

# A Mathematical Measure for Flexibility in Communication Networks

Markus Klügel, Mu He, Wolfgang Kellerer, and Péter Babarczi  
 Chair of Communication Networks, Technical University of Munich, Germany  
 E-mail: {markus.kluegel, mu.he, wolfgang.kellerer, peter.babarczi}@tum.de

**Abstract**—For communication networks research, flexibility of network design and networking solutions is considered a competitive advantage. However, this advantage is typically only claimed on an argumentative level and neither formally supported nor thoroughly investigated. To support the claim of flexibility and to make the flexibility of different solutions comparable, its degree must be quantified and thus made measurable. In this work, we propose a mathematical basis to quantify a degree of flexibility achieved by communication networks. We motivate that flexibility can be cast to the “size” of a set of achievable demand changes. Consequently, we propose the use of mathematical measure theory to quantify achieved networking flexibility. We derive several implications on the basic structure of flexibility, extend the insights towards a utility of flexibility and develop systematic approaches for both analytical and empirical measurement of flexibility. We apply the insights to several use-cases, showing that flexibility is in fact not as straight-forward to argue as it seems at first glance.

## I. INTRODUCTION

In the last decades, communication networks have interweaved with all areas of our society, influencing areas as different as social media, industrial production and health-care. New application requirements create a need for dynamic changes of the used networking resources, e.g., to react to social events or to shifts of communication demands. It is commonly accepted that the ossification of existing networks – in particular, of the Internet – constitutes a lack of flexibility to adapt networks to changing requirements efficiently and on a sufficiently small time-scale.

In recent years, several concepts have emerged to provide more flexibility in networks through virtualization and control-plane programmability [1]. The split between data plane and control plane, proposed by software defined networking (SDN) [2], is regarded as the basic concept to allow flexible network adaptation. In addition, network virtualization (NV) [3] allows the sharing of physical network resources by different, independent networks. Further, the softwarization of network functions via network function virtualization (NFV) [4], replacing previously used middle-boxes, enables dynamic adaptations that add to flexibility. Altogether, these paradigms act as an enabler towards more flexible network operation.

On the other hand, legacy network implementations, too, are to some extent flexible. They are able to dynamically adapt their system states in reaction to changing requirements, e.g., by using appropriate protocols. Although increased flexibility

is often claimed in context of SDN, NV and NFV, a formal argument for support is mostly missing in the respective literature. Instead, flexibility is claimed from an intuitive notion of adaptivity and neglects potential trade-offs that might be induced by adding more complexity into already complex networking systems. To formally analyze the impact of flexibility on communication networks, to strengthen arguments for and against it and enable meaningful trade-off analysis, flexibility needs to be quantified in a formally clean fashion. Clearly, when such a quantification is possible, this enables follow-up arguments: (i) it allows the relative comparison of systems and their ordering in the sense of being more or less flexible; (ii) it allows establishment of flexibility scaling-laws, similar to algorithmic complexity discussions; (iii) it allows explicit design of systems for increased flexibility; (iv) and most importantly, it allows a tractable trade-off analysis with respect to introduced cost, time and system complexity.

In this work we propose a metric to capture flexibility in a quantitative fashion, instead of by intuitive arguments. We argue that mathematically, flexibility can be cast to the “size” of an appropriately defined set of achievable demand changes. Consequently, we propose the use of mathematical measure functions to capture it. We show that the use of such measures complies with common intuition of the word “flexible”. We further develop an empirical measurement procedure that can serve to compare two systems with respect to flexibility.

The rest of the paper is organized as follows. In Section II we enumerate basic research results on flexibility. The main contributions of the paper are introduced in Section III and Section IV, which contain the theoretical and empirical results on the measure, respectively. Our numerical evaluations of concrete use cases are presented in Section V. Finally, the possible extensions of the proposed metric is discussed in Section VI and then the paper is concluded in Section VII.

## II. RELATED WORK

Flexibility has emerged recently as a core target for communication network designs and is explicitly or implicitly touched by many works on SDN, NV and NFV, but even the understanding of the word “flexible” strongly differs among different papers. The works that directly discuss or target flexibility itself, and that we are aware of, can be reduced to [5]–[8]. In [5], [6] the authors discuss the flexibility of traffic engineering solutions in data centers, arguing against the assumption that wired connections are inflexible. To support

their argument, the used flexibility metric is the throughput performance for increasing traffic, compared to a “throughput proportional” behavior. Works [7], [8] propose to measure flexibility by using an acceptance ratio over a set of induced change requests. While this measure is compatible with the notion we develop here, the authors don’t introduce any formal argument of why it reflects the flexibility of a system.

While not commonly discussed in networking, flexibility analysis is a tool that has been used in other scientific contexts [9]–[20], such as manufacturing systems or management science, where an increased interest exists already for more than four decades. The main contributors seen here are works on flexible manufacturing systems (FMS) and decision theory. An FMS [9]–[14], [20] can be re-configured to match changing requirements, e.g., to change production volume, production flow or produced product. As shown in the survey [10], many works follow the naive approach of defining a flexibility function and demanding different properties from intuition, e.g., that it should increase with production volume but decrease with required production time. As argued in [10], each function falls into one out of five main streams, which are generalized in [12], however without resulting in a single, consistent metric. This is criticized in [14], as some of the introduced metrics are even shown to produce inconsistent relative orderings of more or less flexible manufacturing systems. Exceptions are the works [9], [13], which define flexibility more formally as the weighted efficiency of machines over a possible task set [9] and the distance that can be traversed by an FMS in a given state space [13].

In contrast to FMS, decision theory considers the impact of decision flexibility [15]–[19], which is consistently defined and treated throughout the literature. Here, the focus is on companies that need to make decisions, influencing an unknown or only partly known future. Decision flexibility is then defined either as the amount of future options enabled by a current decision, or their revenue to the decision maker.

Combining both views of FMS and decision theory, *flexibility is in general related to an option set*, i.e., a set of possible tasks, achievable system states or available decisions. Further, it is often extended towards the *quality of available options*, e.g., in terms of machine efficiency or decision reward.

#### A. Contribution of this Work

After discussing the related literature on flexibility, the gap which we try to fill is obvious: while other fields have several decades of history on formal flexibility investigations, comparable approaches for communication networks can hardly be found – in a time where flexibility is one of the key promises given by network researchers. We aim at starting the formal analysis of network flexibility by proposing a corresponding metric, making it measurable. Obviously, there is no such thing as “the” definition of flexibility, such that there also is no single way to assess it. However, we aim at providing a clean flexibility notion that (i) is applicable to communication networks, (ii) is consistent with the intuitive usage of the term,

that (iii) does not lead to mathematical inconsistencies and (iv) can be assessed analytically as well as empirically.

To be specific, we target flexibility as size of a feasible demand change set, which is in line with the definition of decision flexibility and major parts of the understanding gained for FMS [9], [13]. In contrast to most existing works, we do not directly incorporate the utility of flexibility into the metric but rather try to capture the essence of flexibility itself. However, we discuss in Section VI how the utility of flexibility can be evaluated. Furthermore, we propose a measurement procedure for empirical assessment of flexibility, which is consistent with the analytical framework.

### III. NETWORK FLEXIBILITY

Just as most high-level terms, such as “creativity”, “intelligence” or “fairness”, the meaning of the term flexibility is not easily defined in a clear fashion, nor can any short definition capture the full meaning of the term. However, to specify what a flexibility analysis of communication networks would be about, we must specify what we refer to with the term *in this context*. From our literature review, we realize that the main features of interest for communication networks are threefold, referring to (i) variety of adaptation possibilities, (ii) speed of adaptation and (iii) overhead/cost of adaptation. Without claiming completeness of this definition, we refer to flexibility in the following fashion:

*Definition 1: Network flexibility is the ability of a communication network to adapt its state to changing conditions promptly and with little effort.*

A state here can be, e.g., the routes used by communication flows or network resource usage patterns. The effort can be related to any cost metric, such as overhead, complexity or monetary cost, and the conditions can be externally given or actively demanded. In an abstract sense, different states can be realized by a given system implementation, which is bound to specific protocols, hardware and software modules. A change of state can then happen on different time-scales and with differing effort and lead or not lead to a proper adaptation to changing conditions. It is this interaction of changing conditions, state change, speed of change and change effort that impact our view on flexibility.

#### A. System Model

We consider a communication system that can be described by a *system state*  $\mathbf{S} \in \mathcal{S}$ , where  $\mathcal{S}$  contains all possible states that the system can realize. The state can reflect, e.g., the routes used by communication flows, end-to-end connectivity patterns, network resource usage patterns or the positions of network functions or controllers. We assume it to be given in an appropriate descriptive form, which can be a set, tuple, matrix or vector but is in principle not limited to such.

Furthermore, we define a *demand set*  $\Omega$  that captures demands posed to the network. Demands are requirements on the network state and can be used to model, e.g., connection requests, rate or Quality of Service (QoS) demands. Each demand  $d_i \in \Omega$  is associated with a set of valid states

$\mathcal{V}(d_i) \subseteq \mathcal{S}$ , in which the demand is satisfied. Over time, the demands will vary and the system will adapt its state in order to satisfy them. A *demand change* is an event denoted by the tuple of initial demand  $d_i$  and new demand  $d_j$ , i.e.,  $d_{i,j} = (d_i, d_j)$  for  $d_i, d_j \in \Omega$ . Each change  $d_{i,j}$  demands that the system is adapted from  $\mathbf{S}_i \mapsto \mathbf{S}_j$ , where  $\mathbf{S}_i \in \mathcal{V}(d_i)$ ,  $\mathbf{S}_j \in \mathcal{V}(d_j)$ , respectively<sup>1</sup>.

Assume a *system implementation*  $X \in \mathcal{X}$ , where  $\mathcal{X}$  is the set of possible implementations, which is bound to specific algorithms, protocols, hardware and software modules. Due to its nature,  $X$  can realize any system state out of a set  $\mathcal{S}_X \subseteq \mathcal{S}$ . Consequently,  $\forall d_i \in \Omega$  there is a set of valid states,  $\mathcal{V}_X(d_i) = \mathcal{S}_X \cap \mathcal{V}(d_i)$ , that can be achieved by  $X$ .

An example set-up is shown in Fig. 1, which consists of a network with connection requests among the nodes. Here, the demands  $d_i$  are the connection requests and the valid system states  $\mathcal{V}(d_i)$  for demand  $d_i$  are the possible realizations (i.e., routings) of the connection. In Fig. 1 the given implementation  $X$  realizes each connection according to a shortest path routing algorithm and hence will deterministically select a realization with the minimum number of hops from  $\mathcal{V}(d_i)$  for every demand. As result,  $\mathcal{V}_X(d_i)$  contains the single system state corresponding to the chosen shortest path. A demand change  $d_{1,4} = (d_1, d_4)$  will induce a system state change from  $\mathbf{S}_1 \mapsto \mathbf{S}_4$ , the realization of which requires a certain amount of time and comes at a certain cost. In general, the time required for the system to react is described by the *reaction time*  $\tau_X : \mathcal{S}_X \times \mathcal{S}_X \mapsto \mathbb{R}_+$ , which maps each possible state change of an implementation  $X$  to its appropriate time value (e.g., time elapsed until the new flow is established in Fig. 1). Furthermore, each demand change is associated with an *adaptation cost*, which is described by the mapping  $c_X : \mathcal{S}_X \times \mathcal{S}_X \mapsto \mathbb{R}_+$  and reflects the effort of realizing the state change (e.g., the number of forwarding table changes).

For the sake of simplicity and readability, throughout this paper we restrict ourselves to the simple model introduced above. In particular, we consider the case that  $\mathcal{V}_X(d_i) \in \mathcal{S}_X$  is either empty or a single element, i.e., the demands are formulated in a way that only a single state will satisfy it. This applies, e.g., to systems where the demand itself is a system state or to systems that adapt themselves towards a unique optimum state, as in the given example. In this model  $\tau_X$  and  $c_X$  become deterministic, memoryless mappings on  $\Omega \times \Omega \mapsto \mathbb{R}_+$ , respectively. As we will discuss in Section VI, the introduced framework can be extended towards more general valid sets  $\mathcal{V}_X(d_i)$  and system behaviors.

### B. Formal Flexibility Definition

A quantifier for flexibility should exhibit some properties to match the way we use the terminology. As we want to be able to sort systems and implementations with respect to flexibility, its quantifier should allow an ordering (i.e., a “ $\leq$ ” operation), in which more flexible system should achieve a

<sup>1</sup>A demand change (e.g., in flow rate) does not necessarily induce a state change if the actual realization (e.g., flow routing) is in both  $\mathcal{V}(d_i)$  and  $\mathcal{V}(d_j)$ .

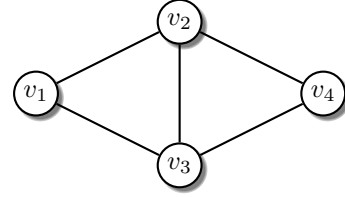


Fig. 1. Network with demand set  $\Omega = \{d_1 = (v_1, v_2), d_2 = (v_1, v_3), d_3 = (v_1, v_4), d_4 = (v_2, v_3), d_5 = (v_2, v_4), d_6 = (v_3, v_4)\}$ . Valid system states for  $d_1$  and  $d_4$  are  $\mathcal{V}(d_1) = \{v_1 - v_2, v_1 - v_3 - v_2, v_1 - v_3 - v_4 - v_2\}$  and  $\mathcal{V}(d_4) = \{v_2 - v_3, v_2 - v_1 - v_3, v_1 - v_3 - v_4 - v_2\}$ . The subset of valid states a shortest path routing algorithm  $X$  can realize are  $\mathcal{V}_X(d_1) = \{\mathbf{S}_1\} = \{v_1 - v_2\}$  for  $d_1$  and  $\mathcal{V}_X(d_4) = \{\mathbf{S}_4\} = \{v_2 - v_3\}$  for  $d_4$ .

larger flexibility value. Furthermore, it should be possible to have a “totally inflexible” system, which intuitively should be assigned value zero. Because such an inflexible system is less flexible than any other system, flexibility should have non-negative values.

From our previous discussion, flexibility is related to the amount of demand changes that a system can support, to the time scale at which it can serve a demand and to the effort associated with it. Using the system model described in Section III-A, we define the set of *achievable demand changes* by the considered system implementation  $X$  under given reaction time constraint  $T$  and cost constraint  $C$  as:

$$\mathcal{A}_X(T, C) = \{d_{i,j} \in \Omega \times \Omega : i \neq j; \mathcal{V}_X(d_i), \mathcal{V}_X(d_j) \neq \emptyset; \tau_X(d_{i,j}) \leq T; c_X(d_{i,j}) \leq C\}.$$

The first line only ensures that  $d_{i,j}$  is a valid demand change, i.e., it can be performed by implementation  $X$ . Next, we assess the flexibility of a system implementation as the “size” of the set<sup>2</sup>  $\mathcal{A}_X(T, C)$ .

*Definition 2:* Given a system implementation  $X$  with the set of achievable demand changes  $\mathcal{A}_X(T, C)$  with respect to time and cost constraints  $T$  and  $C$ . The flexibility of  $X$  is defined as  $\mu(\mathcal{A}_X(T, C))$ , where  $\mu$  is an appropriate measure on  $\Omega \times \Omega$ .

### C. Implications

The size of achievable demand changes indeed meets our intuitions about flexibility. This we show with the following observations, which follow directly from Definition 2 and the properties of mathematical measures:

*Observation 1:* Implementation  $X$  that cannot react to any demand change has zero flexibility (i.e., is inflexible), as  $\mathcal{A}_X(T, C) = \emptyset$  and hence  $\mu(\mathcal{A}_X(T, C)) = 0$ .

The strictness of Observation 1, demanding that the system cannot react to any demand change, in fact implies that only few systems can be claimed to be completely inflexible.

*Observation 2:* Implementation  $X$  is more flexible than  $Y$  if it can react to more demand changes under time and cost constraints  $T, C$ , indicated by  $\mu(\mathcal{A}_Y(T, C)) \leq \mu(\mathcal{A}_X(T, C))$ . Trivially, this is the case when  $\mathcal{A}_Y(T, C) \subseteq \mathcal{A}_X(T, C)$ .

<sup>2</sup>Although for set size usually operator  $|\cdot|$  is used, we will use set function  $\mu(\cdot)$  instead to emphasize the variety of measures which can be applied.

*Observation 3:* If implementation  $X$  and  $Y$  can realize different demand changes, an implementation  $Z = X \cup Y$  can be constructed, that selects among  $X$  and  $Y$  the one that can realize a given demand change within the constraints, with ties broken arbitrarily. It holds  $\forall T, C$  that  $\mu(\mathcal{A}_Z(T, C)) \geq \max\{\mu(\mathcal{A}_X(T, C)), \mu(\mathcal{A}_Y(T, C))\}$ .

*Proof:* This follows from  $\mathcal{A}_Z(T, C) = \mathcal{A}_X(T, C) \cup \mathcal{A}_Y(T, C)$  which induces  $\mu(\mathcal{A}_Z(T, C)) = \mu(\mathcal{A}_X(T, C)) + \mu(\mathcal{A}_Y(T, C)) - \mu(\mathcal{A}_X(T, C) \cap \mathcal{A}_Y(T, C))$  [21] and hence  $\mu(\mathcal{A}_Z(T, C)) \geq \max\{\mu(\mathcal{A}_X(T, C)), \mu(\mathcal{A}_Y(T, C))\}$ . ■

The arguments given in Observations 2 and 3 in fact reflect the perspective that is found often in literature when flexibility is claimed: By enabling re-configuration of the network, its functions, flows or similar, the resulting system is one that can do anything it could do before, and more. Thus, it must be more flexible. From a formal perspective, this argument is only partly applicable to real systems, as it assumes that the reconfiguration itself does not induce any substantially increased delay or cost. As soon as the reconfiguration induces a non-negligible delay or cost overhead, there are time or cost constraints under which the re-configuration itself is a drawback that can actually make the network *less* flexible. When network operation requires reaction times or cost efficiency corresponding to these tight constraints, the impact of this decrease needs to be investigated, otherwise increased flexibility can be an empty promise.

While the evaluation over a “size” of achievable demand changes has an intuitive relation to flexibility, the introduction of time and cost constraints is at first a little counter-intuitive. However, their use is in accordance with the multi-dimensionality of flexibility that has been observed already in the discussed literature. Formally, we establish the “dimensions” *adaptivity*, *reactivity* and *cost-efficiency*: Consider two system implementations  $X, Y$  with associated sets  $\mathcal{A}_X(T, C)$  and  $\mathcal{A}_Y(T, C)$ , respectively<sup>3</sup>. Then, we define the following:

*Definition 3:* We say that implementation  $X$  is at least as *reactive* as  $Y$  if

$$\forall T : \mu(\mathcal{A}_Y(T, \infty)) \leq \mu(\mathcal{A}_X(T, \infty)), \quad (1)$$

that  $X$  is at least as *cost-efficient* as  $Y$  if

$$\forall C : \mu(\mathcal{A}_Y(\infty, C)) \leq \mu(\mathcal{A}_X(\infty, C)), \quad (2)$$

and that  $X$  is at least as *adaptive* as  $Y$  if:

$$\mu(\mathcal{A}_Y(\infty, \infty)) \leq \mu(\mathcal{A}_X(\infty, \infty)). \quad (3)$$

The reactivity property states that, disregarding cost, implementation  $X$  can react to demand changes at least as fast as implementation  $Y$ . Analogy holds for cost-efficiency with respect to cost. The notions of reactivity and cost-efficiency can be interpreted as Pareto-superiority in the time and cost dimensions. Finally, adaptivity states that, independent of time and cost, implementation  $X$  can react to at least as many demand changes as implementation  $Y$ . One can argue that  $X$  is more flexible than  $Y$  if it is better in any of these

properties. Hence, each constraint combination  $(T, C)$  of the flexibility evaluation can be interpreted as an emphasis on the adaptivity, reactivity and cost-efficiency property, with more stringent constraints enforcing better reactivity and cost-efficiency. Therefore, in the rest of the paper, we will focus only on the general term flexibility, but keep in mind that the findings are applicable to all of these dimensions with an appropriate constraint selection. Furthermore, for ease of notation, we will drop inputs  $X, T, C$  when they do not explicitly contribute to understanding and write  $\mu(\mathcal{A})$  instead of  $\mu(\mathcal{A}_X(T, C))$ .

#### D. Flexibility Measures

We have motivated that flexibility can be quantified by a measure  $\mu$  on  $\mathcal{A}_X(T, C)$ , which reflects the size of the set of achievable demand changes. However, with this definition flexibility is by no means a unique metric. For example, the use of different measure types, such as the Dirac measure versus the counting measure, will result in different flexibility values and even might lead to different relative orderings of systems. A special case is that if  $\mu$  is a valid measure, then  $\mu'(\mathcal{A}) = c\mu(\mathcal{A})$  for any constant  $c > 0$  fulfills all properties of a measure, too [21]. In this case, although  $\mu'$  and  $\mu$  will have different absolute values, they will induce the same ordering of implementations with respect to flexibility and hence can be considered equivalent. Therefore, in order to eliminate at least this scaling effect, we can normalize the measure with respect to an arbitrary base set  $\mathcal{B} \subseteq \Omega \times \Omega$  that satisfies  $0 < \mu(\mathcal{B}) < \infty$ , by defining:

$$\mu_{\mathcal{B}}(\mathcal{A}) := \frac{1}{\mu(\mathcal{B})} \mu(\mathcal{A}) \quad (4)$$

It is easily seen that any scaling factor cancels out for  $\mu_{\mathcal{B}}$ :

$$\mu'_{\mathcal{B}}(\mathcal{A}) = \frac{1}{c\mu(\mathcal{B})} c\mu(\mathcal{A}) = \frac{1}{\mu(\mathcal{B})} \mu(\mathcal{A}) = \mu_{\mathcal{B}}(\mathcal{A}). \quad (5)$$

Different choices for  $\mathcal{B}$  can in general lead to different intuitive interpretations. The direct “amount” of achievable demands in  $\mathcal{A}_X(T, C)$  can be represented as the choice of a set with  $\mu(\mathcal{B}) = 1$ , and then has the following intuition:

$$\mu(\mathcal{A}) = \mu_{\mathcal{B}}(\mathcal{A}) = \# \text{ achievable demand changes.} \quad (6)$$

Although we consider Eq. (6) as a measure for flexibility, the outcome will be a number between  $[0, \mu(\Omega^2)]$ , which is a rather arbitrary output.

Another option is the normalization by the maximum number of achievable demand changes over all possible implementations  $X \in \mathcal{X}$ , i.e.,  $\mathcal{B} := \mathcal{A}^*$ , where  $\mathcal{A}^* = \arg \max_{\mathcal{X}} \mu(\mathcal{A}_X(T, C))$ . This measure has the intuitive meaning of *flexibility degree*  $\varphi(\mathcal{A}) := \mu_{\mathcal{A}^*}(\mathcal{A})$ , which is

$$\varphi(\mathcal{A}) = \frac{\mu(\mathcal{A})}{\mu(\mathcal{A}^*)} = \frac{\# \text{ achievable changes by } X}{\# \text{ max. achievable changes over } \mathcal{X}}.$$

Obviously  $\varphi(\mathcal{A}) \in [0, 1]$ , which motivates the term degree. A flexibility degree of one corresponds to a system being “100%”, i.e., maximum flexible. However, we note that it is not easy to find  $\mu(\mathcal{A}^*)$  in general.

<sup>3</sup>We denote with “ $\infty$ ” if a constraint is infinitely relaxed on that parameter.

#### IV. EMPIRICAL FLEXIBILITY EVALUATION

By the introduced theoretical framework in Section III, measuring the flexibility of system implementation  $X$  requires the identification of set  $\mathcal{A}_X(T, C)$  – i.e., the set of possible demand changes under time constraint  $T$  and cost constraint  $C$ . In this section we discuss how to empirically measure and compare the flexibility of different system implementations.

##### A. Observed Flexibility

An intuitive way to empirically observe the measure of  $\mathcal{A}_X(T, C)$  for given  $(T, C)$  constraints is introduced in the following. Given a system under test whose flexibility shall be estimated. Furthermore, we can identify a set of demand changes  $\mathcal{D} \subseteq \Omega \times \Omega$  with  $\mu(\mathcal{D}) < \infty$  such that  $\mathcal{A}_X(T, C) \subseteq \mathcal{D}$  holds. We select an infinite length *challenge sequence*  $\mathbf{C} = \{d_{i_1, j_1}, d_{i_2, j_2}, d_{i_3, j_3}, \dots\}$ , which may contain arbitrary demand changes from  $\mathcal{D}$ . We argue that challenging the system with this sequence and observing its reaction for each demand change, the flexibility measure can be evaluated. First, for each demand change  $d_{i,j} \in \mathcal{D}$ , we assume that it occurs in  $\mathbf{C}$  with a relative frequency

$$\nu(d_{i,j}) = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \mathbb{1}\{d_{i,j} = d_{i_k, j_k}\}. \quad (7)$$

$\mathbb{1}\{\cdot\}$  therein is the indicator function, which is one if the logical statement is true and zero otherwise. We refer to  $\nu$  as the *challenge profile*, which is imposed onto the system under test by sequence  $\mathbf{C}$ . By counting the number of changes that are in  $\mathcal{A}$ , we measure:

$$\mu^\nu(\mathcal{A}) = \int_{d_{i,j} \in \mathcal{D}} \nu(d_{i,j}) \mathbb{1}\{d_{i,j} \in \mathcal{A}\} d\mu = \int_{d_{i,j} \in \mathcal{A}} \nu(d_{i,j}) d\mu,$$

where the second equality holds because  $\mathcal{A} \subseteq \mathcal{D}$  and the integral  $\int \cdot d\mu$  denotes Lebesgue integration, which is an abstraction of integration towards arbitrary measurable set systems [21]. Intuitively, the system under test reacts to challenge profile  $\nu$  and exhibits a certain degree of flexibility, referred to as *observed flexibility*. Now, because  $\nu$  has all properties of a normalized density with cumulative measure of  $\int_{\mathcal{D}} \nu(d_{i,j}) d\mu = 1$ , it can be interpreted as a probability density function, reflecting the probability of a demand change being challenged by the sequence  $\mathbf{C}$ . This leads to

$$\begin{aligned} \mu^\nu(\mathcal{A}) &= \int_{d_{i,j} \in \mathcal{D}} \nu(d_{i,j}) \mathbb{1}\{d_{i,j} \in \mathcal{A}\} d\mu \\ &= \mathbb{E}\{\mathbb{1}\{d_{i,j} \in \mathcal{A}\} \mid \mathbf{C}\} = \Pr\{d_{i,j} \in \mathcal{A} \mid \mathbf{C}\}. \end{aligned} \quad (8)$$

$\mathbb{E}\{\cdot\}$  therein is the expectation value and  $\Pr\{\cdot\}$  the probability of an event occurring. That is, the observed flexibility corresponds to the expectation of the challenges contained in  $\mathbf{C}$  being in  $\mathcal{A}$  under the specific challenge profile  $\nu$ . From the definition of the set  $\mathcal{A}$ , the expectation is nothing but the probability that the system can react to the challenged demand changes within the target time and cost constraints  $(T, C)$ .

##### B. Uniform and Non-Uniform Challenge Profiles

A design choice for estimating flexibility is the sequence  $\mathbf{C}$ . In particular, if it can be chosen such that the challenge profile is uniform, i.e.,  $\forall d_{i,j} \in \mathcal{D} : \nu(d_{i,j}) = 1/\mu(\mathcal{D})$ , then the observed flexibility is:

$$\begin{aligned} \mu^\nu(\mathcal{A}) &= \int_{d_{i,j} \in \mathcal{D}} \nu(d_{i,j}) \mathbb{1}\{d_{i,j} \in \mathcal{A}\} d\mu \quad (9) \\ &= \frac{1}{\mu(\mathcal{D})} \int_{d_{i,j} \in \mathcal{A}} 1 d\mu = \frac{\mu(\mathcal{A})}{\mu(\mathcal{D})} = \mu_{\mathcal{D}}(\mathcal{A}), \end{aligned} \quad (10)$$

which is proportional to  $\mu(\mathcal{A})$ . Hence, we get back to a normalized version of the exact flexibility defined in Section III-D, with the selection of  $\mathcal{B} = \mathcal{D}$ . As we have argued,  $\mu_{\mathcal{D}}(\mathcal{A})$  is equivalent to  $\mu(\mathcal{A})$  in that it induces the same relative ordering.

On the other hand, if  $\nu$  is not uniform, we will have a weighted flexibility value with more emphasis on the challenges that occur more often. This better matches observed system behavior in a running environment, as the system might be challenged with the same demand changes (e.g., in flow demands, connection requests, virtual network requests, etc.) multiple times, while others might not be requested at all. Although such flexibility value will be distorted with respect to our original definition, we can claim it to be the flexibility that the system exhibits when it has to respond only to a certain challenge profile.

##### C. Estimating Flexibility

In practice, the challenge sequence will be finite, which further distorts the resulting values because the targeted relative frequencies might not be matched precisely. However, the empirical flexibility evaluation boils down to estimating the event probability from (8). Consider a challenge sequence  $\hat{\mathbf{C}}$  of finite length  $N$ , with elements randomly chosen out of  $\mathcal{D}$  with uniform distribution. By using the empiric mean as estimator for the expectation, an estimate for  $\mu_{\mathcal{D}}(\mathcal{A})$  can be created, which is

$$\hat{\mu}_{\mathcal{D}}(\mathcal{A}) = \frac{\hat{\mu}(\mathcal{A})}{\hat{\mu}(\mathcal{D})} = \frac{\sum_{k=1}^N \mathbb{1}\{d_{i_k, j_k} \in \mathcal{A}\}}{N}. \quad (11)$$

The intuitive meaning of this estimated flexibility can be reduced to

$$\hat{\mu}_{\mathcal{D}}(\mathcal{A}) = \frac{\# \text{ of supported challenges}}{\# \text{ of posed challenges}}. \quad (12)$$

Indeed, equation (12) takes exactly the form proposed in [7], [8], such that we are able to re-motivate it. The given results lead to an overall estimation flow of flexibility as is given in Algorithm 1. Due to the behavior of the empiric mean, the estimate becomes arbitrarily precise for  $N \rightarrow \infty$ :

$$\lim_{N \rightarrow \infty} \hat{\mu}_{\mathcal{D}}(\mathcal{A}) = \mathbb{E}\{\mathbb{1}\{d_{i,j} \in \mathcal{A}\}\} = \mu_{\mathcal{D}}(\mathcal{A}). \quad (13)$$

A measure defined this way will always be out of the interval  $[0, 1]$ , which follows the intuition of the flexibility degree

---

**Algorithm 1** Estimate Flexibility  $\hat{\mu}_{\mathcal{D}}(\mathcal{A})$ 

---

- 1: Input:  $X, \Omega, T, C, \mathcal{D}, N$
  - 2: Variables:  $\Sigma_{\mathcal{A}} := 0$
  - 3: Create sequence  $\hat{\mathbf{C}}$  with  $N$  elements out of  $\mathcal{D}$
  - 4: **for**  $k \in \{1, \dots, N\}$  **do**
  - 5:   Impose demand change  $d_{i_k, j_k}$
  - 6:   Observe  $\delta_X := \begin{cases} 1, & \text{if } d_{i_k, j_k} \text{ is realized} \\ 0, & \text{else} \end{cases}$
  - 7:   Measure  $\tau_X(d_{i_k, j_k}), c_X(d_{i_k, j_k})$
  - 8:    $\Delta := \mathbb{1}\{\delta_X = 1 \wedge \tau_X(d_{i_k, j_k}) \leq T \wedge c_X(d_{i_k, j_k}) \leq C\}$
  - 9:    $\Sigma_{\mathcal{A}} := \Sigma_{\mathcal{A}} + \Delta$
  - 10: **end for**
  - 11: Output  $\hat{\mu}_{\mathcal{D}}(\mathcal{A}) := \Sigma_{\mathcal{A}}/N$
- 

defined in Section III-D. However, in this case it is not guaranteed that 100% flexibility is reachable, because  $\mathcal{D}$  might contain non-achievable demand changes in contrast with  $\mathcal{A}^*$ .

## V. EXPERIMENTAL RESULTS

### A. Dynamic Controller Placement in SDN

Consider an SDN network described by a graph  $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ , that is managed by  $1, 2, \dots, |\mathcal{N}|$  controllers, which program the network switches in a centralized fashion to serve the flow demands in the network. As SDN decouples control plane and data plane, such a centralized network management is claimed flexible in literature. However, the control also introduces additional latency, i.e., a flow setup time, whenever the forwarding path of a new flow needs to be configured [7]. Our evaluation target thus is the impact of this latency.

We model this set-up as follows. The network state is given by the current paths on which the flows are embedded, as well as the positions of the controllers in the network. We assume that the controller positions are optimized towards a minimum average flow set-up time at each instant. So when the flows are changing, the network might react with a controller migration. A demand  $d_i \in \Omega$  represents a set of flows, specified by a set of triplets  $(s, t, r)$ , each indicating source  $s \in \mathcal{N}$ , target  $t \in \mathcal{N}$  and the requested number of flows  $r \in \mathbb{N}_+$  between  $s$  and  $t$ . The set of all possible new flows is defined as  $\Omega = \mathcal{N} \times \mathcal{N} \times \mathbb{N}_+$ . Assume a time-slotted network operation, such that in each slot a set of new flows  $d_i$  are in the network and their forwarding paths need to be set up by the control plane. Because the network optimizes the controller positions, the associated state  $\mathcal{V}_X(d_i)$  is the optimal controller placement that minimizes the average flow setup time given the demand  $d_i$ , as well as the implemented paths. Because the used paths and controller positions are uniquely defined,  $\mathcal{V}_X(d_i)$  contains a single element  $\forall d_i \in \Omega$ . We can model different numbers of controllers as different system implementations  $X_n$  and realize that  $\mathcal{V}_{X_n}(d_i)$  of different system implementations are completely different, respectively. However, they can be compared due to the same demand sets.

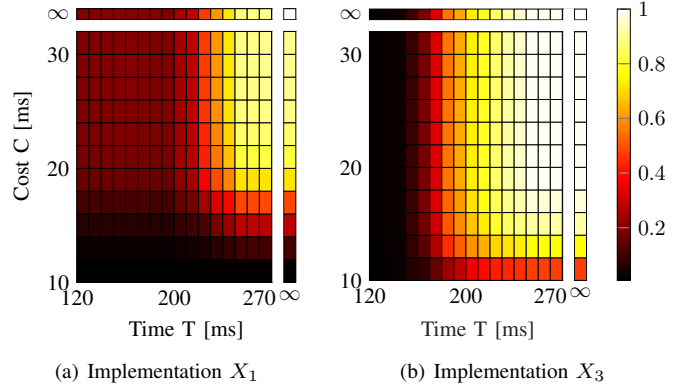


Fig. 2. Flexibility comparison of different system implementations (1 or 3 controllers) with varying constraints  $T$  and  $C$ . Each square represents the average flexibility of 50 simulation runs on 100 long challenge sequences.

To react to a demand change  $d_{i,j}$ , the control plane first needs to find the optimal placement  $\mathcal{V}_X(d_j)$  that induces a certain cost  $c_X(d_{i,j}) = c_X(d_j)$ , defined as the optimal average flow setup time with  $\mathcal{V}_X(d_j)$ . If the current optimal placement  $\mathcal{V}_X(d_j)$  is different from the previous optimal placement  $\mathcal{V}_X(d_i)$ , a certain delay  $\tau_X(d_{i,j})$  is induced to record the control plane adaptation time [8]. The adaptation is counted as a success if and only if  $c_X(d_{i,j}) < C$  and  $\tau_X(d_{i,j}) < T$ .

The numerical results of this use case are generated using the flow of Algorithm 1 on the Abilene network topology. Intuitively, one would expect that adding more controllers introduces higher flexibility in adapting different sets of new flows. Fig. 2 shows the heatmap of the flexibility value with different time and cost constraint values.  $X_1$  and  $X_3$  represent the system implementation with 1 controller and 3 controllers respectively. In general, the flexibility value of each system implementation increases when  $T$  or  $C$  increases, meaning that relaxed constraints allow more adaptation successes. Besides,  $X_3$  is more flexible for most of the  $C$  and  $T$  constraint combinations. A counter-intuitive observation, however, is that for tight  $T$  and relaxed  $C$ ,  $X_1$  is more flexible than  $X_3$  (upper-left corner of the heatmap). This is because in this region, the optimal placement of  $X_1$  does not vary much but a fixed position is optimal for a majority of the demands, whereas that of  $X_3$  changes more frequently and induces adaptation time that violates the time constraint. This effect corresponds to the reactivity and cost-efficiency perspective, which are given in the lines indicated with  $\infty$ . Comparing these, we can claim that  $X_3$  is more cost-efficient than  $X_1$ . However, for low time constraints  $X_1$  is more reactive, as it does not invest the time to re-optimize its state. Finally, both systems are equally adaptive, as is seen in the upper right corner, as both can, in principle, serve any of the challenged demands.

### B. Protection Routing

In this use case we consider routing under use of different protection approaches. Each flow is routed from source to target node and in addition, the flow should be protected against single-link failures. The protection is implemented in

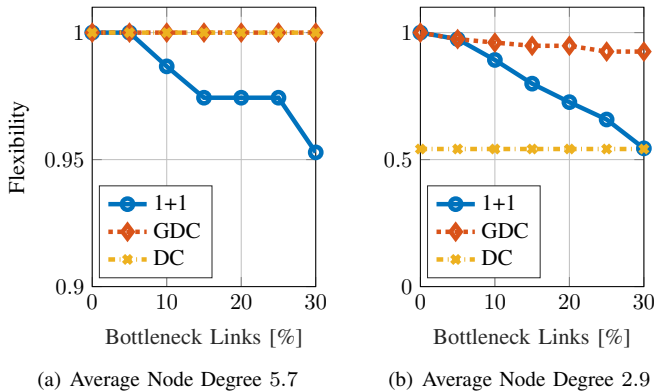


Fig. 3. 80 node networks with increasing number of bottleneck links.

a network with running flows already embedded, such that the network of free capacity, which can be considered as available network topology on which protection should be established, is varying. Flexibility is evaluated as the ability to provide protection under different underlying capacities. In particular, we consider three different protection schemes [22], namely:

- $1+1$  dedicated path protection, which requires a disjoint path-pair between the source  $s_i$  and target node  $t_i$ . The whole user data, split into two parts  $A$  and  $B$  and requiring two capacity units, is sent along both paths.
- *Diversity coding (DC)*, requires three disjoint paths between the source and target node. On each path, only a single capacity unit is required, as data halves  $A$ ,  $B$  and redundancy data  $A \text{ XOR } B$  is sent along the paths.
- *Generalized diversity coding (GDC)* [22], which combines the previous two approaches in order to circumvent their weaknesses, namely the bottleneck links with one unit of free capacity for  $1+1$  (it can use only links with 2 capacity units), and the topological limit of DC (lack of three disjoint paths), respectively.

We can formally argue that GDC is the most adaptive approach from the three according to Observation 3, as it contains both  $1+1$  and DC as special cases owing to its adaptable protection structure. In order to show this quantitatively, in our experimental evaluation we took a snapshot from the considered topology, and represent the state of the network as the free capacities on the links and the currently embedded protection paths. A demand is given as a source target pair,  $d_i = (s_i, t_i)$  and the case of no demand is denoted as  $d_0$ , respectively. The demand sequence  $\hat{C} = \{d_{0,1}, d_{0,2}, \dots, d_{0,n}\}$  contains single-shot demand changes. Time and cost constraints are relaxed to infinity here, as we are interested in the adaptivity perspective.

Fig. 3 contains our simulation results, which were obtained by increasing the number of bottleneck links on two topologies with different average node degrees. One can observe, that in the denser network with increased node degree in Fig. 3(a), where three disjoint paths exist between the node-pairs more often, diversity coding is equally as flexible as GDC. Furthermore, as the number of bottleneck links increases (i.e., with higher traffic load), the number of protectable node-

pairs for  $1+1$  decreases, making it the less flexible choice. On the other hand, in the sparser network in Fig. 3(b) only between the 54% of node-pairs exist three disjoint paths, which makes DC the less flexible choice up to 30% of bottleneck links. In the scenario with 30% bottleneck links the flexibility of  $1+1$  decreases to 54% as well. Hence, indeed in this limited situation GDC is 38% more flexible than any of its counterparts. In general, GDC is never less flexible than any of the other two schemes, which backs up our intuition.

### C. Traffic Flexibility in Data Centers

The introduction of flexibility enables a formal argument for and against flexibility of systems. This shall be demonstrated on the argumentative line of [6]. In this work, the authors argue in favor of static network topologies in data centers, which they claim to be in fact more flexible than reconfigurable topologies that use optical or radio links to bypass a fat-tree topology. The argument is that while reconfigurable topologies are obviously *more flexible in adapting the topology*, their performance in serving end-to-end traffic is not automatically better. For comparison, the authors choose different topologies to connect data centers apart from the fat tree and, on the other side, a re-configurable topology. For the latter, the number of ports is reduced, following the argument that both implementations should require equal cost  $C$ , whereas re-configurable ports in general cost more. The provided throughput of all interconnects is compared for skewed traffic matrices while scaling the traffic and it is shown with simulations that for low scaling factors, which is claimed to be the region of interest, the static topologies support more traffic. The authors conclude that static data center topologies can be *more flexible in terms of supporting throughput* than re-configurable ones.

We argue that our introduced flexibility metric conforms and backs up both claims. First, note that the dynamic implementation  $X$  can realize every possible topology demand changes which the static implementation  $Y$  can with less adaptation time and cost. Hence,  $\mathcal{A}_Y(T, C) \subseteq \mathcal{A}_X(T, C)$ , which is stated in Observation 2 and shows that the adaptive system is clearly more flexible. On the other hand, if the sequence  $\hat{C}$  is based on demands  $d_i$  that corresponds only to traffic matrices with low traffic, as these are in both  $\mathcal{V}_X(d_i)$  and  $\mathcal{V}_Y(d_i)$ , an adaptive strategy cannot serve these demands with lower time and cost than a static one. Hence, with such a sequence  $\hat{C}$  a static implementation can realize all demand changes that an adaptive can. Hence, it is considered to be more flexible in this respect, as stated in [6] as well. However, considering the whole spectrum of traffic matrices, investigating different setups or introducing strict timing or low budget constraints might give different order on the flexibility of these approaches. Therefore, in Section VI we discuss how these different aspects can be incorporated in our flexibility evaluation.

## VI. FURTHER DISCUSSION

### A. Non-Uniform Demand Utilities

In the flexibility evaluation discussed so far every demand change was considered equally important. However, the dif-

ference between flexibility of an implementation and its utility might be excessive when non-uniform utilities are evaluated. Consider a utility function where only two demand changes  $d_{i,j}$  and  $d_{j,i}$  are actually assigned a non-zero value. Hence, any two system implementation  $X$  and  $Y$  capable of reacting to these two changes will have the same utility. In contrast, their flexibility can very well differ if  $X$  is able to handle only these two demand changes while  $Y$  is flexible enough to react within constraints  $(T, C)$  to any possible demand. Thus, utility is independent of the flexibility property itself. In fact, there might be cases in which the utility perspective would encourage the use of less flexible implementations.

We will now show how the utility of different demands can be incorporated in our flexibility metric. Given a utility function  $u : \Omega \times \Omega \mapsto \mathbb{R}_+$  that reflects to how valuable the ability to change a certain demand from one to another is assumed. In many cases, the utility of a demand change is mostly impacted by the utility of its target demand, which can be appropriately modeled. Then, for given constraints  $(T, C)$ , we define the utility of flexibility (UoF) as:

$$\Xi(\mathcal{A}_X(T, C)) = \int_{d_{i,j} \in \mathcal{A}} u(d_{i,j}) d\mu. \quad (14)$$

Equation (14) reflects to the total utility enabled by allowing the switching between demands.

Estimation of utility can be established analog to estimation of flexibility, by designing a challenge sequence out of a set  $\mathcal{D} \subseteq \mathcal{A}$  with profile  $\nu$  and defining an observed utility

$$\begin{aligned} \Xi_\nu &= \int_{d_{i,j} \in \mathcal{D}} u(d_{i,j}) \nu(d_{i,j}) \mathbb{1}\{d_{i,j} \in \mathcal{A}\} d\mu \\ &= \mathbb{E}\{u(d_{i,j}) \mid d_{i,j} \in \mathcal{A}, \mathbf{C}\}. \end{aligned} \quad (15)$$

When  $\nu$  is uniform,  $\Xi_\nu = \Xi/\mu(\mathcal{D})$  holds, which again allows comparison among different implementations. An estimation process can then be created by replacing Line 9 in Algorithm 1 with  $\Sigma_{\mathcal{A}} := \Sigma_{\mathcal{A}} + u(d_{i_k, j_k}) \cdot \Delta$ . Finally, we close with the observation that the observed flexibility for non-uniform challenges from Section IV-A can be expressed as UoF if utility  $u(d_{i,j})$  is selected as the relative frequency  $\nu(d_{i,j})$  of occurrence the sequence  $\mathbf{C}$ .

### B. Joint Flexibility of Different System Set-Ups

When estimating the flexibility of a system under test, we typically want to consider the outcome of many specific set-ups in a Monte-Carlo fashion to ensure the validity of empirical analysis. However, the normalization set  $\mathcal{B}$  exhibits dependencies on the set-up, which have to be addressed accordingly.

Consider a set  $\mathcal{K}$  of system set-ups, in which we want to measure the flexibility of two system implementations  $X$  and  $Y$ , respectively. The different set-ups could correspond to different network graphs, on which two flow embedding algorithms are evaluated for their flexibility, for example. Intuitively, we want to evaluate the overall performance, e.g., by averaging flexibility over set-ups, summing them up or

considering the maximum and/or minimum achieved flexibility. For a consistent comparison over different system set-ups, it must be ensured that  $\forall k, k' \in \mathcal{K} : \mu(\mathcal{B}_k) = \mu(\mathcal{B}_{k'})$ , i.e., that the same normalization is used. This is intuitive at first glance but often needs to be explicitly ensured, e.g., if  $\mathcal{B}_k$  depends on the currently used set-up, such as the underlying topology. When the normalization sets differ, the used measures will in fact vary for different set-ups  $k, k'$ , which can lead to inconsistent orderings. Obviously, averaging over such different measures need not give any insightful results. To guarantee consistent result, an overall measure can be defined using  $\mu(\mathcal{B}_{\text{tot}}) = \sum_k \mu(\mathcal{B}_k)$ . As a result, flexibility can be evaluated with a consistent measure over all different set-ups, leading to the outcome of

$$\begin{aligned} \mu_{\mathcal{B}_{\text{tot}}}(\mathcal{A}) &= \sum_{k=1}^{|\mathcal{K}|} \frac{\mu(\mathcal{A}_k)}{\mu(\mathcal{B}_{\text{tot}})} = \sum_{k=1}^{|\mathcal{K}|} \frac{\mu(\mathcal{A}_k)}{\mu(\mathcal{B}_k)} \frac{\mu(\mathcal{B}_k)}{\mu(\mathcal{B}_{\text{tot}})}, \\ &= \sum_{k=1}^{|\mathcal{K}|} w_k \mu_{\mathcal{B}_k}(\mathcal{A}_k), \end{aligned} \quad (16)$$

$$\text{where } w_k = \frac{\mu(\mathcal{B}_k)}{\mu(\mathcal{B}_{\text{tot}})}; \quad \sum_k w_k = 1.$$

In words, the total measure is a weighted sum over the measures of each set-up, where the weights denote the relative sizes of the normalization sets  $\mathcal{B}_k$ .

### C. Probabilistic State Changes

We are aware that the given system assumptions under which we argue our main derivation – i.e., deterministic behavior of  $\tau_X$  and  $c_X$  and the restriction of valid sets  $\mathcal{V}_X(d_i)$  to a single state – are rather simple. However, extension towards more sophisticated system models is possible.

In order to capture a more general system, we need to relax the assumption that  $\mathcal{V}_X(d_i)$  contains only one state and further adapt the description of system reaction to demand change. We can describe the behavior of state evolution in a time-slotted fashion as  $\mathbf{S}[\tau]$ , where each slot  $\tau$  denotes a time interval  $[(\tau-1)T, \tau T]$  of width  $T$ , which is the chosen time constraint. Implementation  $X$  is then captured by a probabilistic mapping on the slot-to-slot behavior, i.e., by  $p_X : \mathcal{S}_X \times \Omega \mapsto \mathcal{F}_{\mathcal{S}_X}$ , where  $\mathcal{F}_{\mathcal{S}_X}$  is the set of probability densities on  $\mathcal{S}_X$ . In words, considering that the implementation  $X$  is currently in state  $\mathbf{S}_i \in \mathcal{S}_X$  and the current demand is  $d_i \in \Omega$ , the mapping  $p_x(\mathbf{S}_i, d_i)$  defines the probability for each state  $\mathbf{S}' \in \mathcal{S}_X$  that the system will adapt itself towards  $\mathbf{S}'$  within the time constraint  $T$  corresponding to one slot. This formulation is similar to a Markov-chain, however, the state transitions depend on  $d_i$  and hence are not automatically memoryless. If the system is designed to satisfy the demand  $d_i$ , it will adapt towards a state in  $\mathcal{V}_X(d_i)$  within a slot with high probability. If  $\mathbf{S} \in \mathcal{V}_X(d_i)$  already, the state might remain in  $\mathbf{S}$  or change itself towards other states in  $\mathcal{V}_X(d_i)$ . Also, if  $\mathcal{V}_X(d_i)$  is empty, i.e., the demand is infeasible to satisfy for implementation  $X$ , the system might stay in its current state or end up in another in the course of trying to satisfy the demand.



All of these opportunities can be captured by an appropriate mapping  $p_X$ .

Now, as  $\mathcal{V}_X(d_i)$  may contain more than one state, when a demand change  $d_{i,j}$  is considered it is not certain from which state  $\mathbf{S}_i \in \mathcal{S}$  the system will start adapting, nor at which state  $\mathbf{S}[\tau]$  it will be in slot  $\tau$  after the demand change. Further, the reaction time  $\tau_X$  and cost  $c_X$  are no longer deterministic but random variables. Finally, the system behavior is not memoryless with respect to a challenge sequence  $\mathbf{C} = \{d_{i_1,j_1}, d_{i_2,j_2}, \dots\}$ . This is because when the history before demand change  $d_{i_k,j_k}$  is changed, the starting state at the event of the next demand change might also change, leading to different time and cost values.

Even though the relaxation of assumptions makes the presented evaluation significantly more complex, the principle way of evaluating flexibility can be extended. First, because  $\tau_X, c_X$  become random variables, the indication of  $\tau_X \leq T, c_X \leq C$ , used for the definition of  $\mathcal{A}_X(T, C)$ , is not binary but rather relates to a probability. Consequently, the flexibility measure  $\mu(\mathcal{A}_X(T, C))$  needs to be extended to an expectation  $\mathbb{E}\{\mu(\mathcal{A}_X(T, C))\}$  of the respective measure function<sup>4</sup>. As for the history dependence of  $\mathbf{S}[\tau]$  on  $\mathbf{C}$ , we can introduce the assumption that there is an upper bound  $K$ , such that states before  $\tau - K$  have a negligible impact on the state probabilities in slot  $\tau$ . Using this, we can re-define the demands  $\Omega' := \Omega^K$  to be not a demand but a *demand sequence* of length  $K$ . The elements of  $\Omega'$  are then again independent of each other. By creating a challenge sequence  $\mathbf{C}' = \{d'_{i_1,j_1}, d'_{i_2,j_2}, \dots\}$  with  $\forall k: d'_{i_k,j_k} \in \Omega' \times \Omega'$  and defining  $\tau_X$  and  $c_X$  with respect to the time and cost of each  $d'_{i_k,j_k}$ , the derivation and Algorithm 1 can be applied. Finally, by using the probability perspective of Eq. (8), it can be shown that Algorithm 1 will then lead to an estimation of  $\mathbb{E}\{\mu(\mathcal{A}_X(T, C))\}$ .

## VII. CONCLUSIONS AND FUTURE WORK

In this work we derived the basic elements of evaluating flexibility in a network set-up. We defined what we consider as “flexibility” and argued that it is related to the size of the set of achievable demand changes that a system implementation can react to. We further showed that any mathematical measure function can, in general, serve as a flexibility metric. Building on these analytical results, a measurement flow for empirical assessment of flexibility was proposed. We applied the metric to the set-ups of path protection, flow embedding and traffic engineering in data centers and showed it to reflect our intuition on flexibility, as well as to enable concise argumentation.

In our measure we consider only the adaptation cost to demand changes. However, having a more flexible system might induce some additional equipment, operational and security costs, while the complexity of the algorithms might be considered as a pro-action cost as well. Such a cost analysis of flexibility will be done as a future work.

<sup>4</sup>The original measure can be extended to an appropriate probability space.

## ACKNOWLEDGMENTS

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 647158 - FlexNets “Quantifying Flexibility for Communication Networks”). The work of P. Babarzi was supported by the Post-Doctoral Research Fellowship of the Alexander von Humboldt Foundation.

## REFERENCES

- [1] M. He, A. M. Alba, A. Basta, A. Blenk, and W. Kellerer, “Flexibility in softwarized networks: Classifications and research challenges,” *IEEE Communications Surveys Tutorials*, pp. 1–39, 2019.
- [2] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “Openflow: enabling innovation in campus networks,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.
- [3] N. Nikaein, E. Schiller, R. Fraud, K. Katsalis, D. Stavropoulos, I. Alyafawi, Z. Zhao, T. Braun, and T. Korakis, “Network store: Exploring slicing in future 5G networks,” in *Proc. 10th Int. Wkshp. on Mobility in the Evolving Internet Archit.* ACM, 2015, pp. 8–13.
- [4] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, “Network function virtualization: Challenges and opportunities for innovations,” *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 90–97, 2015.
- [5] A. Singla, “Fat-FREE Topologies,” in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks - HotNets '16*. New York, New York, USA: ACM Press, 2016, pp. 64–70.
- [6] S. Kassing, A. Valadarsky, G. Shahaf, M. Schapira, and A. Singla, “Beyond fat-trees without antennae, mirrors, and disco-balls,” in *Proc. of SIGCOMM '17*. New York, NY, USA: ACM Press, 2017, pp. 281–294.
- [7] M. He, A. Basta, A. Blenk, and W. Kellerer, “How flexible is dynamic SDN control plane?” in *Proc. of IEEE Conference on Computer Communications Workshop SWFAN*, May 2017, pp. 689–694.
- [8] W. Kellerer, A. Basta, P. Babarzi, A. Blenk, M. He, M. Klügel, and A. M. Alba, “How to measure network flexibility? A proposal for evaluating softwarized networks,” *IEEE Communications Magazine*, vol. 56, no. 10, pp. 186–192, 2018.
- [9] P. H. Brill and M. Mandelbaum, “On measures of flexibility in manufacturing systems,” *Int. J. Prod. Res.*, vol. 27, pp. 747–756, May 1989.
- [10] A. K. Sethi and S. P. Sethi, “Flexibility in manufacturing: A survey,” *Int. J. Flexible Manuf. Syst.*, vol. 2, no. 4, pp. 289–328, 1990.
- [11] S. Benjaafar, “Modeling and analysis of flexibility in manufacturing systems,” Ph.D. dissertation, Purdue University, 1992.
- [12] I. J. Chen and C.-H. Chung, “An examination of flexibility measurements and performance of flexible manufacturing systems,” *Int. J. Prod. Res.*, vol. 34, no. 2, pp. 379–394, Feb 1996.
- [13] S. K. Das, “The measurement of flexibility in manufacturing systems,” *Int. J. Flexible Manuf. Syst.*, vol. 8, no. 1, pp. 67–93, Jan 1996.
- [14] R. E. Giachetti, L. D. Martinez, O. A. Sáenz, and C.-S. Chen, “Analysis of the structural measures of flexibility and agility using a measurement theoretical framework,” *Int. J. Prod. Econ.*, vol. 86, no. 1, pp. 47–62, Oct 2003.
- [15] T. Marschak and R. Nelson, “Flexibility, Uncertainty and Economic Theory,” *Metroeconomica*, vol. 14, no. 1-2-3, pp. 42–58, Feb 1962.
- [16] M. Mandelbaum, “Flexibility in decision making: an exploration and unification,” Ph.D. dissertation, University of Toronto, 1978.
- [17] J. Lasserre and F. Roubellat, “Measuring decision flexibility in production planning,” *IEEE Trans. Autom. Control*, vol. 30, no. 5, pp. 447–452, May 1985.
- [18] M. Mandelbaum and J. Buzacott, “Flexibility and decision making,” *Eur. J. Oper. Res.*, vol. 44, no. 1, pp. 17–27, 1990.
- [19] S. Benjaafar, T. L. Morin, and J. J. Talavage, “The strategic value of flexibility in sequential decision making,” *Eur. J. Oper. Res.*, vol. 82, no. 3, pp. 438–457, 1995.
- [20] W. Golden and P. Powell, “Towards a definition of flexibility: in search of the Holy Grail?” *Omega*, vol. 28, no. 4, pp. 373–384, Aug 2000.
- [21] T. Tao, *An introduction to measure theory*. American Mathematical Society Providence, RI, 2011.
- [22] P. Babarzi, J. Tapolcai, A. Pašić, L. Rónyai, E. R. Bérczi-Kovács, and M. Médard, “Diversity coding in two-connected networks,” *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2308–2319, Aug 2017.