Automated Decision Making based on Pareto Frontiers in the Context of Service Placement in Networks

Michael Seufert, Stanislav Lange, Markus Meixner
University of Würzburg, Institute of Computer Science, Chair of Communication Networks, Würzburg, Germany
{seufert | stanislav.lange | markus.meixner}@informatik.uni-wuerzburg.de

***Abstract*—Virtualization paradigms like cloud computing, software defined networking (SDN), and network functions virtualization (NFV) provide advantages with respect to aspects like flexibility, costs, and scalability. However, management and orchestration of the resulting networks also introduce new challenges. The placement of services, such as virtual machines (VMs), virtualized network functions (VNFs), or SDN controllers, is a multi-objective optimization task that confronts operators with a multitude of possible solutions that are incomparable among each other. The goal of this work is to investigate mechanisms that enable automated decision making between such multi-dimensional solutions. To this end, we investigate techniques from the domain of multi-attribute decision making that aggregate the performance of placements to a single numeric score. A comparison between resulting rankings of placements shows that many techniques produce similar results. Hence, placements that achieve good rankings according to many approaches might be viable candidates in the context of automated decision making. In order to illustrate the functionality of the different scoring mechanisms, we perform a case study on a single network graph and a fixed number of objectives and service instances. Additionally, we present aggregated results from broad evaluations on the Internet Topology Zoo and a larger number of objectives as well as varying numbers of service instances. These allow making more reliable statements about the mechanisms' performance and agreement.**

***Index Terms*—Cloud Service, Placement, Orchestration, Multi-Objective Optimization, Pareto Frontier**

I. Introduction

Network and cloud operators benefit from virtualization paradigms in terms of costs, flexibility, scalability, and vendor independence. In contrast to the prevalent deployment of dedicated computing resources for services like firewalls, load balancers, SDN controllers, or cloud applications, these services can nowadays be virtualized and hosted on commercial off-the-shelf (COTS) hardware deployed anywhere in the network.

However, management and orchestration techniques are required in order to achieve and maintain a high degree of flexibility and assert that QoS and QoE constraints are met. In particular, the placement of service instances within the network can have a significant impact on both, user and operator satisfaction. Since goals like low latency among service instances and low latency between services and end users can be competing, finding suitable placements corresponds to a multi-objective optimization task.

In addition to the increased complexity of algorithms that can solve such problems, the solutions they return can not always be compared with each other due to different domains and units of the objectives. Especially in the context of automated and dynamic service migration and instantiation, however, algorithms need to choose one distinct solution.

The contribution of this work is the investigation of the level of agreement between the rankings of multiple automated decision making algorithms. This is done in a three step approach. First, four methods for determining the relative importance of different objectives are selected and compared with each other. In contrast to approaches that determine such weights a priori, the methods used in this work take into account characteristics of the solutions that are returned by the multi-objective optimization algorithm. Second, four mechanisms for aggregating the performance of a multi-dimensional solution into a single score are selected. Finally, the rankings of solutions that result from different combinations of weighting and aggregation techniques are characterized. On the one hand, analyzing solutions that consistently achieve high ranks according to many approaches might lead to more efficient methods for identifying viable placements. On the other hand, the comparison can help to derive guidelines for choosing the appropriate ranking mechanism for a particular problem.

In a case study, we demonstrate the particular behavior of the investigated mechanisms for an exemplary network and three objective functions, i.e., three optimization goals. Furthermore, we extend our work from [1] by an extensive analysis of 58 real-world network topologies from the Internet Topology Zoo [2], a total of five objective functions, and varying numbers of service instances that are placed. By aggregating the results of these analyses, we can compare the different weighting and ranking methods in terms of aspects like agreement and consistency across problem instances.

The remainder of this work is structured as follows. After an overview of related work in Section II, the data set is presented alongside the resulting problem instances in Section III. The selected methods for assessing the weight of each objective dimension are introduced and compared in Section IV. These methods are then used as input for aggregation algorithms that assign a single score to each placement. In Section V, the four selected aggregation algorithms are discussed and compared with respect to the rankings of placements they produce. Finally, Section VI concludes the work.

II. Related Work

Resouce management in clouds [3], in particular, the placement of cloud services in data centers has become an increasingly important problem. Typically, many parameters and metrics regarding resource utilization and performance have to be taken into account within the cloud and the network. Thus, different methodologies are proposed in literature to place the virtual machines efficiently. In the context of placing virtual network functions (VNF), [4] investigates a weighted sum approach, while [5] uses a linear program to find an optimal placement. Also [6], [7] propose linear programs for chains of VNF, while [7] adds a Pareto analysis to investigate the trade-offs between the different dimensions.

Related problems, which have been discussed recently, are the placement of virtual machines in distributed architectures [8] as well as the placement of SDN controllers. Both are also multi-objective optimization problems, which have to take into account a large set of parameters and metrics. Weighted sums (e.g., [9]) and linear programs (e.g., [10]) are widely used. Additionally, the Pareto frontier is analyzed when different alternatives are incomparable. Due to state explosion, the problem of obtaining the Pareto frontier is frequently tackled heuristically [11], [12]. However, no automated decisions can be taken from Pareto frontiers. Based on our work in [1], we present an extended evaluation of several mechanisms that can be used to approach this problem.

Therefore, we will transform the Pareto frontiers into a ranked list of alternatives. To compare the rankings when the underlying order of alternatives is unknown, we will mainly rely on correlation coefficients and techniques based on probabilistic ranking models. In [13], the rank correlation between the pairs of ranking is calculated using either Spearman's ρ or Kendall's τ. [14] proposes a measure of agreement between rankings based on removal of disputable elements. A basic model for order statistics was developed by Thurstone [15], and [16] constructed an equivalent model based on choice probabilities. Mallow [17] presented simplified and analytically tractable models induced by paired comparison. [18] investigates concordance between different judges (i.e., rankings) based on Mallow's model to detect outlier rankings. [19] proposes to compare the distribution of ranks by box plots and derive a degree of discordance based on the interquartile range. The goodness of fit of simple ranking models is investigated in [20], and metric based ranking models are discussed in [21]. A classification of probabilistic ranking models can be found in [22].

III. Data Set Description

In order to investigate the practical feasibility of the different weighting and ranking methods that are discussed in this work, realistic input data is required. To this end, we use 58 different network graphs from the Internet Topology Zoo [2] and use the freely available POCO tool [23] to exhaustively evaluate all possible service placements with respect to a total of up to five objective functions. While results are consistent among different networks, some characteristics depend on statistics like the number of nodes and the diameter of the graph. On the one hand, we present detailed results and statistics for the Internet2 OS3E topology which is chosen as an exemplary representative. This allows for accurate insights into the functionality of the different weighting and scoring mechanisms. On the other hand, we provide aggregated results and statistics regarding the whole data set in order to identify topology-independent relationships between the various mechanisms.

A. Internet2 OS3E

Table I provides an overview of the Internet2 graph as well as the resulting problem instance. In order to keep the solution space small enough to visually illustrate the effects and behavior of the presented methods, only four service instances are placed in the network and the number of objectives is limited to three. Although this results in a total of $46,376$ distinct placements, only ten of those are Pareto optimal and thus relevant during the decision making process. In the context of larger search spaces, e.g., when placing more services or dealing with networks that have more nodes, an exhaustive evaluation of all possible placements might not be feasible due to time and resource constraints. For such cases, a trade-off between accuracy and runtime can be achieved by employing heuristic approaches that can approximate the Pareto frontier [11].

TABLE I: Information regarding the Internet2 OS3E topology and the corresponding problem instance used in the case study.

Property	**Value**
Number of nodes	34
Number of placed services	4
Number of distinct placements	$46,376$
Number of Pareto optimal placements	10
Objective functions	Mean latency to end users $\pi^{\text{avg latency}}$ Maximum latency to end users $\pi^{\text{max latency}}$ Imbalance between service instances $\pi^{\text{imbalance}}$

As mentioned in the previous paragraph, three different objective functions are taken into account when assessing the performance of each placement. These include two latency-related measures, namely, the mean and average latency between services and end users. We use the longitude and latitude information that is provided for each node to calculate the Euclidean distance between nodes and approximate the latency of each link. For multi-hop paths, the latency is defined as the sum of latencies of all involved links. Furthermore, the load imbalance between service instances is defined as the difference between the number of end users assigned to the instance with the highest and lowest amount of end users, respectively. Several statistical properties of these objective functions are presented in Table II. Additionally, Figure 1 displays the cumulative distribution function of objective values that are attained across all placements.

Due to the fact that the latency measures are continuous, they yield significantly more distinct values, resulting in smooth CDF curves. In contrast, the imbalance is always an integer value which is constrained by the number of nodes in the topology. Hence, individual steps are visible in the plot. Since the average latency between end users and services is calculated from 34 individual latencies, outliers are smoothed out and the resulting variance is relatively low. The values of the maximum latency objective have a higher variance and fewer distinct values since the maximum does not necessarily change between similar placements that share multiple controller locations.

TABLE II: Various statistics of the objective functions that are used in the case study.

Objective	Number of distinct values	Mean	Variance
$\pi^{\text{avg latency}}$	$45,311$	0.195	0.001
$\pi^{\text{max latency}}$	244	0.491	0.013
$\pi^{\text{imbalance}}$	29	0.305	0.019

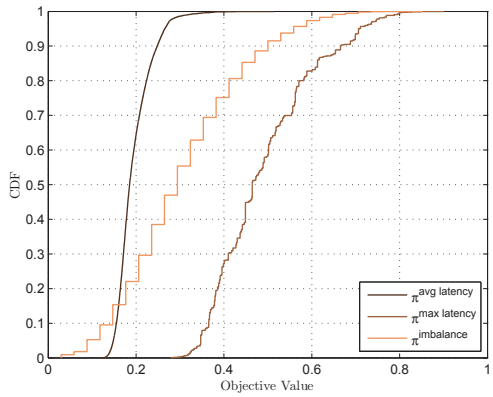


Fig. 1: Empirical CDFs of objective values that are attained in the example scenario.

B. Internet Topology Zoo

For the evaluation of networks from the Internet Topology Zoo, we chose networks whose size n ranges between 25 and 50 nodes. This ensures that the exhaustive evaluation of all possible placements with POCO can be performed within a reasonable time frame. Using each of the resulting 58 topologies, we calculated placements of $k \in \{3, 4, 5\}$ service instances and evaluated them with respect to a total of five objectives, resulting in a total of 174 problem instances. In addition to the abovementioned imbalance and latency measures, the average and maximum latency between each pair of service instances is also taken into account. These objectives are referred to as $\pi^{\text{avg inter-latency}}$ and $\pi^{\text{max inter-latency}}$, respectively. In the following, we present various aggregated statistics of the set of problem instances that are discussed in this work.

Similarly to Table II, Figure 2 presents the distribution of the number of distinct values that are attained by each objective function per problem instance. In this context, a problem instance is characterized by the number of placed instances, k, and the network graph. Qualitatively, the statistics across all problem instances are similar to those in the Internet 2 graph, i.e., the continuous average latency measures have the largest number of distinct values. They are followed by the objectives that consider the maximum latency, which are also continuous. Finally, the imbalance is integer-valued and restricted by the number of nodes in the network, n. Hence, it attains the lowest amount of distinct values.

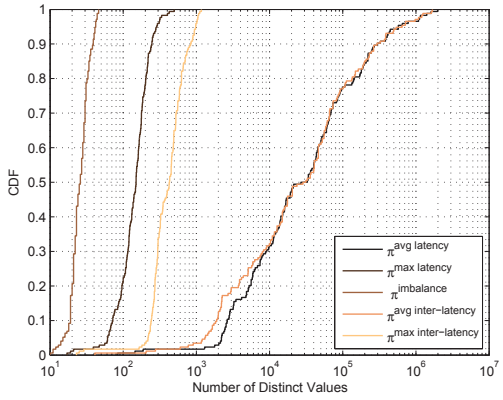
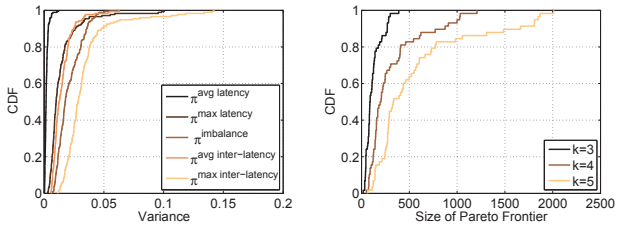


Fig. 2: Empirical CDFs of the number of distinct function values per problem instance for each objective function.

In addition to the number of distinct values, the variance plays an important role when quantifying the relative importance of an objective. For all problem instances, Figure 3 displays the distribution of the variance of each objective across all possible placements for a problem instance. Although the average latency measures attain the highest number of distinct values, they show the lowest variance. The reason for this characteristic is that the averages are formed from many individual latencies and do not differ much between placements. Furthermore, the variance of the average inter-instance latency is higher than that of the average latency between end users and services. This stems from the fact that $\pi^{\text{avg inter-latency}}$ is based on fewer individual latencies and can take on extreme values when all instances are placed close to each other in a cluster or are distributed at the edge of the network, respectively. As discussed in Section III-A, objectives that quantify the maximum latency have a higher variance due to the wider range of attained values (cf. Figure 1). Similarly, the imbalance measure is based on the maximum load difference between service instances and takes on a large number of values, resulting in a high variance. For all objectives, the $90\,\%$ quantiles of the variance distribution are below 0.05.

To further motivate the need for mechanisms that map a multi-obective result vector to a single score, the distributions

of the number of Pareto-optimal placements for different numbers of placed service instances, k, are illustrated in Figure 3b. For a given number m of Pareto-optimal placements on the x-axis, the value on the y-axis represents the fraction of problem instances whose five dimensional Pareto frontier includes up to m elements. The different numbers of service instances, k, are denoted by differently colored curves. The number of placed service instances has a direct impact on the total number of distinct placements, which can be calculated as the binomial coefficient $\binom{n}{k}$. Hence, the size of the Pareto frontier also increases due to the increase of incomparable pairs of objective vectors, in particular. While a human decision maker might compare and choose from a couple of alternatives in an objective space with few dimensions, comparing one thousand and more different solutions in a practical time frame seems unlikely.



(a) Variance per objective function across all problem instances investigated in this work.

(b) Number of Pareto optimal placements for different numbers of service instances that are placed.

Fig. 3: Empirical CDFs of the variance per objective and the size of the Pareto frontier.

IV. Weighting Methods

In order to aggregate the performance of a placement that is evaluated with respect to multiple objective functions into a single value, the mechanisms that are analyzed in this work require weights for each considered dimension. Hence, we first discuss methods for obtaining these weights based on the set of placements and the corresponding objective values.

In the following, the weight of the $j-$th objective is denoted as w_j and weights are normalized, i.e., $\sum_{j=1}^{m} w_j = 1$ in case of m objective functions. Additionally, objective values are also normalized prior to applying the weighting mechanisms. The observed values for n placements and m objective dimensions are stored in an $n \times m$ matrix A which is transformed into the normalized matrix R according to Equation 1.

$$r_{ij} = \frac{a_j^{max} + a_j^{min} - a_{ij}}{a_j^{max} + a_j^{min}} \tag{1}$$

In this equation, $a_j^{min} = \min_i a_{ij}$ and $a_j^{max} = \max_i a_{ij}$ refer to the minimum and maximum values of the j-th objective, respectively.

A. Uniform Weighting

As a baseline naïve approach, we use a weighting mechanism that does not take into account any observed data and assigns equal weights to every objective, i.e., $w_j^{uni} = \frac{1}{m}$.

B. Entropy-Based Weighting

In information theory, (the Shannon) entropy is used as a means to quantify the amount of information that is stored in a message [24]. The key idea behind the entropy-based weighting method consists of assigning higher weights to objective dimensions that carry more information, i.e., those that have a higher number of distinct values and low individual occurrence probabilities for each value. Based on [25], the weights are calculated in three steps. First, observed values are normalized for each dimension (cf., Equation 2).

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^{n} r_{ij}},\ j \in \{1, \ldots, m\} \tag{2}$$

Then, the entropy is determined by means of

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} p_{ij} \ln p_{ij},\ j \in \{1, \ldots, m\}. \tag{3}$$

Finally, the weight is calculated as

$$w_j^{ent} = \frac{1 - e_j}{\sum_{i=1}^{m}(1 - e_i)},\ j \in \{1, \ldots, m\}. \tag{4}$$

C. Weighting Based on the Coefficient of Variation

Intuitively, objectives whose values cover a wide range of different values tend to have a higher impact on the total resulting performance of a placement than objectives that attain only few values or values that are very close to each other. Hence, we investigate the suitability of the coefficient of variation for quantifying the relative importance of an objective. The coefficient of variation is defined as the ratio between the standard deviation and the mean of observed values. Thus, the weights are calculated according to Equation 5. σ_j and μ_j refer to the standard deviation and mean of the j-th objective, respectively.

$$w_j^{cv} = \frac{\frac{\sigma_j}{\mu_j}}{\sum_{i=1}^{m} \frac{\sigma_i}{\mu_i}},\ j \in \{1, \ldots, m\} \tag{5}$$

D. Weighting Based on the Standard Deviation

Similarly to the weighting approach that is based on the coefficient of variation, this mechanism uses the standard deviation in order to calculate the relative weights.

$$w_j^{sd} = \frac{\sigma_j}{\sum_{i=1}^{m} \sigma_i},\ j \in \{1, \ldots, m\} \tag{6}$$

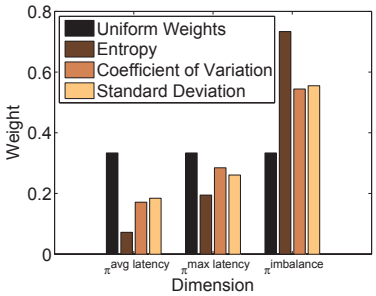


Fig. 4: Relative weights of objectives according to different weighting mechanisms for Internet2 OS3E topology.

E. Comparison

In order to allow for a comparison between the different weighting mechanisms, Figure 4 presents the weights of individual objectives according to the four weighting approaches for the Internet2 OS3E topology. The x-axis denotes the three objectives, the height and color of the bars represent the weight and weighting method, respectively.

While the weights that are returned by the different mechanisms differ in terms of absolute values, the relative order of objectives is consistent. Having the lowest variance and the narrowest interquartile range, the latency between end users and services is assigned the lowest weights. As discussed in Section III, the maximum-based measure has a higher variance and thus also results in higher weights when compared to its average-based counterpart. The highest weights are assigned to the imbalance measure. This can be explained by the high variance that is observed for the imbalance objective.

A comparison of the absolute weights that are assigned by the weighting methods shows that the mechanisms that are based on standard deviation and the coefficient of variation return similar values. This phenomenon can be explained by the fact that objective values are normalized prior to applying the weighting methods. Thus, the normalization using the mean that is applied in the context of the latter does not have a large impact on the final weights. Finally, the entropy-based weighting approach yields the widest range of weights, i.e., between less than 0.1 and more than 0.6. This indicates a higher sensitivity towards the objectives' variance, which seems to be the main influence factor on the resulting weight for all weighting methods that take into account observed objective values.

Figure 5 shows the weight coefficients for each weighting method applied to the problem instances of the Topology Zoo and five dimensions. Figure 5a shows weights for uniform weighting, Figure 5b for entropy-based weighting, and the bottom plots show the resulting weights based on coefficient of variation (Figure 5c) and standard deviation (Figure 5d), respectively. The x-axis of each plot represents the IDs of the different problem instances. Each bar shows the weights of each dimension according to the investigated weighting method. From bottom to top, the weights of average latency (black), maximum latency (dark brown), imbalance (light brown), average inter-latency (orange), and maximum inter-latency (yellow) are stacked. For better visibility, the legend was omitted for Figures 5b–5d, but is the same as in Figure 5a for all plots.

The uniform weighting assigns each dimension the same weight, which can be seen in Figure 5a. The entropy-based weighting in Figure 5b gives high scores to the maximum inter-latency, which receives a weight of around 0.5 for most of the topologies. As in the case study of the Internet2 OS3E topology, the high variance of the maximum inter-latency (cf. Figure 3a) is responsible for the high weights. The second highest scores are given either to average inter-latency or imbalance, which are highly fluctuating depending on the particular problem instance. Maximum latency and average latency receive the smallest weights due to the small variance. As already observed for the Internet2 OS3E topology, also for the Topology Zoo problem instances the weighting based on coefficient of variation resembles much the weighting based on standard deviation, and both weight distributions are less skewed than entropy-based weights. While the weighting based on coefficient of variation generally gives higher weights to maximum inter-latency and lower weights to average latency, the resulting weights of the standard deviation method are closer to the uniform weighting. In this case, maximum inter-latency, average inter-latency, and imbalance have weights of around 0.25 each, and maximum latency and average latency share the remaining 0.25 almost equally.

To sum up, the weighting method has a significant impact. Aside from uniform weighting, the other weighting algorithms emphasize the characteristics of the dimensions differently, which results in divergent weightings. This is not only observed for the case study of the Internet2 OS3E graph, but also for the problem instances of the Topology Zoo. Especially, the entropy-based weighting results in the most skew weights and shows a high variability for different topologies. In contrast, weighting based on coefficient of variation and standard deviation results in less skew and consistent weights over all graphs.

V. Ranking Methods and Results

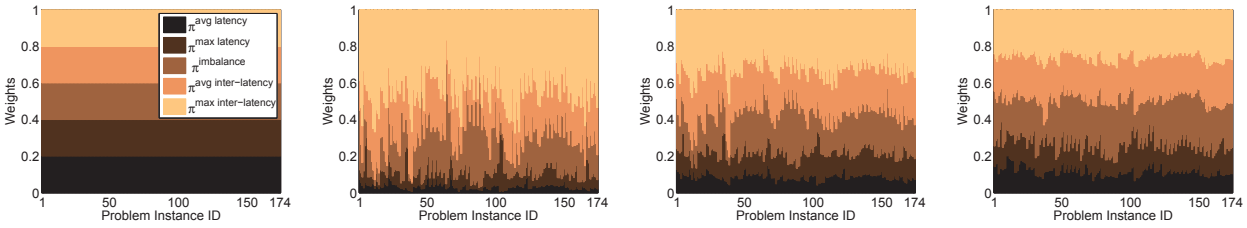
A. Ranking Methods

To aggregate the scores a_{ij} of the different attributes j of the placement i to an overall ranking score ρ_i, four well-known multi-attribute decision methods will be considered.

First, we consider Simple Additive Weighting (SAW) [26], which computes the overall score by adding the normalized attribute scores $r_{ij} = \frac{a_j^{min}}{a_{ij}}$ multiplied by the weights w_j.

$$\rho_i^{SAW} = \sum_j w_j \cdot r_{ij}$$

A similar ranking method is Multiplicative Exponent Weighting (MEW) [27], which calculates the overall score as the product of the normalized attribute scores $r_{ij} = \frac{a_j^{min}}{a_{ij}}$,



(a) Uniform weighting (b) Entropy-based weighting (c) Weighting based on coefficient of variation (d) Weighting based on standard deviation

Fig. 5: Relative weights of objectives according to different weighting mechanisms for the Topology Zoo. The legend of Figure 5a applies to all plots.

which are given the respective weight as exponent.

$$\rho_i^{MEW} = \prod_j r_{ij}^{w_j}$$

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [26] normalizes the attributes $r_{ij} = \frac{a_{ij}}{\sum_i a_{ij}^2}$, and computes the distances to an optimal placement with all best weighted normalized attribute values $v_j^{min} = \min_i(w_j r_{ij})$, and to a worst placement composed of all worst weighted normalized attribute values $v_j^{max} = \max_i(w_j r_{ij})$. Then, the separation between the optimal and the worst placement is computed by

$$s_i^{min} = \sqrt{\sum_j (w_j r_{ij} - v_j^{min})^2}$$

$$s_i^{max} = \sqrt{\sum_j (w_j r_{ij} - v_j^{max})^2}$$

The resulting ranking ρ_i is the relative closeness to the ideal solution:

$$\rho_i^{TOPSIS} = \frac{s_i^{max}}{s_i^{min} + s_i^{max}}$$

VIKOR [28] relies on the best and worst attribute values, a_j^{min} and a_j^{max}. Then, for each placement, scores are calculated by two strategies:

$$S_i = \sum_j w_j \frac{a_j^{min} - a_{ij}}{a_j^{min} - a_j^{max}}, \quad R_i = \max_j \left(w_j \frac{a_j^{min} - a_{ij}}{a_j^{min} - a_j^{max}} \right)$$

The final ranking score for each placement is then computed with a parameter γ, $0 \leq \gamma \leq 1$, for the weight of each strategy, and the best and worst values of S_i and R_i, i.e., $S^{min} = \min_i S_i$, $S^{max} = \max_i S_i$, $R^{min} = \min_i R_i$, $R^{max} = \max_i R_i$:

$$\rho_i^{VIKOR} = \gamma \frac{S_i - S^{min}}{S^{max} - S^{min}} + (1-\gamma) \frac{R_i - R^{min}}{R^{max} - R^{min}}$$

We set $\gamma = 0.5$ to give equal weight to both strategies.

Together with the four weighting methods presented in Section IV, this gives 16 different ranking methods, i.e., weighting-ranking combinations, for the multi-objective placement problem. Due to the vast amount of distinct placements, we will apply the 16 methods only to the subset of Pareto-optimal placements, i.e., the set of placements in which no attribute can outperform any other attribute.

B. Case Study for Internet2 OS3E Topology

First, the performance of the weighting-ranking combinations is investigated for the Internet2 OS3E topology and three dimensions, i.e., rankings of the ten Pareto-optimal points are compared. Table III lists the highest and lowest correlations between different combinations in terms of Kendalls's τ and Spearman's ρ rank order correlation coefficients. It can be seen that generally high correlations can be achieved between all ranking algorithms. In contrast, small negative correlation can be seen only for VIKOR with uniform weights. Thus, this might give some evidence that the investigated algorithms mainly agree on the inherent order of the elements.

TABLE III: Highest and lowest correlations between different combinations of weighting and ranking methods on Internet2 OS3E topology.

Method 1	**Method 2**	τ	ρ
(w^{ent}, ρ^{SAW})	(w^{ent}, ρ^{MEW})	1.00	1.00
(w^{ent}, ρ^{SAW})	(w^{ent}, ρ^{TOPSIS})	1.00	1.00
(w^{ent}, ρ^{SAW})	(w^{ent}, ρ^{VIKOR})	1.00	1.00
(w^{sd}, ρ^{MEW})	(w^{sd}, ρ^{TOPSIS})	1.00	1.00
(w^{ent}, ρ^{MEW})	(w^{ent}, ρ^{TOPSIS})	1.00	1.00
(w^{ent}, ρ^{MEW})	(w^{ent}, ρ^{VIKOR})	1.00	1.00
(w^{uni}, ρ^{MEW})	(w^{uni}, ρ^{TOPSIS})	1.00	1.00
(w^{ent}, ρ^{TOPSIS})	(w^{ent}, ρ^{VIKOR})	1.00	1.00
(w^{sd}, ρ^{SAW})	(w^{uni}, ρ^{VIKOR})	−0.11	−0.16
(w^{ent}, ρ^{SAW})	(w^{uni}, ρ^{VIKOR})	−0.11	−0.15
(w^{ent}, ρ^{MEW})	(w^{uni}, ρ^{VIKOR})	−0.11	−0.15
(w^{ent}, ρ^{TOPSIS})	(w^{uni}, ρ^{VIKOR})	−0.11	−0.15
(w^{ent}, ρ^{VIKOR})	(w^{uni}, ρ^{VIKOR})	−0.11	−0.15

Another metric for measuring the agreement between rankings was proposed by Gordon [14]. Gordon's α is defined as the number of objects, which are contributing to the agreement between the rankings: $\alpha := N - \delta$. Thus, it can be computed as the difference between the length of the ranking N and the minimum number of objects δ, which have to be removed to ensure a perfect agreement between the reduced rankings.

Gordon's α confirms the high correlation coefficients, as there are many pairs of rankings with a perfect agreement of $\alpha = N = 10$, see Table IV. The lowest value of α is 4, which shows that still the ranking order is not completely inverted by any weighting-ranking combination.

TABLE IV: Highest and lowest Gordon α scores for weighting-ranking combinations on Internet2 OS3E topology.

Method 1	**Method 2**	α
(w^{ent}, ρ^{SAW})	(w^{ent}, ρ^{MEW})	10
(w^{ent}, ρ^{SAW})	(w^{ent}, ρ^{TOPSIS})	10
(w^{ent}, ρ^{SAW})	(w^{ent}, ρ^{VIKOR})	10
(w^{sd}, ρ^{MEW})	(w^{sd}, ρ^{TOPSIS})	10
(w^{ent}, ρ^{MEW})	(w^{ent}, ρ^{VIKOR})	10
(w^{uni}, ρ^{MEW})	(w^{sd}, ρ^{TOPSIS})	10
(w^{ent}, ρ^{TOPSIS})	(w^{ent}, ρ^{VIKOR})	10
(w^{ent}, ρ^{SAW})	(w^{uni}, ρ^{VIKOR})	4
(w^{uni}, ρ^{SAW})	(w^{sd}, ρ^{VIKOR})	4
(w^{sd}, ρ^{MEW})	(w^{uni}, ρ^{VIKOR})	4
(w^{uni}, ρ^{MEW})	(w^{uni}, ρ^{VIKOR})	4
(w^{sd}, ρ^{TOPSIS})	(w^{uni}, ρ^{VIKOR})	4
(w^{uni}, ρ^{TOPSIS})	(w^{uni}, ρ^{VIKOR})	4

Probabilistic ranking models give another approach to comparing the obtained rankings. Luce [16] constructs probabilities for a ranking $\rho = (i_1, i_2, \ldots, i_N)$ from conditional probabilities. Thus, after $r-1$ stages, p_{i_r} is defined as the probability that the element i_r is the most preferred element from the set of remaining elements $B = \{i_r, \ldots, i_N\}$. By repeating the choice, this gives the probability of the rating ρ as:

$$P(\rho) = \prod_{r=1}^{N-1} \frac{p_{i_r}}{\sum_{j \in B} p_j}$$

The highest Luce probabilities are obtained by a ranking, which was created by the combinations $(w^{uni}|w^{sd}, \rho^{MEW})$ and $(w^{uni}|w^{cv}|w^{sd}, \rho^{TOPSIS})$. This means, this ranking gives high ranks to the elements, which are most preferred by all weighting-ranking combinations. Note that this ranking is also the modal ranking in the resulting set of rankings. All four entropy based algorithms output the same ranking, which reaches the second highest Luce probabilities. Towards the other end, the SAW and VIKOR algorithms and the standard deviation (sd) and coefficient of variation weighting (cv) output rankings with low probabilities (with the above mentioned exceptions).

Mallow's Φ-model is based on paired comparison of the ranked elements. It can be formulated as

$$P_{\rho_0,\theta}(\rho) = \left(\sum_{\rho} \theta^{X(\rho_0,\rho)} \right)^{-1} \cdot \theta^{X(\rho_0,\rho)}, \ 0 \leq \theta < \infty,$$

in which $X(\rho_0, \rho)$ is Kendall's τ distance, i.e., the number of disagreements between ρ_0 and ρ. ρ_0 is an a priori set location parameter (e.g., the modal ranking or an averaged ranking), and θ is a measure of variation, which will be fitted from the rankings with a table given in [18]. Following the methodology presented by Feigin and Cohen in [18], the model also allows to detect outlier rankings. Using the averaged ranking as location parameter and fitting θ accordingly, the highest probability is obtained by the ranking of (w^{cv}, ρ^{MEW}). The second highest probabilities are achieved by the modal ranking, which already accounted for the highest Luce probabilities. Again the entropy rankings have the third highest probability. This means that these three rankings are closest to the averaged ranking, which was chosen as location parameter. Using the modal ranking as location parameter, the order of the first and second rating would change, but the entropy rating would still receive the third highest probability. The outlier detection, which mainly depends on the fitting of θ, indicates that (w^{cv}, ρ^{MEW}) is an outlier ranking with a too high probability, and (w^{uni}, ρ^{SAW}) and (w^{uni}, ρ^{VIKOR}) are outliers with a too low probability close to 0. In particular, this means that the disagreements for (w^{uni}, ρ^{SAW}) and (w^{uni}, ρ^{VIKOR}) are exceptionally high compared to the averaged or modal ranking.

Following the approach described in [19], Figure 6 shows a boxplot of the ranks of the different placements sorted by median. It can be seen that there are small boxes for the first five placements, which means that there is a large agreement among the different weighting-ranking combinations. Only for the last five placements, there is some disagreement among the different rankings. Still several outliers can be observed, however, taking a detailed look at the data, most outlier ratings stem from uniform weighting of the attributes. Thus, this weighting method seems to be inappropriate for ranking the placements.

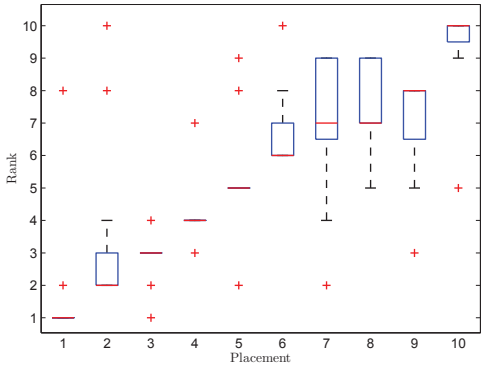
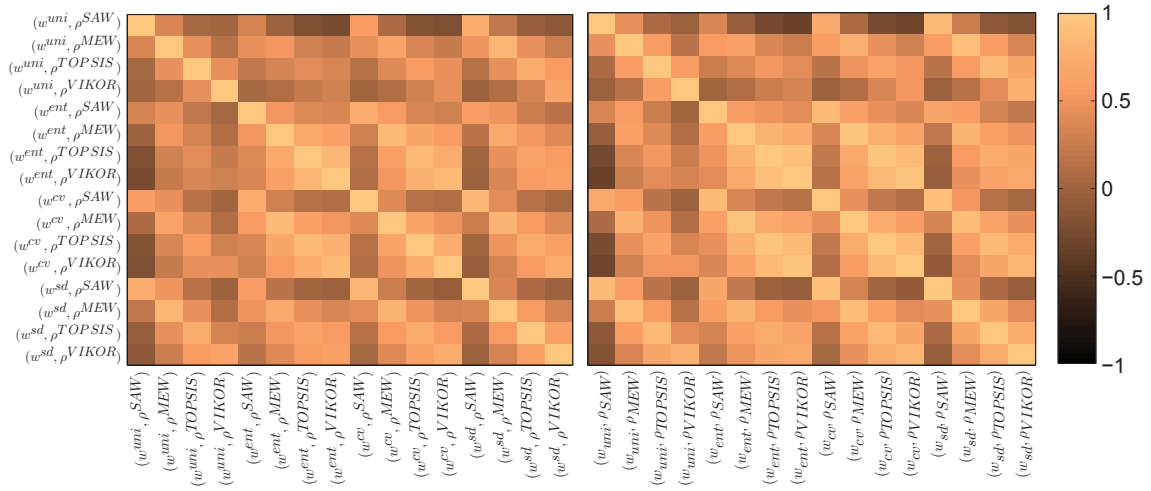


Fig. 6: Pareto optimal placements and their ranks according to the presented ranking mechanisms.

To sum up, the different ranking methods showed a high agreement, especially for the top-ranked placements. This means, among the investigated methods, no weighting-ranking combination stands out and most of them are well suited to combine the Pareto-optimal placements into a single score. Nevertheless, the results suggest that the use of uniform weights can lead to outlier rankings, which do not reproduce the majority rankings.



(a) Kendall correlation coefficient τ (b) Spearman correlation coefficient ρ

Fig. 7: Average correlation coefficients between rankings of different weighting-ranking combinations for all topologies.

C. Broad Evaluation on the Topology Zoo

The presented case study is generalized by applying the methodology to the 174 problem instances of the Topology Zoo and five dimensions. The aggregated performance over all instances provides better insights on the performance of each of the weighting-ranking combinations. Figure 7a shows the average Kendall τ correlation coefficient for all pairs of weighting-ranking combinations. Figure 7b shows the corresponding Spearman ρ correlation coefficients. In both figures, the color at area (i, j) indicates the average correlation coefficient between combination i and combination j over all problem instances according to the color scale on the right. A perfect correlation is indicated by a yellow area (e.g., $(i, i)\ \forall\, i$), while darker colors show a lower correlation. A correlation coefficient of 0 is shown in brown, while dark brown to black colors represent negative correlations.

It can be seen that both average correlation coefficients give similar results for all pairs of combinations. The darker colors of row/column 1, 5, 8, and 13 indicate that the rankings generated by SAW generally have little or even negative correlations with the other rankings. The average Kendall and Spearman correlations are especially low for combination 1, i.e., (w^{uni}, ρ^{SAW}). Moreover, it can be seen that uniform weighting (1 – 4) results in lower average correlations. The highest correlations can be observed among the combinations with entropy-based weighting (5 – 8), which means that those combinations generally output similar rankings.

Figure 8 shows the aggregated performance of all weighting-ranking combinations according to the Luce probabilities. For each combination, it shows the number of problem instances for which its ranking had the highest Luce probability. It can be seen that combination 10, i.e., (w^{cv}, ρ^{MEW}), outputs the ranking with the highest Luce probability on almost half of the problem instances. In general, it can be seen that the ranking of combinations with MEW (2, 6, 10, 14) often has the highest Luce probability. According to that evaluation, SAW (1, 5, 9, 13) performs second best. The other ranking methods perform much worse. Regarding the weighting algorithm, weightings based on coefficient of variation (9–12) and entropy (5–8) perform best. They often output rankings with high Luce probabilities, while uniform weighting and weighting based on standard deviation give rankings with low probabilities.

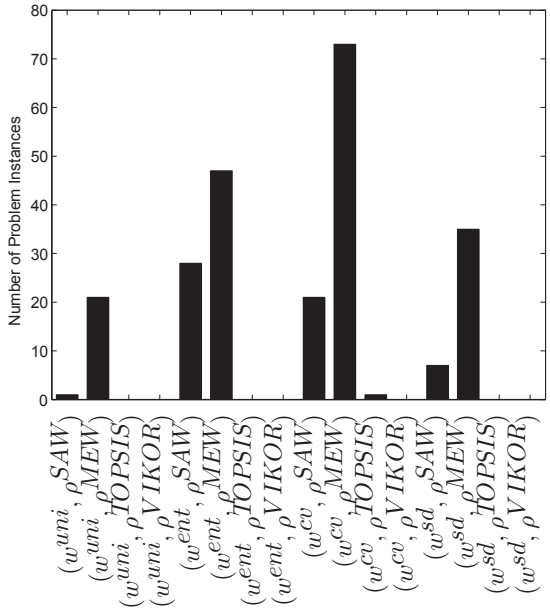


Fig. 8: Number of topologies for which weighting-ranking combination output the ranking with highest Luce probability.

All in all, the problem instances of the Topology Zoo reveal findings similar to those from the case study with the Internet2 OS3E topology. Uniform weightings are likely to produce

rankings, which show little correlations to the rankings of other weighting-ranking combinations. Also the usage of the SAW ranking algorithm provides lower correlation coefficients, but the resulting rankings have the highest Luce probabilities on some problem instances. A worse performance is observed for the TOPSIS and VIKOR algorithms, which could not output rankings with high Luce probabilities. Instead, with respect to these probabilities, all MEW combinations, especially (w^{cv}, ρ^{MEW}), should be considered for ranking the Pareto-optimal placements.

VI. Conclusion

In this work, we applied multi-objective decision making methods to the problem of selecting the best placement for a cloud service from a set of Pareto-optimal placements. Therefore, we investigated four methods to determine the relative importance of individual objectives (i.e., uniform, entropy-based, coefficient of variation-based, and standard deviation-based weighting), and four methods for aggregating the performance of solution sets that are returned by multi-objective optimization algorithms (i.e., simple additive weighting, multiplicative exponent weighting, TOPSIS, and VIKOR). We demonstrated, both for a single case study and for a broad evaluation on a large set of problem instances, that most combinations of weighting and aggregation algorithms perform sufficiently good for the investigated problem and have a high level of agreement, especially on the top-ranked placements. Only the usage of uniform weights was shown to cause outlier rankings, which, nevertheless, can provide a complementary view on the ranked placements. The best weighting-ranking combination was multiplicative exponent weighting based on coefficient of variation (i.e., (w^{cv}, ρ^{MEW})), which outputs rankings with the highest Luce probabilities for the largest set of problem instances. In case of all approaches, the weights and ranks for a given set of placements can be efficiently calculated by implementing the presented equations. However, operators might need to employ heuristic approaches for large problem instances in order to find feasible placements in a timely manner. Eventually, the goal will be to derive guidelines for choosing the appropriate ranking mechanism for Pareto-optimal placements in the cloud service placement problem.

Acknowledgment

This work was partly funded in the framework of the EU ICT project INPUT (H2020-2014-ICT-644672), the DFG grant QoE-DZ (TR257/41), and the CELTIC EUREKA project SENDATE-PLANETS (Project ID C2015/3-1), which is partly funded by the German BMBF (Project ID 16KIS0474). The authors alone are responsible for the content.

References

[1] M. Seufert, S. Lange, and M. Meixner, "Automated decision making methods for the multi-objective optimization task of cloud service placement," in *1st International Workshop on Programmability for Cloud Networks and Applications (PROCON)*, Würzburg, Germany, 2016.

[2] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The Internet Topology Zoo," *IEEE JSAC*, vol. 29, no. 9, 2011.

[3] B. Jennings and R. Stadler, "Resource Management in Clouds: Survey and Research Challenges," *Journal of Network and Systems Management*, vol. 23, no. 3, pp. 567–619, 2015.

[4] M. Bari, S. R. Chowdhury, R. Ahmed, R. Boutaba *et al.*, "On Orchestrating Virtual Network Functions," in *11th International Conference on Network and Service Management (CNSM)*, Barcelona, Spain, 2015.

[5] M. C. Luizelli, L. R. Bays, L. S. Buriol, M. P. Barcellos, and L. P. Gaspary, "Piecing Together the NFV Provisioning Puzzle: Efficient Placement and Chaining of Virtual Network Functions," in *IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Ottawa, ON, Canada, 2015.

[6] H. Moens and F. De Turck, "VNF-P: A Model for Efficient Placement of Virtualized Network Functions," in *10th International Conference on Network and Service Management (CNSM)*, Rio de Janeiro, Brazil, 2014.

[7] S. Mehraghdam, M. Keller, and H. Karl, "Specifying and Placing Chains of Virtual Network Functions," in *3rd International Conference on Cloud Networking (CloudNet)*, Luxembourg, Luxembourg, 2014.

[8] Z. Á. Mann, "Allocation of virtual machines in cloud data centersa survey of problem models and optimization algorithms," *ACM Computing Surveys (CSUR)*, vol. 48, no. 1, p. 11, 2015.

[9] M. F. Bari, A. R. Roy, S. R. Chowdhury, Q. Zhang, M. F. Zhani, R. Ahmed, and R. Boutaba, "Dynamic Controller Provisioning in Software Defined Networks," in *9th International Conference on Network and Service Management (CNSM)*, Zurich, Switzerland, 2013.

[10] R. Guerzoni, R. Trivisonno, I. Vaishnavi, Z. Despotovic, A. Hecker, S. Beker, and D. Soldani, "A Novel Approach to Virtual Networks Embedding for SDN Management and Orchestration," in *Network Operations and Management Symposium (NOMS)*, Krakow, Poland, 2014.

[11] S. Lange, S. Gebert, T. Zinner, P. Tran-Gia, D. Hock, M. Jarschel, , and M. Hoffmann, "Heuristic Approaches to the Controller Placement Problem in Large Scale SDN Networks," *IEEE Transactions on Network and Service Management - Special Issue on Efficient Management of SDN and NFV-based Systems*, vol. 12, 2015.

[12] S. Lange, S. Gebert, J. Spoerhase, P. Rygielski, T. Zinner, S. Kounev, and P. Tran-Gia, "Specialized Heuristics for the Controller Placement Problem in Large Scale SDN Networks," in *International Teletraffic Congress (ITC 27)*, Ghent, Belgium, 2015.

[13] M. G. Kendall, *Rank correlation methods.* Griffin, 1948.

[14] A. Gordon, "A measure of the agreement between rankings," *Biometrika*, vol. 66, no. 1, pp. 7–15, 1979.

[15] L. L. Thurstone, "A law of comparative judgment." *Psychological review*, vol. 34, no. 4, p. 273, 1927.

[16] R. D. Luce, *Individual choice behavior.* John Wiley, 1959.

[17] C. L. Mallows, "Non-null ranking models. i," *Biometrika*, vol. 44, no. 1/2, pp. 114–130, 1957.

[18] P. D. Feigin and A. Cohen, "On a model for concordance between judges," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 203–213, 1978.

[19] A. Cohen, "Analysis of large sets of ranking data," *Communications in Statistics-Theory and Methods*, vol. 11, no. 3, pp. 235–256, 1982.

[20] A. Cohen and C. L. Mallows, "Assessing goodness of fit of ranking models to data," *The Statistician*, pp. 361–374, 1983.

[21] P. Diaconis, "Group representations in probability and statistics," *Lecture Notes-Monograph Series*, vol. 11, pp. i–192, 1988.

[22] D. E. Critchlow, M. A. Fligner, and J. S. Verducci, "Probability models on rankings," *Journal of mathematical psychology*, vol. 35, no. 3, pp. 294–318, 1991.

[23] D. Hock, M. Hartmann, S. Gebert, M. Jarschel, T. Zinner, and P. Tran-Gia, "Pareto-Optimal Resilient Controller Placement in SDN-based Core Networks," in *25th International Teletraffic Congress (ITC)*, 2013.

[24] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 1948.

[25] J. Huang, "Combining entropy weight and TOPSIS method for information system selection," in *IEEE Conference on Cybernetics and Intelligent Systems*, 2008.

[26] C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making - Methods and Applications: A State-of-the-art Survey*. Springer Science & Business Media, 1981, vol. 186.

[27] S. H. Zanakis, A. Solomon, N. Wishart, and S. Dublish, "Multi-attribute Decision Making: A Simulation Comparison of Select Methods," *European Journal of Operational Research*, vol. 107, no. 3, pp. 507–529, 1998.

[28] S. Opricovic and G.-H. Tzeng, "Compromise Solution by MCDM Methods: A Comparative Analysis of VIKOR and TOPSIS," *European Journal of Operational Research*, vol. 156, no. 2, pp. 445–455, 2004.