Assessing the Risk of Violating SLA Dependability Requirements in Software-Defined Networks

Andrzej Kamisinski*, Bjarne E. Helvik**, Andres J. Gonzalez§, Gianfranco Nencioni**

* Department of Telecommunications, AGH University of Science and Technology, Kraków, Poland
Email: kamisinski@kt.agh.edu.pl

** IIK, NTNU — Norwegian University of Science and Technology, Trondheim, Norway
Email: {bjarne.e.helvik, gianfranco.nencioni}@ntnu.no

§ Research Department, Telenor ASA, Tromsø, Norway
Email: andres.gonzalez@telenor.com

Abstract—Dependability of computer and communication networks is an important aspect of the customer-provider relationship in the telecommunication industry. Considering the increasing interest in Software-Defined Networking (SDN) technologies, as well as unclear understanding of dependability in the context of traffic flows in such networks, it is not clear how to define the related objectives in Service Level Agreements (SLAs), and how to estimate the risk of violation of the included dependability requirements. In this paper, we present a solution to both issues and we evaluate it in different scenarios by simulation. The results show that the proposed method is feasible and may help service providers to select the preferred recovery technique in SDN based on the estimated risk of violation of the SLA dependability requirements and known Service Level Objectives (SLOs).

I. INTRODUCTION

An increasing demand for diverse network services with different dependability requirements has prompted Internet Service Providers (ISPs) to modernize and expand their national and global infrastructure. To ensure that the services are delivered at the required levels with respect to dependability, ISPs constantly monitor their networks and respond to all events affecting the performance of their services [1], [2]. As part of this task, they also try to estimate the risk of violation of the dependability requirements specified in Service Level Agreements (SLAs) signed with their customers to prepare and deploy adequate protection measures on time, thus avoiding major service disruption and the related penalty [3]–[5]. Each violation of the dependability-related Service Level Objectives (SLOs) defined in the corresponding SLA may lead to significant monetary consequences, affecting the ISP’s reputation.

With the introduction of Software-Defined Networks (SDNs) [6], [7], in which the control plane is decoupled from the data plane, the research community and telecommunication industry became interested in the flexibility they offer, and in the expected simplification of network management tasks. However, researchers soon realized that it is necessary to identify the potential sources of failures in such networks, as well as their impact on the overall dependability of a system [8]–[13]. Further, it is not clear how to construct dependability-related SLOs in the case of SDNs, and how to estimate the related SLA violation risk. To the best of our knowledge, no previous work has addressed this issue, while it needs to be solved before ISPs start using SDN in their network infrastructure. The currently used metrics in SLAs mainly refer to service downtime which has not been explicitly defined in the case of SDNs. The existing proposals cover non-SDN networks in which an ICT service is either available or not [3]–[5], [14]. At the same time, in SDNs, traffic flows established before a failure may still be successfully forwarded through a network, while the new flows between the same pair of nodes may be rejected due to the unavailability of the logically-centralized controller. Thus, the existing solutions that cannot handle this case are not directly applicable to SDNs and should be revisited in this context.

The objective of this paper is to provide ISPs and their customers with a valid and conceptually simple solution that allows them to i.) define the dependability-related SLOs for traffic flows in SDNs, and ii.) assess the respective risk of violation of the SLA dependability requirements and take appropriate measures in advance to ensure that this risk remains within an acceptable range. Thus, the main contributions of this paper can be summarized as follows:

- the main factors affecting the dependability of SDN networks from the perspective of customers and their ability to utilize their offered services are identified;
- service degradation is defined as the key measure of decreased dependability in SDN environments;
- a method for the assessment of the SLA violation risk with respect to the dependability requirements for traffic flows in SDNs is proposed and evaluated by simulation.

In this paper, we follow [15] with respect to the definitions of dependability, availability, and reliability of systems.

The remainder of this paper is structured as follows: in Section II, the general architecture of SDN is presented, and the challenges of the SLA-based business relationship model are discussed with respect to the ability to assess the dependability of traffic flows in SDNs. Section III introduces the service degradation metric used in the assessment of risk of violation of the SLA dependability requirements for traffic flows in SDN, while Section IV describes the evaluation strategy, summarizes the results, and identifies the advantages and disadvantages of the proposed approach. Finally, Section V


II. SDN ARCHITECTURE AND SLAs

The concept of SDN relies on the assumption that the control plane is decoupled from the data plane. This is a significant design approach that not only simplifies the management of network devices, but it also brings in new challenges in terms of the overall dependability of the network. For example, the survivability of the control plane has been discussed in [16] in the context of the Generalized Multiprotocol Label Switching (GMPLS) or Automatically Switched Optical Networks (ASON) model. To make a clear connection between the considered type of network and the proposed solution described in Section III, we present the general architecture of SDN in Figure 1.

As presented in the figure, customers are located in the access networks. The customers sign SLAs with their service providers for specific contract periods. The contract period formally defines the agreed time for which the SLA is legally binding [5]. Every customer may sign one or more SLAs with a provider, depending on the type and importance of the services it would like to use (e.g., delay-sensitive voice communication, financial transactions, or best effort web traffic), as well as the respective quality and dependability-related requirements, i.e., the SLOs. The service provider is expected to provide the service in such a way that the corresponding SLOs are satisfied. In the opposite case, the provider has to compensate the customer according to the agreed scheme. Typical business relationships that fit this model, and which we consider in this paper, are established between service providers and either individual or corporate customers.

Customers send traffic flows between pairs of network nodes. At any point in time, the paths for some flows may have already been established in the network, while the other flows might have just arrived at the first SDN switch on their paths and have to be configured by the logically-centralized SDN controller. Note that in SDNs, some flows which have been inactive for a specified time may be marked as expired.

In the considered SDN network model, switches mainly forward customers’ traffic. At the same time, it is important to note that control traffic can also be transmitted using the same infrastructure (In-Band control)\(^1\), instead of using dedicated links (Out-of-Band control). In such a case, the reliability of operation of the control plane within an SDN depends on the operation of the same data plane that is to be controlled, which significantly extends the consequences of failures in the data plane and is undesirable with respect to the overall dependability of the system. Switches are connected to one or more logically-centralized controllers and communicate with them using a control protocol, such as OpenFlow [6]. The controllers provide an open interface for custom applications, for example, traffic engineering or network measurement applications. All components in the data plane, the controller plane, and the application plane are managed by network engineers, possibly with the aid of automated tools (note the presence of the respective connections in Figure 1).

Currently, it is not clear how to construct dependability-related SLOs in the case of SDN networks. One of the main differences between the existing computer and communication networks and SDNs is that for each new flow in an SDN, a working connection from each switch that will belong to the configured flow’s path to at least one controller is needed. An immediate conclusion is that if no controller is reachable from a switch when a new flow arrives, the customer observes (partial) service downtime, even while the previously-established traffic flows are still transmitted successfully (service uptime).

Unlike the existing proposals, the concept presented in this paper provides a solution to this issue. We introduce our method in Section III.

III. ASSESSMENT OF THE SLA VIOLATION RISK IN SDN

In this section, we present a method that enables ISPs to estimate the risk of violating an SLA as a result of exceeding the maximum agreed service degradation. The term *service degradation* is an important concept that we propose and it is defined as the fraction of the total number of traffic flows of a customer that were not successfully delivered to the intended destinations during the observation period. Note that by *observation period*, we mean a predefined time interval over which the service degradation metric is computed. In particular, several observation periods may exist within the SLA contract period [5].

To better illustrate the proposed idea, let us consider the example scenario shown in Figure 2. The figure captures an arbitrarily-selected period of 60 seconds during which a single customer sends several traffic flows through an SDN backbone.

\(^1\)Actually, this is a likely scenario, especially in the case when control messages have to be transmitted over long distances, which means that reserving some links or transmission channels exclusively for control traffic would be much more expensive than sharing the resources with customers’ traffic. On the other hand, whenever the reliability of such communication channels is critical to network operation, it might be reasonable to use dedicated channels on short-distance links at the cost of increased capital and operational expenditures.
network. The black curve with square points, \( n_a(t) \), reflects the number of flows of that customer at time \( t \). Among these flows are both the new (i.e., unregistered or expired) and the previously-configured flows. The number of correctly-transmitted flows is represented by \( n_c(t) \) and the blue curve with square points. Once a failure occurs in the network, some flows may not be delivered to the intended destinations, which is reflected by the red curve with triangular points, \( n_a(t) - n_c(t) \). The area below this curve divided by the length of the observation period provides information about the average number of the customer’s flows that were not delivered to the intended destinations in the period \([0, t]\). Once this value is computed, we can determine the service degradation for the entire observation period, and then the new estimation of the SLA violation risk with respect to the dependability-related SLOs for traffic flows.

While working on the solution, we have made the following general assumptions:

- service degradation must be measurable by both service providers and customers;
- failures affecting the already established flows have the same impact on the estimated service degradation as failures related to new flows (i.e., the service degradation is estimated based on an assumption that all flows are equally important for the customer);
- we have enough network resources to respond to failures through flow rerouting; at the same time, the recovery does not have to be successful.

In addition, we do not consider the volume of traffic flows as an indicator of their importance. Note that one flow of very small volume can be far more important than several high-volume flows.

The symbols used in the formulation are presented in Table I. To simplify the explanation (without losing generality), we assume that at each time \( t \), new traffic flows receive unique indices in the range of 1 to \( n_n(t) \), while the existing flows are assigned higher indices between \( n_n(t) + 1 \) and \( n_n(t) \). In addition, we assume that the customer has signed a single SLA with the service provider\(^3\). In the first step, we count the number of customer’s flows that can be delivered to the intended destinations at time \( t \), which is reflected by the \( n_c(t) \) function as follows:

\[
n_c(t) = \sum_{i=1}^{n_n(t)} p(t,i) c(t,i) + \sum_{i=n_n(t)+1}^{n_a(t)} p(t,i) \quad (1)
\]

Based on Equation (1), the overall service degradation \( D(\tau) \) corresponding to the signed SLA and the related observation period \( \tau \) is computed as follows:

\[
D(\tau) = \begin{cases} 
1 - \int_{0}^{\tau} n_c(t) \, dt & \text{if } \exists t \in [0,\tau] \; n_a(t) > 0 \\
0 & \text{otherwise}
\end{cases} \quad (2)
\]

Note that \( \int_{0}^{\tau} n_c(t) \, dt \) is the number of flows of the selected customer that were delivered to the intended destinations during the observation period \( \tau \), while \( \int_{0}^{\tau} n_a(t) \, dt \) is the number of all flows offered by the customer during that period.

Once the service degradation \( D(\tau) \) is computed for each of the consecutive observation periods, it is possible to determine the related Cumulative Distribution Function (CDF). The SLA success probability \( S(\tau, \alpha) \) and the SLA violation risk \( W(\tau, \alpha) \) with respect to the agreed service degradation threshold \( \alpha \) and length of an observation period \( \tau \) are then:

\[
S(\tau, \alpha) = \Pr \{ D(\tau) \leq \alpha \} \quad (3)
\]

\[^3\text{The analysis would follow the same steps individually for each SLA signed with the given customer. We believe that such an approach provides greater flexibility with respect to groups of flows that have different requirements.}\]
The actual values were selected at random according to the uniform distribution. The numerical values were selected in the range [0, 1], whereas the actual values were selected according to the exponential distribution with the mean value of 0.1 s, while the duration of each flow followed the Pareto distribution with the mean value of 60 s and the shape parameter equal to 1.5. The average flow demand was set to 1 Mbit/s, whereas the actual values were selected at random from range [0.75; 1.25] Mbit/s according to the uniform distribution. The numerical values were selected in such a way that the generated traffic did not cause link congestion during fault-free network operation. Moreover, we assumed that the service provider maintains sufficient capacity on links to deal with failures in the network, so that at least a small fraction of each affected flow could still be transmitted, if only the respective reachability requirements were satisfied. During the steady-state period of each simulation run, the total number of traffic flows in the network oscillated around an average value. Further, it was assumed that every 10000 s on average, a node in the network would fail. The time between consecutive failures was selected at random according to the exponential distribution. In the case of links, failures occurred every 1000 s on average. The Mean Time To Repair (MTTR) for nodes and links was set to 1000 s and 100 s, respectively. The actual values were selected at random according to the Pareto distribution with the shape parameter equal to 1.5.

In each simulation run, the service degradation metric was computed every month (τ = 1 month), while the risk of violation of the dependability-related SLOs for each SLA was computed after 5 consecutive months, based on Equation (4).

\[
W(\tau, \alpha) = 1 - S(\tau, \alpha)
\] (4)

See Figure 3 for an illustration. Every SLA, for which the risk of violating the service degradation requirement is estimated using the proposed method, should be extended with at least the α and τ parameters agreed on with the customer.

Based on the recent work applicable to typical computer and communication networks [14], it is possible to make risk-aware decisions on the preferred use of different dependability provisioning techniques in the case of SDNs. Although the implementation of specific solutions in real networks is usually complex, we provide the possible directions as the first step, and we use simulation to illustrate the potential use cases.

IV. EVALUATION

The evaluation of the proposed solution was based on discrete-event, flow-level network simulation. The main objective of the simulation study was to confirm the capability of the proposed method to estimate the SLA violation risk with respect to the dependability SLOs in Software-Defined Networks in different scenarios. It is shown that based on the collected results, suggestions can be made about the additional protection measures that should be deployed in the network to reduce the risk of violation of the dependability-related SLOs of each individual SLA to an acceptable level.

A. Evaluation Scenarios

To illustrate the possible use cases for the proposed solution, the following two evaluation scenarios were considered:

- Scenario I: homogeneous service degradation threshold α across all SLAs (standard SLAs);
- Scenario II: differentiated service degradation threshold (standard SLAs: αs, business SLAs: αb).

Standard SLAs were only signed by customers who were sending traffic between nodes selected at random according to the uniform distribution. However, to reflect the fact that companies often have remote premises in different cities, business SLAs have been introduced in Scenario II. Business SLAs assumed that traffic flows are transmitted within the predefined business groups, with higher SLA dependability requirements than traffic associated with standard SLAs. Each business group was defined as a set of 4 different nodes selected at random and representing the corresponding company’s premises in different cities. In this case, the source and destination nodes were selected at random from the same business group according to the uniform distribution. Further, it was assumed that there are 100 SLAs per backbone node in the network. In the case of Scenario II, each backbone node represented the home location for 70 standard SLAs and 30 business SLAs signed with local customers.

The evaluation was based on the modified version of the real-world backbone network topology shown in Figure 4. The considered network included two logically-centralized SDN controllers (C1 and C2). Each link of the network had the capacity of 1 Gbit/s. The flow inter-arrival time was selected according to the exponential distribution with the mean value of 0.1 s, while the duration of each flow followed the Pareto distribution with the mean value of 60 s and the shape parameter equal to 1.5. The average flow demand was set to 1 Mbit/s, whereas the actual values were selected at random from range [0.75; 1.25] Mbit/s according to the uniform distribution. The numerical values were selected in such a way that the generated traffic did not cause link congestion during fault-free network operation. Moreover, we assumed that the service provider maintains sufficient capacity on links to deal with failures in the network, so that at least a small fraction of each affected flow could still be transmitted, if only the respective reachability requirements were satisfied.

In each simulation run, the service degradation metric was computed every month (τ = 1 month), while the risk of violation of the dependability-related SLOs for each SLA was computed after 5 consecutive months, based on Equation (4).
For error control, 10 independent simulation runs were executed. Each simulation run consisted of the following phases:

- the transient period of 512 s (estimated using the method presented in [18]) — to make sure that samples were collected during the steady-state period of the simulation;
- 5 consecutive observation periods, each of length \( \tau \);
- the termination phase of at least 40 s.

In each of the considered scenarios, different values of the service degradation threshold \( \alpha \) were considered. All traffic flows were forwarded along the shortest paths and it was assumed that the only available resilience provisioning mechanism was flow rerouting.

In real computer and communication networks, rerouting traffic flows may take different amounts of time — usually from tens of milliseconds to several seconds. The duration of this process depends on several factors, such as: the selected rerouting strategy (e.g., convergence of a specific routing protocol, IP Fast Reroute-based schemes), layer of operation, network technology, network topology, and available resources. As the operation of specific routing protocols and recovery mechanisms is beyond the scope of this paper, we assumed that rerouting a flow after node or link failure imposes the related service downtime of 1 s.

### B. Evaluation Results

The evaluation results corresponding to Scenario I are shown in Figures 5-6. The first figure represents an example simulation run and shows the Cumulative Distribution Function (CDF) of the SLA violation risk with respect to the service degradation requirement \( \alpha \) (further referred to as SLA violation risk; see Equation (4)). It may be noticed that lower values of \( \alpha \) resulted in higher SLA violation risk. Thus, if service providers decide to guarantee low service degradation in their agreements with customers, they should already be prepared to activate enough redundant network resources immediately when such demand occurs. The second figure presents the estimated CDFs of the maximum SLA violation risk and the arithmetic mean of the SLA violation risk computed for all SLAs in each simulation run. Note that the curves in Figure 6(a) overlap and have a steep slope for \( x = 1 \) for all considered values of \( \alpha \), which shows that the maximum SLA violation risk was always equal to 1. Furthermore, it is clearly visible that the maximum computed SLA violation risk among all SLAs may deviate significantly from the arithmetic mean of the risk values determined for the same set of SLAs. Thus, it is a critical factor that should be monitored.

The impact of the maximum computed SLA violation risk on the decision making process may even be stronger when some SLAs assume different service degradation thresholds than the other SLAs, which was the main focus of Scenario II.

In this case, we considered three different combinations of service degradation thresholds for standard and business SLAs. As business communication is usually associated with higher dependability guarantees, we assumed that the corresponding service degradation threshold \( \alpha_b \) should always be lower than the corresponding threshold for standard SLAs, \( \alpha_s \). The simulation results have shown that in the considered evaluation scenario, the maximum SLA violation risk and the arithmetic mean of the SLA violation risk computed with respect to the exceeded service degradation threshold were much higher for business SLAs, as they had stronger dependability guarantees than standard SLAs (see Figures 7 and 8; note that there are two overlapping curves in Figure 8(a)). Thus, to avoid penalties due to violated dependability-related SLOs of business SLAs, the service provider should either deploy more effective recovery mechanisms for the related traffic flows, or negotiate higher service degradation thresholds with its business customers. However, different recovery mechanisms may have different cost. An idea of how to select the optimal recovery strategy based on the related cost and the estimated SLA violation risk has been presented in [14]. Now, being able to estimate the risk of violation of the dependability-related SLOs for each SLA in SDNs, service providers may plan future expenditures more effectively.

### V. Conclusion

This paper presents a method for the assessment of the SLA violation risk with respect to the dependability-related SLOs defined for traffic flows in SDNs. To clarify the understanding of dependability in the context of traffic flows in SDNs, the main related factors are identified, and then service degrada-
The presented work is the first step to understand how to select the preferred recovery technique based on the estimated SLA violation risk related to the known dependability SLOs. The simulation results show that the proposed solution is feasible and may help service providers to manage the SLA violation risk. The results represent all simulation runs ($N = 10$) and all 2730 standard SLAs (70 per each network node).

Fig. 7. Scenario II: The estimated CDF of (a) the maximum SLA violation risk and (b) the arithmetic mean of the SLA violation risk with respect to the service degradation requirement $\alpha_s$. The results represent all simulation runs ($N = 10$) and all 2730 standard SLAs (70 per each network node).

Fig. 8. Scenario II: The estimated CDF of (a) the maximum SLA violation risk and (b) the arithmetic mean of the SLA violation risk with respect to the service degradation requirement $\alpha_s$. The results represent all simulation runs ($N = 10$) and all 1170 business SLAs (30 per each network node).

The authors would like to thank Eirik Følstad and Andrzej Jajszczyk for their valuable comments about the proposed method, SLAs, and SDNs. A. Kamisiński was supported by the collaboration project between Telenor Research and NTNU QUAM Research Lab (research), and AGH University of Science and Technology under contract no. 11.11.230.018 (presentation at the conference).

REFERENCES