120]. A
5 framework was proposed to jointly optimize the energy consumption of servers
6 and balance the load. Therefore, a heuristic algorithm was adopted to solve
7 the multi-objective optimization problem in a feasible time.

8 In [111], Troia et al. addressed the dynamic provision of slicing physical 5G
9 networks into logical networks in order to provide customized data transport
10 to each service. Then, ILP models were proposed to minimize “the probability
11 of blocking traffic requests, energy consumption of physical network devices
12 and interruption of service due to the reconfiguration of the slices”.

13 Harutyunyan et al. [41] also proposed a 5G slicing mechanism, that rep-
14 represents each slice as VFNs which compose a SFC. The VNFs of a SFC can
15 be allocated in hierarchical data centers with a different number of computing
16 resources. The main objective was to minimize the bandwidth consumption
17 in the transport network, the service provisioning cost, and the number of
18 VNF migrations. The placement problem was formulated as an MILP and
19 a heuristic was proposed to address the scalability problem. In [42], a SFC
20 placement solution for 5G networks was also proposed. It mainly considered
21 minimizing the delay of request services, service provisioning cost, and the
22 VFN migrations. Three different ILP formulations were presented to optimize
23 each objective function. They also provided an end-to-end delay model suit-
24 able for SFCs for 5G mobile networks. In order to avoid the scalability issue
25 of ILP formulation, a heuristic was used.

26
27 Jin et al. [49] also considered cellular networks, but the SFC was a general
28 signaling processing flow of virtual Mobility Management Entity (vMME) from
29 LTE (4G). Several cost functions were minimized in the proposed problem: the
30 total signaling communication overhead cost; signaling communication over-
31 head cost on backhauled; and migration overhead cost. Since the placement
32 problem is NP-hard, the authors proposed three different heuristic approaches
33 based on genetic algorithms to solve the problem.

34 The work presented in [89] proposed a solution to jointly optimize posi-
35 tioning and routing of SFCs to minimize the operational cost and network
36 congestion. The authors modeled the optimization problem as an ILP one and
37 proposed an approximation algorithm called Candidate Path Selection (CPS).
38 CPS solves the offline problem with a bounded approximation ratio. The au-
39 thors also proposed an online algorithm (OCPS) that deals with the lack of
40 future information and the rapidly fluctuating demand.

41 Eramo et al [27] proposed a solution to minimize operation costs (revenue
42 loss and energy consumption). This problem was broken down into two oth-
43 ers: the first one allocates a set of SFCs seeking to maximize the amount of
44 accepted bandwidth considering peak hour bandwidth demand, i.e., SFCs de-
45 manding more bandwidth are prioritized over SFCs demanding less bandwidth;
46 and the second one tries to consolidate VNFs on servers to minimize energy
47 consumption costs during the hours when traffic demand is lower, taking into
48 account also the revenue loss due SFC reconfiguration (migration imposes QoS
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

degradation to SFCs). Three heuristics were proposed: one for the first problem and two for the second one. The first heuristic follows a greedy approach based on link and node stress. The second one greedily migrates VNFs from servers with low power consumption and high bandwidth demand to servers with high power consumption and low bandwidth usage. Finally, the third algorithm uses a Discrete Time Markov Decision Process to determine the policy to be applied at each hour interval to reduce the total operational cost.

The work presented in [92] proposed a Tabu search scheme to deploy SFC in fog devices taking into account energy and resource constraints in a fog/cloud architecture. The main objective was to minimize the SFC delay and to improve load balancing (distribution) across network nodes. The authors proposed a solution based on both a heuristic and a metaheuristic. The Tabu search was used to define the SFC placement, while graph-based shortest-path heuristics were used to define the connections of nodes where the VNFs were deployed.

Tajiki et al. [106] addressed the problem of SFC placement in fog networks considering the trade-off between reliability, QoS, and energy consumption. To solve the corresponding problem, they mathematically formulated an optimization problem called the Optimal Fog-Supported Energy-Aware SFC rerouting algorithm (OFES) and proposed a near-optimal heuristic called Heuristic OFES (HFES).

5.3 Scenarios

The primary studies selected in this systematic review considered different application scenarios for the SFC placement, and mainly concentrated on networks of Internet Service Providers (ISP), 5G environments, smart city, Mobile Edge Networks (MEC), fog computing, hybrid networks, and Content Delivery Networks (CDN). Some works do not consider a specific scenario or application to evaluate the proposed solution. The works consider either randomly generated topologies, topologies defined in simulators, or prototypes, but without specifying a specific scenario. We classify these works as “other”. Figure 8 shows the number of articles for the different scenarios addressed.

Most of the papers (35 out of 60) considered different network topologies based on the different worldwide ISPs [64, 110, 88, 34, 62, 53, 54, 14, 13, 29, 112, 58, 101, 73, 15, 116, 60, 109, 78, 30, 116, 89, 94, 17, 106, 122, 115, 65, 37, 20, 36, 117, 10, 120, 59]. In this context, we can highlight applications considering the network topologies of China’s Telecom [14], USNET [94, 115, 122, 29], NSFNET [94, 115, 122, 73, 116, 54, 112, 29], The German backbone network [110, 15], Pan-European [15], Chinese CERNET2 [120], the ATT north America network [59], and Google data centers [65], which also acts as an ISP ⁵ have also been the scene for SFCs. However, the ISP network can be generated through simulation software. For instance, in [10], bhamare et al. used the

⁵ <https://fiber.google.com/>

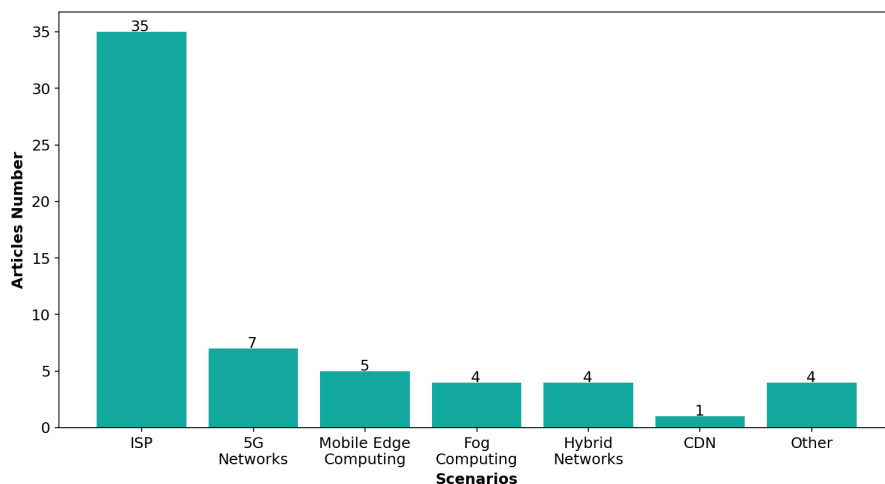


Fig. 8 Articles per scenario

GUESS software to generate the network topology used in the experiments. The topology consists of user clusters, where a user cluster is an ISP network with a 3-tier hierarchy of routers: access, aggregation, and core routers. The choice for this type of scenario reflects the authors' need to evaluate solutions considering a large numbers of geographically distributed nodes.

The second most popular scenario, present in seven papers (or 11.6%), was 5G networks [67, 107, 8, 103, 44, 41, 99]. These articles were motivated by the need to meet two main 5G networks' requirements: low energy consumption and low latency.

The smart city scenario was considered in [55], [49], [85], and [111]. This type of scenario has interesting dynamic characteristics due to the high mobility of users, which were presented as one of the major challenges in the SFC placement [49]. Specifically, this work [49] considered the SFC placement to meet the scenarios of vehicular networks and shared bicycles. Besides, authors seek to meet other scenarios, such as smart home and smart metering. A surveillance camera solution was presented in [85], considering face recognition services distributed over different nodes. The studies in [55] and [111] chose the application of solutions based on generic topologies, with nodes representing IoT gateways.

We also identified four papers that proposed solutions applied to Mobile Edge Computing (MEC) scenarios [113, 35, 21, 42]. One of the main characteristics of scenarios based on the MEC's concept is the ability to bring computing and storage resources close to the edge of the network, to the end-users, by pairing small servers with base stations. Generally, these servers have limited computing power and storage space, requiring that the service function allocation process be optimized to meet demand and fulfill such restrictions. In this sense, [35] addressed the process of minimizing the resources used in

1 this scenario to meet a set of restrictions. In [113, 21, 42], authors considered
2 reducing latency in the SFC placement process.
3

4 Three papers considered scenarios based on fog computing [80, 50, 92, 76,
5 119]. Unlike MEC-based scenarios in which processing is undertaken by the
6 edge server that receives requests, the concept of fog computing allows data
7 to be processed within a fog node or an IoT gateway placed within the user's
8 LAN. Such node provides higher computing power and acts as a centralized
9 unit at the edge of the network, serving a set of small distributed servers. In
10 this scenario, [80, 50] have as their main motivation to optimize the use of
11 resources in the face of Fog's limited computing power. In [92], the objective
12 was to minimize delays in the process of SFC placement. The paper presented
13 by Zamani and Sharifian [119] considered the SFC placement in a fog-to-
14 cloud scenario with the purpose to minimize the overall latency and bandwidth
15 consumption. In Nguyen *et al.* [76], a SFC placement was presented in an IoT
16 edge-to-cloud scenario to minimize the operational cost of the scenario.
17

18 In [47], the proposed solution considers SFC placement in a Content De-
19 livery Networks (CDN) application. A characteristic of this scenario is the
20 delivery of content considering cache servers at the edge of the network with
21 different content update policies. One of the main challenges in this scenario
22 was to minimize the SFC costs while the QoS of all service requests are jointly
23 satisfied, given an exponential increase in content consumption. Hence, the
24 solution presented in [47] aimed to minimize operating costs in this type of
25 scenario.

26 Four other articles considered generic scenarios for evaluating their pro-
27 posed solutions [108, 61, 27, 93]. Although they did not present details of the
28 scenarios considered in the articles, they all consider that the nodes were geo-
29 graphically distributed. Tashtarian *et al.* [108] considered a network topology
30 composed of six physical machines and nine links (a figure is presented in the
31 paper to illustrate the topology). However, no more details about the topology
32 generation were provided. In [27], the network topology is composed of five
33 core nodes, five edge nodes, and six access nodes. No more details about the
34 network topology generation were provided. The CloudSimSDN was used in
35 [93] for the generation of the network topology used in the evaluation. The au-
36 thors consider a network composed of multiple distributed cloud data centers,
37 but do not follow a specific scenario or real topology. Different from previous
38 works, Li *et al.* [61] considered both simulated and prototype topologies in
39 their evaluation. The simulated topology is composed of six nodes (not fol-
40 lowing a specific topology), while the prototype is composed of five connected
41 containers, that represent the servers in the network.
42
43
44
45

46 5.4 Metrics

47

48 The articles found in this systematic review evaluated their results based on
49 different metrics to achieve the defined objectives, such as minimizing energy
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 consumption, maximizing throughput, among others. Table 1 shows the met-
2 metrics used by each primary study of our systematic review.

3 We observed that energy consumption was considered as a restriction in ten
4 studies. Another twenty-six studies considered the quality of service metrics to
5 evaluate the solutions. Among these, eighteen studies considered delay, seven
6 considered available bandwidth, and only one considered network congestion
7 as evaluation metric.

8 One of the most popular metrics among the articles was the solution's
9 operational cost, being considered by twenty-five out of sixty papers found.
10 Another three articles considered revenue as an evaluation metric.

11 Metrics related to the use of resources were also considered, totaling twenty-
12 one articles. Here the use of CPU, memory, and network links were observed
13 as constraints to achieve the defined objectives. In the same vein, eight studies
14 analyzed the number of nodes used by the proposed solution and the number
15 of instances of service functions necessary to meet the demand. Another three
16 studies analyzed load balancing.

17 The reliability of the services to be allocated was also taken into account by
18 twenty-seven studies, in which four of them considered availability, twenty-one
19 considered acceptance rate, and two the utilization rate. Another three articles
20 presented the evaluation of throughput as a metric to measure the proposed
21 solution's scalability.

22 Finally, twenty-four articles evaluated the average execution and conver-
23 gence times of the algorithms presented in the SFC placement process.

24 **6 Discussion**

25 Although the NFV paradigm was proposed by ETSI as early as 2012 [22], one
26 can see that it is still a relevant research topic, mainly considering the man-
27 agement of SFCs. Most works found in this systematic review are very recent,
28 being published mostly in 2020, 2019, and 2018. The high number of recently
29 published papers only reinforces the importance of the NFV paradigm in the
30 development of new network services and architectures. Indeed, new network
31 based applications arose in the last years, which demand specific network func-
32 tions such as IoT, vehicular networks, e-health, smart city applications, among
33 others. In addition, with the growth of smartphone popularity, the softwariza-
34 tion of the cellular networks, more precisely in the next generation (5G), is
35 a trend to develop new network architectures and functionalities for mobile
36 systems [23].

37 Observing the scenarios considered in the works analyzed in this systematic
38 review, traditional ISP networks were the most frequent distributed scenario
39 evaluated in the literature. It is justifiable because ISP infrastructures are suit-
40 able for software-based solutions, since ISP providers are already familiar with
41 the cloud computing paradigm [98]. [Currently, the Internet Service Provider
42 Partners \(ISPP\) infrastructure is already composed of distributed data centers
43 supported by SDN switches and commodity servers \[45\].](#)

Table 1 Metrics used in the evaluation of solutions proposed by recent articles.

	Metrics		Reference	Description
Energy	power consumption	con-	[44], [111], [92], [55], [103], [120], [106], [122], [115], [27]	The energy consumption of the proposed solution was used as an evaluation metric.
QoS	delay		[76], [80], [35], [85], [37], [92], [119], [99], [101], [20], [36], [15], [10], [8], [78], [30], [116], [113]	Delay, bandwidth, and network congestion were metrics used by the studies to assess the quality of services, defining them as constraints in evaluating the proposed solutions.
	bandwidth		[80], [14], [85], [58], [116], [122], [115]	
	congestion		[89]	
Cost	operational cost		[76], [107], [34], [62], [53], [54], [65], [35], [50], [29], [123], [20], [36], [21], [44], [8], [120], [117], [60], [47], [109], [89], [94], [49], [27]	Costs involving some solutions are also considered. Here, the operating cost considers the cost of infrastructure equipment necessary to meet demand and budget constraints and maximizing expected revenue.
	revenue		[53], [61], [60]	
Resource utilization	resource utilization	uti-	[108], [64], [88], [76], [14], [13], [35], [50], [99], [58], [20], [73], [116], [8], [117], [59], [41], [17], [106], [27], [42]	The use of available resources was one of the most evaluated metrics among the studies found. Here the use of CPU, memory, and link are observed as restrictions to achieve the defined objectives, in addition to the number of nodes distributed geographically, and the number of VNFs instantiated to meet the demand.
	number of nodes or instances	of VNF	[58], [116], [122], [115], [34], [50], [112], [78]	
Load balancing	load balancing		[13], [112], [120]	Load balancing restrictions were also presented as evaluation metrics in articles found in the literature.
Reliability	availability		[88], [55], [123], [106]	The studies also considered metrics that measure the reliability of systems. The constraints defined by the algorithms were defined based on the availability, acceptance rate, and the services' utilization rate.
	acceptation rate		[112], [61], [34], [62], [53], [14], [37], [29], [36], [73], [15], [116], [21], [67], [60], [41], [78], [116], [113], [42], [93]	
	utilization rate		[112], [78]	
Scalability	throughput		[62], [78], [30]	Some studies presented the evaluation of throughput as a metric to measure the scalability of the proposed solution.
Algorithm	execution time		[108], [64], [34], [54], [13], [85], [37], [29], [99], [55], [123], [15], [117], [59], [41], [109], [78], [30], [116], [113], [27], [42]	The average execution times and convergence times of the algorithms in the SFC placement process were used as metrics to evaluate the proposed solutions.
	convergence time		[76], [53]	

1 We also found works that presented SFC placement solutions for emer-
2 gent technologies such as IoT, fog computing, smart city, CDN, and MEC.
3 In addition to being highly distributed, some of these scenarios are composed
4 of devices with limited computational resources that hinder the SFC place-
5 ment, such as fog computing, MEC, and IoT. These scenarios have strict re-
6 quirements which must be ensured by the network operator. For instance,
7 IoT devices have energy constraints and limited computational resources [106]
8 [50], 5G networks enforce requirements about latency, users mobility, connec-
9 tion density, peak data rate, etc [118]. Therefore, the SFC placement solutions
10 must consider all these constraints and specificity scenarios in order to propose
11 strategies to efficiently orchestrate SFCs.
12

13 Regarding the solutions, most works presented a mathematical formulation
14 based on ILP (or other mathematical programming formulations) for solving
15 the SFC placement problem. Using ILP formulations, the authors assumed
16 that the objective functions have a linear behavior and for optimizing these
17 models. They used traditional ILP solvers or proposed heuristics to deal with
18 bigger scenarios. Some recent work used machine learning strategies also, such
19 as deep learning and reinforcement learning [65, 35, 61, 59]. In case of deep
20 learning, the models can learn intricate structures in high-dimensional data,
21 and can be used to make predictions based on new data [57]. Using reinforc-
22 ement learning, an agent can be created to learn how to map situations to
23 actions in order to maximize a numerical reward signal [105]. Therefore, these
24 approaches can be used to create models for different types of SFC place-
25 ment problems, in different network scenarios, for mono and multi-objective
26 optimization.
27

28 Despite the existence of many works in the literature, there are still several
29 open research challenges that deserve our attention. For instance, [116] high-
30 light the need for implementing their solutions in a real testbed, and from our
31 systematic review, one could note that there is a current gap once the most
32 works have evaluated their proposals by using simulators. The design and im-
33 plementation of a testbed is a time demanding task and requires deep technical
34 knowledge, but it brings more realist results, highlighting the challenges and
35 providing experience to implement the solution in a real scenario.
36

37 Several technologies have emerged to implement SFC management system
38 prototypes, such as Open Source MANO⁶, OpenStack Tacker⁷, ONAP⁸, Ku-
39 bernetes, among others. However, there are emerging technologies that can be
40 used in future implementations. An example is the Programming Protocol-
41 Independent Packet Processors (P4), which allows to describe the pipeline of
42 packets in the network, including the parsing, match-action, and deparsing
43 [40]. The latest version of P4 language [24] is independent of the platform
44 where it will be executed, both of which can evolve independently, and fa-
45 cilitating the operability of the P4 language. In other words, the P4 allows
46

47 ⁶ <https://osm.etsi.org/>

48 ⁷ <https://wiki.openstack.org/wiki/Tacker>

49 ⁸ <https://www.onap.org/>

50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 programming the network behavior for different programmable devices. These
2 aspects are relevant for SFC implementations and prototypes to control and
3 adapt the traffic among the VNFs.
4

5 Taking into account the system model formulation considered in the works, it
6 is common to consider generic aspects regarding the computing and link re-
7 sources of the physical infrastructure. For instance, several works considered
8 the processing resources just as “CPU”. However, depending on the scenario,
9 several processor types can be available in the physical nodes. A good example
10 is the solution proposed by Jin et al. [50] that is focused on SFC placement for
11 mobile games. As defined by the authors, the online game VNFs can require
12 CPU and/or GPUs to execute the tasks, and these processor types have dif-
13 ferent costs. Therefore, this differentiation has an impact on the overall SFC
14 cost and is relevant to be considered.

15 Another example is about the delays considered in the SFC. The majority
16 of work considered only delays about the links and the processing delay by
17 VNFs. However, other delays can be present depending on the scenario. An
18 interesting example is the solution presented in [99] for SFC placement in
19 5G networks. As the main goal of the paper is to minimize the overall SFC
20 delay, several delays specific to mobile networks are considered such as user
21 equipment processing delay, over-the-air transmission delay, the transmission
22 about the base stations, among others. Whether only the delay about the
23 links were considered, a relevant part of the total delay of the SFC would be
24 disregarded in this scenario, thus, the presented solution could not be adapted
25 so well if it were tried in a real scenario.
26
27
28

29 **7 Conclusions**

30

31 This work presented a systematic review on the problem of SFC placement in
32 distributed scenarios. In this context, the various network restrictions imposed
33 by the increasing number of devices and the emergence of applications sensi-
34 tive to such restrictions (e.g., delay, bandwidth, and energy consumption) are
35 pointed out as the main challenges faced by operators. In highly distributed
36 scenarios, these challenges may even be greater in the face of network fluctu-
37 ations.
38

39 In this review, solutions were found whose objectives focus on optimizing
40 operational cost, delay, resource utilization, energy consumption, and revenue,
41 considering different distributed scenarios. Although most of these works con-
42 sider the optimization of only one of these metrics, other studies have formu-
43 lated their optimization problems considering two or more objectives at the
44 same time. As for the formulation of the problem, the most common optimiza-
45 tion framework among the studies is the integer programming technique.
46

47 With regard to problem formulation, most of the algorithms were designed
48 using some heuristic. Besides, dynamic programming and reinforcement learn-
49 ing were also investigated in many of the studies.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 To evaluate the proposed solutions, the authors considered different scenarios,
2 such as Fog Computing, Mobile Edge Computing, and Content Delivery
3 Networks. However, the most popular scenario among the studies was ISP
4 Networks. The scenario involved evaluations that took place considering different
5 metrics. The most commonly presented metric are the network delay
6 of the SFC allocated, the operational costs, and the use of resources for the
7 placement of the SFCs. The acceptance rate was also one of the most popular
8 metrics considered, allowing to measure reliability, in addition to the execution
9 time of the proposed solutions by the studies, which allowed to measure their
10 performance.
11
12
13
14
15

16 References

- 18 1. Alameddine, H.A., Assi, C., Tushar, M.H.K., Yu, J.Y.: Low-latency service
19 schedule orchestration in nfv-based networks. In: 2019 IEEE Conference
20 on Network Softwarization (NetSoft), pp. 378–386. IEEE (2019)
- 21 2. Alleg, A., Ahmed, T., Mosbah, M., Riggio, R., Boutaba, R.: Delay-aware
22 vnf placement and chaining based on a flexible resource allocation approach.
23 In: 2017 13th international conference on network and service
24 management (CNSM), pp. 1–7. iee (2017)
- 25 3. Asad, M., Basit, A., Qaisar, S., Ali, M.: Beyond 5g: Hybrid end-to-end
26 quality of service provisioning in heterogeneous iot networks. *IEEE Access* **8**,
27 192320–192338 (2020)
- 28 4. Badshah, A., Ghani, A., Shamshirband, S., Aceto, G., Pescapè, A.:
29 Performance-based service-level agreement in cloud computing to optimise
30 penalties and revenue. *IET Communications* **14**(7), 1102–1112
31 (2020)
- 32 5. Badshah, A., Ghani, A., Shamshirband, S., Chronopoulos, A.T.: Optimising
33 infrastructure as a service provider revenue through customer
34 satisfaction and efficient resource provisioning in cloud computing. *IET*
35 *Communications* **13**(18), 2913–2922 (2019)
- 36 6. Bari, F., Chowdhury, S.R., Ahmed, R., Boutaba, R., Duarte, O.C.M.B.:
37 Orchestrating virtualized network functions. *IEEE Transactions on Network*
38 *and Service Management* **13**(4), 725–739 (2016)
- 39 7. Barroso, L.A., Hölzle, U.: The datacenter as a computer: An introduction
40 to the design of warehouse-scale machines. *Synthesis lectures on computer*
41 *architecture* **4**(1), 1–108 (2009)
- 42 8. Bhamare, D., Erbad, A., Jain, R., Zolanvari, M., Samaka, M.: Efficient
43 virtual network function placement strategies for cloud radio access networks.
44 *Computer Communications* **127**, 50–60 (2018)
- 45 9. Bhamare, D., Jain, R., Samaka, M., Erbad, A.: A survey on service function
46 chaining. *Journal of Network and Computer Applications* **75**, 138–
47 155 (2016)

10. Bhamare, D., Samaka, M., Erbad, A., Jain, R., Gupta, L., Chan, H.A.: Optimal virtual network function placement in multi-cloud service function chaining architecture. *Computer Communications* **102**, 1–16 (2017)
11. Bonfim, M.S., Dias, K.L., Fernandes, S.F.: Integrated nfv/sdn architectures: A systematic literature review. *ACM Computing Surveys (CSUR)* **51**(6), 1–39 (2019)
12. Bouet, M., Leguay, J., Combe, T., Conan, V.: Cost-based placement of vdpi functions in nfv infrastructures. *International Journal of Network Management* **25**(6), 490–506 (2015)
13. Bouten, N., Mijumbi, R., Serrat, J., Famaey, J., Latré, S., De Turck, F.: Semantically enhanced mapping algorithm for affinity-constrained service function chain requests. *IEEE Transactions on Network and Service Management* **14**(2), 317–331 (2017)
14. Cai, J., Huang, Z., Luo, J., Liu, Y., Zhao, H., Liao, L.: Composing and deploying parallelized service function chains. *Journal of Network and Computer Applications* p. 102637 (2020)
15. Capanera, P., Paganelli, F., Paradiso, F.: Vnf placement for service chaining in a distributed cloud environment with multiple stakeholders. *Computer Communications* **133**, 24–40 (2019)
16. Cardoso, J., Barros, A., May, N., Kylau, U.: Towards a unified service description language for the internet of services: Requirements and first developments. In: 2010 IEEE International Conference on Services Computing, pp. 602–609. IEEE (2010)
17. Carpio, F., Bziuk, W., Jukan, A.: Replication of virtual network functions: Optimizing link utilization and resource costs. In: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 521–526. IEEE (2017)
18. Chamberland, S.: Point of presence design in internet protocol networks with performance guarantees. *Computers & operations research* **32**(12), 3247–3264 (2005)
19. Chamberland, S., St-Hilaire, M., Pierre, S.: On the point-of-presence optimization problem in ip networks. *Canadian Journal of Electrical and Computer Engineering* **30**(3), 137–143 (2005)
20. Chen, H., Wang, X., Zhao, Y., Song, T., Wang, Y., Xu, S., Li, L.: Mosc: A method to assign the outsourcing of service function chain across multiple clouds. *Computer Networks* **133**, 166–182 (2018)
21. Chen, Y.T., Liao, W.: Mobility-aware service function chaining in 5g wireless networks with mobile edge computing. In: ICC 2019-2019 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2019)
22. Chiosi, M., BT, C.D., PETER, W., CENTURYLINK, A.R., JAMES, F., MICHAEL, B., WAQAR, K., MICHAEL, F.: Network functions virtualization—introductory white paper issue 1 network functions virtualization an introduction, benefits, enablers, challenges & call for action contributing organisations & authors (2012)

23. Condoluci, M., Mahmoodi, T.: Softwarization and virtualization in 5g mobile networks: Benefits, trends and challenges. *Computer Networks* **146**, 65–84 (2018)
24. Consortium, P.L.: P4 language and related specifications. <https://p4.org/specs/> (2021)
25. Coutinho, E.F., de Carvalho Sousa, F.R., Rego, P.A.L., Gomes, D.G., de Souza, J.N.: Elasticity in cloud computing: a survey. *annals of telecommunications-Annales des télécommunications* **70**(7-8), 289–309 (2015)
26. Cui, Y., Geng, Z., Zhu, Q., Han, Y.: Multi-objective optimization methods and application in energy saving. *Energy* **125**, 681–704 (2017)
27. Eramo, V., Miucci, E., Ammar, M., Lavacca, F.G.: An approach for service function chain routing and virtual function network instance migration in network function virtualization architectures. *IEEE/ACM Transactions on Networking* **25**(4), 2008–2025 (2017)
28. Fan, J., Guan, C., Zhao, Y., Qiao, C.: Availability-aware mapping of service function chains. In: *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pp. 1–9. IEEE (2017)
29. Fang, L., Zhang, X., Sood, K., Wang, Y., Yu, S.: Reliability-aware virtual network function placement in carrier networks. *Journal of Network and Computer Applications* **154**, 102536 (2020)
30. Feng, H., Llorca, J., Tulino, A.M., Molisch, A.F.: Optimal dynamic cloud network control. *IEEE/ACM Transactions on Networking* **26**(5), 2118–2131 (2018)
31. Forney, G.D.: The viterbi algorithm. *Proceedings of the IEEE* **61**(3), 268–278 (1973)
32. Fulber-Garcia, V., Duarte Jr, E.P., Huff, A., dos Santos, C.R.: Network service topology: Formalization, taxonomy and the custom specification model. *Computer Networks* **178**, 107337 (2020)
33. Ghazouani, S., Slimani, Y.: Towards a standardized cloud service description based on usdl. *Journal of Systems and Software* **132**, 1–20 (2017)
34. Guo, H., Wang, Y., Li, Z., Qiu, X., An, H., Yuan, N., et al.: Cost-aware placement and chaining of service function chain with vnf instance sharing. In: *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*, pp. 1–8. IEEE (2020)
35. Guo, S., Dai, Y., Xu, S., Qiu, X., Qi, F.: Trusted cloud-edge network resource management: Drl-driven service function chain orchestration for iot. *IEEE Internet of Things Journal* (2019)
36. Gupta, L., Jain, R., Erbad, A., Bhamare, D.: The p-art framework for placement of virtual network services in a multi-cloud environment. *Computer Communications* **139**, 103–122 (2019)
37. Gupta, L., Samaka, M., Jain, R., Erbad, A., Bhamare, D., Metz, C.: Colap: A predictive framework for service function chain placement in a multi-cloud environment. In: *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1–9. IEEE (2017)

- 1 38. Gurobi Optimization, I.: Gurobi optimizer reference manual. URL
2 <http://www.gurobi.com> (2018)
- 3 39. Hantouti, H., Benamar, N., Taleb, T., Laghrissi, A.: Traffic steering for
4 service function chaining. *IEEE Communications Surveys & Tutorials*
5 **21**(1), 487–507 (2018)
- 6 40. Harkous, H., Jarschel, M., He, M., Priest, R., Kellerer, W.: Towards un-
7 derstanding the performance of p4 programmable hardware. In: 2019
8 ACM/IEEE Symposium on Architectures for Networking and Commu-
9 nications Systems (ANCS), pp. 1–6. IEEE (2019)
- 10 41. Harutyunyan, D., Fedrizzi, R., Shahriar, N., Boutaba, R., Riggio, R.: Or-
11 chestrating end-to-end slices in 5g networks. In: 2019 15th International
12 Conference on Network and Service Management (CNSM), pp. 1–9. IEEE
13 (2019)
- 14 42. Harutyunyan, D., Shahriar, N., Boutaba, R., Riggio, R.: Latency-aware
15 service function chain placement in 5g mobile networks. In: 2019 IEEE
16 Conference on Network Softwarization (NetSoft), pp. 133–141. IEEE
17 (2019)
- 18 43. He, W., Li, W., Xu, S.: A lyapunov drift-plus-penalty-based multi-
19 objective optimization of energy consumption, construction period and
20 benefit. *KSCE Journal of Civil Engineering* pp. 1–14 (2020)
- 21 44. Hejja, K., Hesselbach, X.: Offline and online power aware resource al-
22 location algorithms with migration and delay constraints. *Computer*
23 *Networks* **158**, 17–34 (2019)
- 24 45. Hong, D.K., Ma, Y., Banerjee, S., Mao, Z.M.: Incremental deployment
25 of sdn in hybrid enterprise and isp networks. In: Proceedings of the
26 Symposium on SDN Research, pp. 1–7 (2016)
- 27 46. ISG, N.: Network functions virtualisation (nfv)-virtualisation require-
28 ments. ETSI Technical Report (2013)
- 29 47. Jahromi, N.T., Kianpisheh, S., Glitho, R.H.: Online vnf placement and
30 chaining for value-added services in content delivery networks. In: 2018
31 IEEE International Symposium on Local and Metropolitan Area Net-
32 works (LANMAN), pp. 19–24. IEEE (2018)
- 33 48. Jiao, S., Zhang, X., Yu, S., Song, X., Xu, Z.: Joint virtual network func-
34 tion selection and traffic steering in telecom networks. In: GLOBECOM
35 2017-2017 IEEE Global Communications Conference, pp. 1–7. IEEE
36 (2017)
- 37 49. Jin, H., Jin, Y., Lu, H., Zhao, C., Peng, M.: Nfv and sfc: a case study of
38 optimization for virtual mobility management. *IEEE Journal on Selected*
39 *Areas in Communications* **36**(10), 2318–2332 (2018)
- 40 50. Jin, H., Zhu, X., Zhao, C.: Computation offloading optimization based
41 on probabilistic sfc for mobile online gaming in heterogeneous network.
42 *IEEE Access* **7**, 52168–52180 (2019)
- 43 51. John, W., Pentikousis, K., Agapiou, G., Jacob, E., Kind, M., Manzalini,
44 A., Risso, F., Staessens, D., Steinert, R., Meirosu, C.: Research directions
45 in network service chaining. In: 2013 IEEE SDN for future networks and
46 services (SDN4FNS), pp. 1–7. IEEE (2013)
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- 1 52. Katsaros, G., Menzel, M., Lenk, A., Revelant, J.R., Skipp, R., Eberhardt,
2 J.: Cloud application portability with toasca, chef and openstack. In: 2014
3 IEEE international conference on cloud engineering, pp. 295–302. IEEE
4 (2014)
- 5 53. Khebbache, S., Hadji, M., Zeghlache, D.: Virtualized network functions
6 chaining and routing algorithms. *Computer Networks* **114**, 95–110 (2017)
- 7 54. Kiji, N., Sato, T., Shinkuma, R., Oki, E.: Virtual network function place-
8 ment and routing for multicast service chaining using merged paths. *Op-
9 tical Switching and Networking* **36**, 100554 (2020)
- 10 55. Kouah, R., Alleg, A., Laraba, A., Ahmed, T.: Energy-aware placement
11 for iot-service function chain. In: 2018 IEEE 23rd International Workshop
12 on Computer Aided Modeling and Design of Communication Links and
13 Networks (CAMAD), pp. 1–7. IEEE (2018)
- 14 56. Kuo, T.W., Liou, B.H., Lin, K.C.J., Tsai, M.J.: Deploying chains of virtual
15 network functions: On the relation between link and server usage. *IEEE/ACM Transactions On Networking* **26**(4), 1562–1576 (2018)
- 16 57. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *nature* **521**(7553),
17 436–444 (2015)
- 18 58. Li, D., Hong, P., Xue, K., Pei, J.: Virtual network function placement
19 and resource optimization in nfv and edge computing enabled networks.
20 *Computer Networks* **152**, 12–24 (2019)
- 21 59. Li, G., Feng, B., Zhou, H., Zhang, Y., Sood, K., Yu, S.: Adaptive ser-
22 vice function chaining mappings in 5g using deep q-learning. *Computer
23 Communications* **152**, 305–315 (2020)
- 24 60. Li, G., Zhou, H., Feng, B., Li, G.: Context-aware service function chain-
25 ing and its cost-effective orchestration in multi-domain networks. *IEEE
26 Access* **6**, 34976–34991 (2018)
- 27 61. Li, G., Zhou, H., Feng, B., Zhang, Y., Yu, S.: Efficient provision of service
28 function chains in overlay networks using reinforcement learning. *IEEE
29 Transactions on Cloud Computing* (2019)
- 30 62. Liu, Y., Pei, J., Hong, P., Li, D.: Cost-efficient virtual network function
31 placement and traffic steering. In: ICC 2019-2019 IEEE International
32 Conference on Communications (ICC), pp. 1–6. IEEE (2019)
- 33 63. Liu, Y., Zhang, H., Chang, D., Hu, H.: Gdm: A general distributed
34 method for cross-domain service function chain embedding. *IEEE Trans-
35 actions on Network and Service Management* **17**(3), 1446–1459 (2020)
- 36 64. Luizelli, M.C., Bays, L.R., Buriol, L.S., Barcellos, M.P., Gasparly, L.P.:
37 Piecing together the nfv provisioning puzzle: Efficient placement and
38 chaining of virtual network functions. In: 2015 IFIP/IEEE International
39 Symposium on Integrated Network Management (IM), pp. 98–106. IEEE
40 (2015)
- 41 65. Luo, Z., Wu, C., Li, Z., Zhou, W.: Scaling geo-distributed network func-
42 tion chains: A prediction and learning framework. *IEEE Journal on Se-
43 lected Areas in Communications* **37**(8), 1838–1850 (2019)
- 44 66. Ma, Y., Liang, W., Xu, Z., Guo, S.: Profit maximization for admitting re-
45 quests with network function services in distributed clouds. *IEEE Trans-
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65*

- actions on Parallel and Distributed Systems **30**(5), 1143–1157 (2018)
67. Martín-Pérez, J., Bernardos, C.J.: Multi-domain vnf mapping algorithms. In: 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1–6. IEEE (2018)
68. Mechtri, M., Ghribi, C., Soualah, O., Zeghlache, D.: Nfv orchestration framework addressing sfc challenges. *IEEE Communications Magazine* **55**(6), 16–23 (2017)
69. Mehraghdam, S., Karl, H.: Placement of services with flexible structures specified by a yang data model. In: 2016 IEEE NetSoft Conference and Workshops (NetSoft), pp. 184–192. IEEE (2016)
70. Mijumbi, R., Serrat, J., Gorricho, J.L., Bouten, N., De Turck, F., Boutaba, R.: Network function virtualization: State-of-the-art and research challenges. *IEEE Communications surveys & tutorials* **18**(1), 236–262 (2015)
71. Mijumbi, R., Serrat, J., Gorricho, J.L., Bouten, N., De Turck, F., Davy, S.: Design and evaluation of algorithms for mapping and scheduling of virtual network functions. In: Proceedings of the 2015 1st IEEE conference on network softwarization (NetSoft), pp. 1–9. IEEE (2015)
72. Mirjalily, G., Zhiquan, L.: Optimal network function virtualization and service function chaining: A survey. *Chinese Journal of Electronics* **27**(4), 704–717 (2018)
73. Mohamad, A., Hassanein, H.S.: On demonstrating the gain of sfc placement with vnf sharing at the edge. In: 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE (2019)
74. Nan, Y., Li, W., Bao, W., Delicato, F.C., Pires, P.F., Zomaya, A.Y.: Cost-effective processing for delay-sensitive applications in cloud of things systems. In: 2016 IEEE 15th international symposium on network computing and applications (NCA), pp. 162–169. IEEE (2016)
75. Neely, M.J.: Queue stability and probability 1 convergence via lyapunov optimization. arXiv preprint arXiv:1008.3519 (2010)
76. Nguyen, D.T., Pham, C., Nguyen, K.K., Cheriet, M.: Placement and chaining for run-time iot service deployment in edge-cloud. *IEEE Transactions on Network and Service Management* **17**(1), 459–472 (2019)
77. Noghani, K.A., Kessler, A., Taheri, J.: On the cost-optimality trade-off for service function chain reconfiguration. In: 2019 IEEE 8th International Conference on Cloud Networking (CloudNet), pp. 1–6. IEEE (2019)
78. Pei, J., Hong, P., Xue, K., Li, D.: Efficiently embedding service function chains with dynamic virtual network function placement in geodistributed cloud system. *IEEE Transactions on Parallel and Distributed Systems* **30**(10), 2179–2192 (2018)
79. Qu, L., Assi, C., Shaban, K.: Delay-aware scheduling and resource optimization with network function virtualization. *IEEE Transactions on Communications* **64**(9), 3746–3758 (2016)
80. Ren, W., Sun, Y., Luo, H., Obaidat, M.S.: A new scheme for iot service function chains orchestration in sdn-iot network systems. *IEEE Systems*

- Journal **13**(4), 4081–4092 (2019)
81. Riera, J.F., Escalona, E., Batalle, J., Grasa, E., Garcia-Espin, J.A.: Virtual network function scheduling: Concept and challenges. In: 2014 international conference on smart communications in network technologies (SaCoNeT), pp. 1–5. IEEE (2014)
 82. Rotsos, C., King, D., Farshad, A., Bird, J., Fawcett, L., Georgalas, N., Gunkel, M., Shiimoto, K., Wang, A., Mauthe, A., et al.: Network service orchestration standardization: A technology survey. *Computer Standards & Interfaces* **54**, 203–215 (2017)
 83. Saha, S., Sarkar, J., Dwivedi, A., Dwivedi, N., Narasimhamurthy, A.M., Roy, R.: A novel revenue optimization model to address the operation and maintenance cost of a data center. *Journal of Cloud Computing* **5**(1), 1–23 (2016)
 84. Sahhaf, S., Tavernier, W., Czentye, J., Sonkoly, B., Sköldström, P., Jocha, D., Garay, J.: Scalable architecture for service function chain orchestration. In: 2015 Fourth European Workshop on Software Defined Networks, pp. 19–24. IEEE (2015)
 85. Santos, J., Wauters, T., Volckaert, B., De Turck, F.: Towards delay-aware container-based service function chaining in fog computing. In: NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium, pp. 1–9. IEEE (2020)
 86. Savi, M., Tornatore, M., Verticale, G.: Impact of processing-resource sharing on the placement of chained virtual network functions. *IEEE Transactions on Cloud Computing* (2019)
 87. Schardong, F., Nunes, I., Schaeffer-Filho, A.: Nfv resource allocation: A systematic review and taxonomy of vnf forwarding graph embedding. *Computer Networks* p. 107726 (2020)
 88. Shang, X., Li, Z., Yang, Y.: Placement of highly available virtual network functions through local rerouting. In: 2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), pp. 80–88. IEEE (2018)
 89. Shang, X., Liu, Z., Yang, Y.: Network congestion-aware online service function chain placement and load balancing. In: Proceedings of the 48th International Conference on Parallel Processing, pp. 1–10 (2019)
 90. Sheikholeslami, F., Navimipour, N.J.: Service allocation in the cloud environments using multi-objective particle swarm optimization algorithm based on crowding distance. *Swarm and Evolutionary Computation* **35**, 53–64 (2017)
 91. Sheoran, A., Sharma, P., Fahmy, S., Saxena, V.: Contain-ed: An nfv micro-service system for containing e2e latency. *ACM SIGCOMM Computer Communication Review* **47**(5), 54–60 (2017)
 92. Siasi, N., Jaesim, A., Ghani, N.: Tabu search for efficient service function chain provisioning in fog networks. In: 2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC), pp. 145–150. IEEE (2019)

- 1 93. Son, J., Buyya, R.: Latency-aware virtualized network function provi-
2 sioning for distributed edge clouds. *Journal of Systems and Software*
3 **152**, 24–31 (2019)
- 4 94. Song, X., Zhang, X., Yu, S., Jiao, S., Xu, Z.: Resource-efficient virtual
5 network function placement in operator networks. In: *GLOBECOM 2017-*
6 *2017 IEEE Global Communications Conference*, pp. 1–7. IEEE (2017)
- 7 95. Soualah, O., Mechtri, M., Ghribi, C., Zeghlache, D.: Energy efficient al-
8 gorithm for vnf placement and chaining. In: *2017 17th IEEE/ACM Inter-*
9 *national Symposium on Cluster, Cloud and Grid Computing (CCGRID)*,
10 pp. 579–588. IEEE (2017)
- 11 96. Souza, R., Dias, K., Fernandes, S.: Nfv data centers: A systematic review.
12 *IEEE Access* **8**, 51713–51735 (2020)
- 13 97. Souza, R., Santos, M., Fernandes, S.: Importance measures for nfv data
14 center: An availability evaluation. In: *Anais do V Workshop Pré-IETF.*
15 *SBC* (2018)
- 16 98. Stiawan, D.: Personal data protection and liability of internet service
17 provider a comparative. *International Journal of Electrical and Computer*
18 *Engineering (IJECE)* **9**(4), 3175–3184 (2019)
- 19 99. Subramanya, T., Harutyunyan, D., Riggio, R.: Machine learning-driven
20 service function chain placement and scaling in mec-enabled 5g networks.
21 *Computer Networks* **166**, 106980 (2020)
- 22 100. Sun, G., Chen, Z., Yu, H., Du, X., Guizani, M.: Online parallelized service
23 function chain orchestration in data center networks. *IEEE Access* **7**,
24 100147–100161 (2019)
- 25 101. Sun, G., Li, Y., Li, Y., Liao, D., Chang, V.: Low-latency orchestration
26 for workflow-oriented service function chain in edge computing. *Future*
27 *Generation Computer Systems* **85**, 116–128 (2018)
- 28 102. Sun, G., Li, Y., Liao, D., Chang, V.: Service function chain orchestration
29 across multiple domains: A full mesh aggregation approach. *IEEE Trans-*
30 *actions on Network and Service Management* **15**(3), 1175–1191 (2018)
- 31 103. Sun, G., Li, Y., Yu, H., Vasilakos, A.V., Du, X., Guizani, M.: Energy-
32 efficient and traffic-aware service function chaining orchestration in multi-
33 domain networks. *Future Generation Computer Systems* **91**, 347–360
34 (2019)
- 35 104. Sun, L., Dong, H., Ashraf, J.: Survey of service description languages and
36 their issues in cloud computing. In: *2012 Eighth International Conference*
37 *on Semantics, Knowledge and Grids*, pp. 128–135. IEEE (2012)
- 38 105. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT
39 press (2018)
- 40 106. Tajiki, M.M., Shojafar, M., Akbari, B., Salsano, S., Conti, M., Singhal,
41 M.: Joint failure recovery, fault prevention, and energy-efficient resource
42 management for real-time sfc in fog-supported sdn. *Computer Networks*
43 **162**, 106850 (2019)
- 44 107. Tang, L., Yang, H., Ma, R., Hu, L., Wang, W., Chen, Q.: Queue-aware
45 dynamic placement of virtual network functions in 5g access network.
46 *IEEE Access* **6**, 44291–44305 (2018)
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- 1 108. Tashtarian, F., Zhani, M.F., Fatemipour, B., Yazdani, D.: Codec: a cost-
2 effective and delay-aware sfc deployment. *IEEE Transactions on Network*
3 *and Service Management* (2019)
- 4 109. Tastevin, N., Obadia, M., Bouet, M.: A graph approach to placement of
5 service functions chains. In: 2017 IFIP/IEEE Symposium on Integrated
6 *Network and Service Management (IM)*, pp. 134–141. IEEE (2017)
- 7 110. Tomassilli, A., Giroire, F., Huin, N., Pérennes, S.: Provably efficient algo-
8 rithms for placement of service function chains with ordering constraints.
9 In: *IEEE INFOCOM 2018-IEEE Conference on Computer Communica-*
10 *tions*, pp. 774–782. IEEE (2018)
- 11 111. Troia, S., Cibari, A., Alvizu, R., Maier, G.: Dynamic programming of
12 network slices in software-defined metro-core optical networks. *Optical*
13 *Switching and Networking* **36**, 100551 (2020)
- 14 112. Xie, Y., Wang, S., Dai, Y.: Revenue-maximizing virtualized network func-
15 tion chain placement in dynamic environment. *Future Generation Com-*
16 *puter Systems* (2020)
- 17 113. Xu, Q., Gao, D., Li, T., Zhang, H.: Low latency security function chain
18 embedding across multiple domains. *IEEE Access* **6**, 14474–14484 (2018)
- 19 114. Xu, Y., Kafle, V.P.: A mathematical model and dynamic programming
20 based scheme for service function chain placement in nfv. *IEICE Trans-*
21 *actions on Information and Systems* **102**(5), 942–951 (2019)
- 22 115. Xu, Z., Zhang, X., Yu, S., Zhang, J.: Energy-efficient virtual network
23 function placement in telecom networks. In: 2018 IEEE International
24 *Conference on Communications (ICC)*, pp. 1–7. IEEE (2018)
- 25 116. Yang, S., Li, F., Trajanovski, S., Chen, X., Wang, Y., Fu, X.: Delay-
26 aware virtual network function placement and routing in edge clouds.
27 *IEEE Transactions on Mobile Computing* (2019)
- 28 117. Yao, Y., Guo, S., Li, P., Liu, G., Zeng, Y.: Forecasting assisted vnf scaling
29 in nfv-enabled networks. *Computer Networks* **168**, 107040 (2020)
- 30 118. Yu, H., Lee, H., Jeon, H.: What is 5g? emerging 5g mobile services and
31 network requirements. *Sustainability* **9**(10), 1848 (2017)
- 32 119. Zamani, A., Sharifian, S.: A novel approach for service function chain
33 (sfc) mapping with multiple sfc instances in a fog-to-cloud computing
34 system. In: 2018 4th Iranian Conference on Signal Processing and Intel-
35 *ligent Systems (ICSPIS)*, pp. 48–52. IEEE (2018)
- 36 120. Zhang, C., Wang, X., Dong, A., Zhao, Y., He, Q., Huang, M.: Energy effi-
37 cient network service deployment across multiple sdn domains. *Computer*
38 *Communications* **151**, 449–462 (2020)
- 39 121. Zhang, C., Wang, X., Dong, A., Zhao, Y., Li, F., Huang, M.: The intel-
40 ligent multi-domain service function chain deployment: Architecture,
41 challenges and solutions. *International Journal of Communication Sys-*
42 *tems* **34**(1), e4665 (2021)
- 43 122. Zhang, X., Xu, Z., Fan, L., Yu, S., Qu, Y.: Near-optimal energy-efficient
44 algorithm for virtual network function placement. *IEEE Transactions on*
45 *Cloud Computing* (2019)
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

1 123. Zhao, D., Sun, G., Liao, D., Xu, S., Chang, V.: Mobile-aware service
2 function chain migration in cloud-fog computing. *Future Generation*
3 *Computer Systems* **96**, 591–604 (2019)
4 124. Zhao, T., Zhou, S., Guo, X., Zhao, Y., Niu, Z.: A cooperative scheduling
5 scheme of local cloud and internet cloud for delay-aware mobile cloud
6 computing. In: *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6.
7 IEEE (2015)
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65