On the Impact of Billing Cycles on Compensations in Network SLAs

Tobias Enderle
Institute of Communication Networks and Computer Engineering
University of Stuttgart
Stuttgart, Germany
tobias.enderle@ikr.uni-stuttgart.de

***Abstract*—Network connections in optical backbone networks are subject to failures and the network operator is obliged to compensate its customers according to the terms of the service level agreement (SLA) if outages have occurred. These compensations are typically made after regular billing cycles. In most SLAs the billing cycle has a length of one calendar month. However, the length of the billing cycle has a significant impact on the compensation that is to be paid by the network operator. To show this, we first derive the exact distribution of the downtime per billing cycle for dedicated path-protected connections with exponential failure and repair times. Based on this, we derive the expected amount of compensation for different SLA compensation policies. We show that, depending on the failure and repair characteristics of the network components, there is a billing cycle for which the expected amount of compensation is maximized and hence is to be avoided. Consequently, the adjustment of the billing cycle to the failure and repair characteristics or vice versa can play an important role in the optimal design of SLAs.**

***Index Terms*—Compensation, Downtime distribution, Interval availability, Network outage, Service level agreement**

I. Introduction

Availability is one of the most important performance indicators for connections in today's networks, especially when it comes to optical backbone networks with their enormous data rates. Outages in the network like link failures disrupt many connections at once and in that way affect many customers. The customers have to be compensated for these outages, e.g., by paying refunds or issuing service credits. Typically, this happens once at the end of the connection's contract period or, for long-running connections, regularly after a fixed billing cycle [1]. Details about the compensation as well as the corresponding guaranteed levels of availability are stipulated in the service level agreement (SLA) between the network operator and its customer. To increase the operator's profit, compensation payments should be avoided or reduced where possible. One way of achieving this is the introduction of redundancy in the network, which increases the system availability and hence leads to less and shorter outages. However, redundancy is not for free and a balance between additional investments and reduced compensations has to be found [2]. Another way, which we pursue here, is to exploit properties of the availability itself to reduce the compensation amount without imposing extra costs.

The term *availability* has many interpretations. Typically, it refers to the *steady-state availability* which is the probability of finding a connection working at any point in time in an infinite time interval. However, as soon as a limited time interval is considered, the relevant metric is the *interval availability*, which is the actual availability of the connection during that very interval. The interval availability is a random variable (RV) because the underlying connection outages are random. Different authors have shown how the probability of delivering a certain interval availability in a limited time interval depends on the underlying failure and repair processes of the connection but also on the interval length itself. In [3], the authors derive the probability of violating the availability guaranteed in the SLA and briefly discuss the impact of the connection holding time or contract period over which the availability is evaluated. They show that the violation probability has a maximizing contract period and that shorter or longer periods reduce the violation probability. In [1] it is shown that the failure and repair processes have a significant impact on the interval availability, even if the steady-state availabilities are the same. The authors of [4] and [5] present important properties of the distribution of the interval availability and also discuss the significance of the contract period. In [6] and [7], the authors build on the theory of interval availability and derive the value at risk (VaR) for the penalty incurred by availability violations. In that way, they connect the interval availability with a business related metric. Finally, the authors of [8], like others, consider the probability of availability violations for different assumptions on the repair time. Furthermore, they briefly touch on the impact of the considered time interval not in the sense of a contract period but in the sense of a repeating billing cycle for a connection. The presented literature suggests that, for a comprehensive SLA design, it is necessary to consider failure and repair characteristics and billing cycles jointly.

In this paper, we focus on the scenario in which the availability of a long-running network connection is evaluated regularly at the end of each billing cycle. Depending on the amount of downtime the connection experienced during a billing cycle, the network operator has to pay compensation to the customer. Unlike many of the works mentioned above, we do not assume that connections have a limited contract period with only a single evaluation of the downtime and compensation at the end of the contract. Our first contribution is the exact distribution of the downtime per billing cycle for dedicated path-protected connections with exponential failure and repair times. Further, the expected amount of compensation for different compensation policies is derived, and it is shown that a precise

978-3-903176-33-1

numerical evaluation is possible using freely available software. Finally, the impact of the billing cycle on the expected amount of compensation is demonstrated in a network study.

II. Distribution of the Total Downtime

The amount of compensation a network operator has to pay depends on the number and length of the connection outages. We assume that compensations for a network connection that runs indefinitely are made regularly after a fixed time interval—the billing cycle. Therefore, the distribution of the total downtime during a billing cycle is of interest.

For the moment, we consider a network connection not as a combination of network components but as a monolithic system. We assume that this system can fail and is repairable, i.e., after a failure and a certain repair time the system is working again. Consequently, the system is either in the up state U (working) or in the down state D (not working) (see Figure 1). Failures occur randomly and also the time it takes to repair the system varies. Hence, the time $T_{\mathrm{U},i}$ the system is in state U between a successful repair (or connection setup) and the i-th failure is a RV with expectation *MTTF* (mean time to failure). Likewise, the time $T_{\mathrm{D},i}$ between the occurrence of the i-th failure and its successful repair (state D) is a RV with expectation *MTTR* (mean time to repair). We assume that all failure and repair times are independent and identically distributed, hence we set $T_\mathrm{U} \equiv T_{\mathrm{U},i}$ and $T_\mathrm{D} \equiv T_{\mathrm{D},i}$ $\forall i$. For a time interval $I = [t_0, t_0 + T]$, the total downtime X is the RV given by the sum of the individual outage times $T_{\mathrm{D},i}$ that occurred in I (if I ends in D, then the last $T_{\mathrm{D},i}$ is counted only partly). In the following, we assume that the failure and repair times T_U and T_D follow negative exponential distributions with failure rate λ and repair rate μ, like e.g. in [9] and [10].

In [11], Takács considers an equivalent system with two alternating states A and B. He provides the cumulative distribution function (CDF) for the total time β spent in state B during an interval of length T as

$$\Omega(t) = P(\beta \le t) \tag{1}$$

$$= \sum_{n=0}^{\infty} H_n(t)\big(G_n(T-t) - G_{n+1}(T-t)\big). \tag{2}$$

In this equation G and H are the CDFs of the individual times spent in the states A and B, respectively, and G_n and H_n are their respective n-th iterated convolutions with themselves. An important assumption for this result is that the system is in state A at the beginning of the considered interval.

For the case of negative exponential failures and repairs, Takács also provides a closed-form solution with $G(t) = 1 - e^{-\gamma t}$ and $H(t) = 1 - e^{-\delta t}$ in [11]:

$$\Omega(t) = e^{-\gamma(T-t)}\Bigg(1 + \sqrt{\gamma\delta(T-t)} \cdot \int_0^t e^{-\delta y} y^{-\frac{1}{2}} I_1\Big(2\sqrt{\gamma\delta(T-t)y}\Big)\,\mathrm{d}y\Bigg) \tag{3}$$

$$\equiv \Omega(t, \gamma, \delta) \tag{4}$$

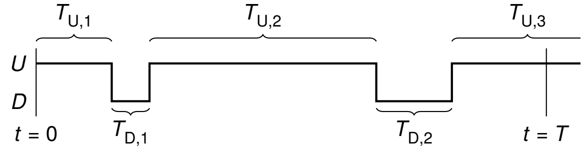


Fig. 1. Alternating system states U (working) and D (not working).

where $I_1(x)$ is the modified Bessel function of the first kind of order 1.

A direct application of the presented results from [11] are network connections with a finite contract period for which compensations are made at the end of the period, e.g. [12]. The states A and B can be mapped to the states U and D, respectively. The interval length T corresponds to the contract period and since the operator ensures that a new connection is working properly when it is handed over to the customer, the condition that the system starts in state A is fulfilled. The distribution of the total downtime X we are looking for is then given by Ω. To indicate that the state U is mapped to the initial state A we use the notation Ω_U.

For long-running connections with a fixed billing cycle instead of a predetermined contract period the case is slightly different. With a fixed billing cycle T, compensations have to be made on a regular basis. Like in the application above, in the first billing cycle, the connection has been set up and, therefore, this cycle starts in the up state. However, for all subsequent billing cycles this assumption is not necessarily true because connection outages can reach from the end of one cycle into the beginning of the next. This is especially the case when the billing cycle T is on the same order as the repair times, i.e., either for small T or large *MTTR* (e.g. 540 hours for submarine cables [13]). Therefore, it is not sufficient to use Ω directly to obtain the distribution of the downtime for a billing cycle. Instead, we use the law of total probability to combine the case of a cycle starting in U with the case of a cycle starting in D.

The probability of starting a billing cycle in the working state U is equal to the steady-state availability of the connection which is given by [14, Ch. 1.2]

$$a = \frac{MTTF}{MTTF + MTTR}. \tag{5}$$

Once a billing cycle is started in state U, the system has almost surely already been in U at the end of the previous billing cycle (see $T_{\mathrm{U},3}$ in Figure 1). Therefore, the time between the start of the cycle and its first failure is only a part of T_U. In general, this residual part does not share the distribution of T_U. However, since we assume T_U to be exponentially distributed, the residual part follows the distribution of T_U as well [14, A6.5]. Hence, with the same mapping as described above (A $\rightarrow$ U, B $\rightarrow$ D, and $\beta \rightarrow X$), Ω_U is the distribution of the downtime, i.e.,

$$P(X \le x | \xi_0 = \mathrm{U}) = \Omega_\mathrm{U}(x) \tag{6}$$

where ξ_0 denotes the state at the beginning of the billing cycle. The probability of starting in the down state D is equal to the unavailability $u = 1 - a$. In this case, we use Ω as well, but we map

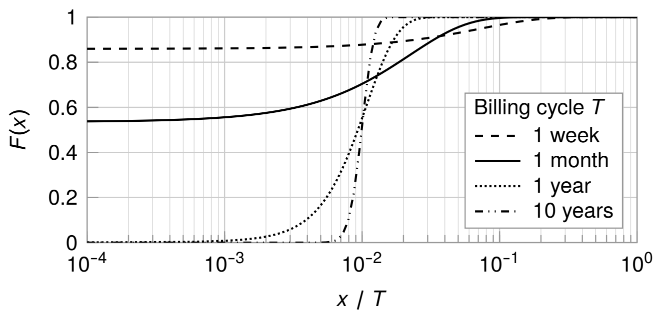


Fig. 2. Cumulative distribution function F of the downtime for different billing cycles T, $MTTR = 12\,\text{h}$, and $MTTF = 1188\,\text{h}$. The x-axis is normalized to the respective billing cycle and in log scale.

D to the initial state A and U to state B. Consequently, β corresponds to the total *uptime* in the billing cycle. Since the sum of up- and downtime is equal to the billing cycle T it follows that

$$P(X \leq x|\xi_0 = \text{D}) = P(\beta > T - x) \tag{7}$$
$$= 1 - P(\beta \leq T - x) \tag{8}$$
$$= 1 - \Omega_\text{D}(T - x). \tag{9}$$

Since the states U and D are mutually exclusive, the CDF of the total downtime X per billing cycle is

$$F(x) = P(X \leq x) \tag{10}$$
$$= aP(X \leq x|\xi_0 = \text{U}) + uP(X \leq x|\xi_0 = \text{D}) \tag{11}$$
$$= a\Omega_\text{U}(x) + (1 - a)\big(1 - \Omega_\text{D}(T - x)\big) \tag{12}$$

and with the notation $\Omega(t,\gamma,\delta)$ in (4) we get

$$F(x) = a\Omega(x, \lambda, \mu) + (1 - a)\big(1 - \Omega(T - x, \mu, \lambda)\big). \tag{13}$$

Figure 2 shows the CDF of the total downtime for different billing cycles, $MTTR = 1/\mu = 12\,\text{h}$, and $MTTF = 1/\lambda = 1188\,\text{h}$. The resulting steady-state availability is $a = 0.99$. It is visible for the billing cycles of one week and one month that as x decreases, $F(x)$ approaches a non-zero probability. This is the probability $P(X = 0)$ that there is no downtime (i.e., no outage occurs) during the interval T. This scenario occurs if the billing cycle is started in the working state U and $T_\text{U} > T$. Consequently, we have for the exponential case

$$P(X = 0) = aP(T_\text{U} > T) = F(0) = ae^{-\lambda T}. \tag{14}$$

That, in turn, means that X is a mixed RV and F has a discontinuity at $x = 0$ (not shown in the figure due to the log scale). For longer billing cycles, this probability vanishes, i.e., it is very unlikely that no failure occurs during a cycle. Since the x-axis is normalized to the billing cycle it can be interpreted as the *interval unavailability*. The steady-state unavailability in this example equals $u = 1 - a = 10^{-2}$. As can be seen in the figure, with an increasing billing cycle, the probability that the interval unavailability is close to the steady-state unavailability increases.

III. Downtime of Unprotected and Dedicated Path-Protected Network Connections

In the previous section, we have derived the distribution of the downtime of a billing cycle for a monolithic system with exponential failure and repair times with known MTTF and MTTR. However, a network connection is a composite system that consists of several components. Each of these components has its own failure and repair process and influences the distribution of the total downtime of the connection. In order to determine the downtime distribution for unprotected connections and connections with dedicated path protection, we propose to reduce the composite system of network components to a monolithic system by series and parallel reductions. For this, we assume that each individual component i has exponential failure and repair processes, and we know their means or rates $MTTF_i = 1/\lambda_i$ and $MTTR_i = 1/\mu_i$.

We assume that an unprotected connection consists of a series of N network components like fibers, amplifiers, (de)multiplexers, cross connects, and transponders. The connection is working only if all components are working. We further assume that component failures and repairs are independent of each other. Therefore, the aggregated failure rate of the series of components is the sum of the individual failure rates [14, Ch. 6.3]

$$\lambda_\text{S} = \sum_{i=1}^{N} \lambda_i = \frac{1}{MTTF_\text{S}}. \tag{15}$$

The steady-state availability of a series is given by the product of the individual availabilities, i.e.,

$$a_\text{S} = \prod_{i=1}^{N} a_i \tag{16}$$

and hence with (5) the repair rate of the serial system is

$$\mu_\text{S} = \lambda_\text{S} \frac{a_\text{S}}{1 - a_\text{S}} = \frac{1}{MTTR_\text{S}}. \tag{17}$$

To handle protected connections, we need to reduce parallel systems. For dedicated path protection with a working and a backup path it is sufficient to consider a parallel system consisting of two subsystems. According to [14, Ch. 6.4], the aggregated failure rate of this system is

$$\lambda_\text{P} = \frac{\lambda_1\lambda_2(\lambda_1+\lambda_2+\mu_1+\mu_2)}{(\lambda_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1(\lambda_1+\mu_2)+\lambda_2(\lambda_2+\mu_1)}. \tag{18}$$

The steady-state availability of the parallel system is given by

$$a_\text{P} = 1 - (1 - a_1)(1 - a_2) \tag{19}$$

and similar to (17) the repair rate of the parallel system is

$$\mu_\text{P} = \lambda_\text{P} \frac{a_\text{P}}{1 - a_\text{P}} = \frac{1}{MTTR_\text{P}}. \tag{20}$$

Using the presented transformations we are able to compute the failure and repair rates—and consequently also the distribution of the downtime—of unprotected and dedicated path-protected network connections.

In the previous two sections, we have derived the distribution of the total downtime in a billing cycle for composite systems like unprotected or protected network connections. In the next section we extend this result in order to compute the expected amount of compensation a network operator has to pay.

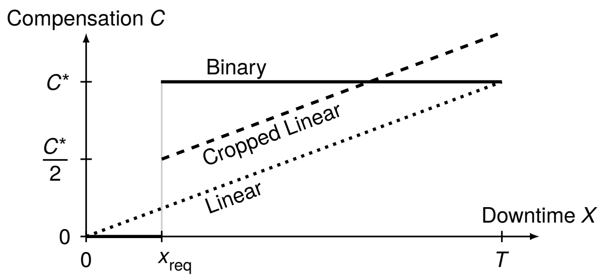


Fig. 3. Compensation policies *Binary*, *Linear*, and *Cropped Linear*.

IV. Expected Amount of Compensation

As introduced above, the random variable X describes the total downtime of a connection during a time interval I of length T. The compensation the network operator has to pay is a function of this downtime. Let C be the RV that describes the amount of compensation for the time interval I and let $g_p : [0,T] \to \mathbb{R}$ be the mapping from the downtime X to the compensation C for the *compensation policy p*. Figure 3 shows three compensation policies that are based on those presented in [6]:

- *Binary (B)* $$g_{\text{B}}(x) \equiv \begin{cases} 0 & \text{for } x \le x_{\text{req}} \\ C^* & \text{for } x > x_{\text{req}} \end{cases} \quad (21)$$
- *Linear (L)* $$g_{\text{L}}(x) \equiv \frac{C^*}{T} x \quad (22)$$
- *Cropped Linear (CL)* $$g_{\text{CL}}(x) \equiv \begin{cases} 0 & \text{for } x \le x_{\text{req}} \\ \frac{C^*}{T}(x - x_{\text{req}}) + \frac{C^*}{2} & \text{for } x > x_{\text{req}} \end{cases} \quad (23)$$

The policies *Binary* and *Cropped Linear* are sensitive to a maximum allowed downtime x_{req} agreed on in the SLA such that shorter downtimes will not lead to compensation payments.

For each billing cycle, the compensation is calculated using the policy function g_p and the measured downtime x in the cycle. The expected amount of compensation, which we denote by C_p, is given by the Riemann-Stieltjes integral

$$E[C] = \int_{x=-\infty}^{\infty} g_p(x) \, \mathrm{d}F(x) \equiv C_p \quad (24)$$

where $F(x)$ is the CDF of X derived in (13).

In order to evaluate (24) numerically, we transform the integral and use properties of the individual compensation policies. For *Binary*, we restrict the integration to the interval $(x_{\text{req}}, T]$ because $g_{\text{B}}(x) = 0$ everywhere else and obtain

$$C_{\text{B}} = C^* \int_{x=x_{\text{req}}}^{T} \mathrm{d}F(x) \quad (25)$$
$$= C^* \big(F(T) - F(x_{\text{req}})\big) \quad (26)$$
$$= C^* \big(1 - F(x_{\text{req}})\big) \quad (27)$$

because $F(T) = 1$ by definition. The same result is obtained by considering the fact that the compensation amount under the *Binary* policy is a discrete RV taking values from the set $\{0, C^*\}$. For *Linear* we employ partial integration [15, Ch. 3.3] and obtain

$$C_{\text{L}} = g_{\text{L}}(T)F(T) - g_{\text{L}}(0)F(0) - \int_{x=0}^{T} F(x) \, \mathrm{d}g_{\text{L}}(x) \quad (28)$$
$$= g_{\text{L}}(T) - \int_{x=0}^{T} F(x) g'_{\text{L}}(x) \, \mathrm{d}x \quad (29)$$
$$= C^* - \frac{C^*}{T} \int_{x=0}^{T} F(x) \, \mathrm{d}x. \quad (30)$$

Here we used the properties that $g_{\text{L}}(0) = 0$, $g_{\text{L}}(T) = C^*$, and $C^*/T = g'_{\text{L}}(x)$ is the constant slope of g_{L}. Similarly, for *Cropped Linear* we arrive at

$$C_{\text{CL}} = C^* \left(\frac{3}{2} - \frac{x_{\text{req}}}{T} - \frac{F(x_{\text{req}})}{2} - \frac{1}{T} \int_{x=x_{\text{req}}}^{T} F(x) \, \mathrm{d}x \right). \quad (31)$$

We use the open-source Python packages *SciPy* [16] and *mpmath* [17] for the numerical evaluation of the provided equations including the remaining integrals in (30) and (31). For most parameters, *SciPy* provides sufficiently precise results. However, whenever *SciPy* reports stability problems (which is the case especially for long billing cycles T), we switch to the arbitrary-precision library *mpmath* to compute stable integrals.

Notice that with the help of the presented transformations, we completely avoid the numerical differentiation of F which improves the stability of the numerical evaluation of $E[C]$ greatly.

In order to verify both the correctness of the derived equations and the precision of the numerical evaluation we have conducted a reference simulation using a discrete-event simulator. In the simulation we consider the simple case of a connection that uses a single fiber with a length of 300 km, and we compare the results for different types of fiber deployments. The deployments include aerial, buried, and submarine fiber, and they differ in terms of the time between failures and the time needed for repair. Table I presents the detailed characteristics based on [13], namely the average length of fiber that suffers one cable cut per year (CC) and the MTTR as well as the MTTF and the resulting steady-state availability for 300 km of fiber. The MTTF of a fiber with length ℓ can be calculated by

TABLE I
Failure and Repair Characteristics and Availability of Different Fiber Deployments (Based on [13]).

Deployment	CC in km	MTTR in h	MTTF in h for 300 km	Availability for 300 km
Aerial	20	6	584	0.9898
Buried (conservative)	275	24	8030	0.9970
Buried (nominal)	300	12	8760	0.9986
Buried (optimistic)	628	9	18 338	0.9995
Submarine	5300	540	154 760	0.9965

$$MTTF = \frac{CC \cdot 365\,\text{days} \cdot 24\,\text{h/day}}{\ell}. \quad (32)$$

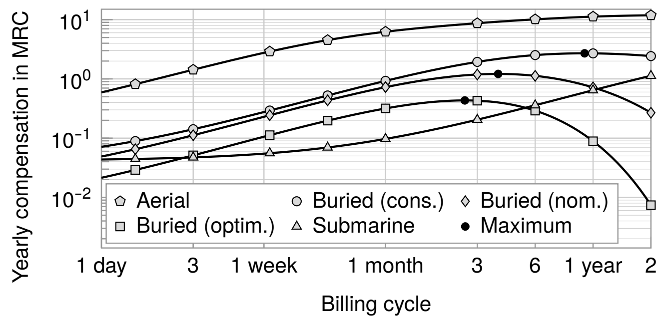
We assume that the required availability per billing cycle in the SLA is $a_{\text{req}} = 0.995$, which yields a cycle length-dependent maximum allowed downtime of $x_{\text{req}} = T(1 - a_{\text{req}})$. Since we study different billing cycles T, the compensation policies need to be scaled accordingly, which means that C^* depends on T. We assume that C_T^* for a billing cycle T is a multiple of the monthly recurring charge (MRC) a customer would pay for a connection with a cycle of one month, i.e., $C_T^* = 1\,\mathrm{MRC} \cdot T/1\,\mathrm{month}$. As an example, using the policy *Binary*, the compensation for not fulfilling a_{req} in a billing cycle of one month is 1 MRC while for a cycle of one year (12 months) it is 12 MRC.

In the simulation, we randomly generate failure and repair times according to the parameters of Table I, and we log the corresponding outage times as well as the resulting compensations per billing cycle. The number of simulated fiber cuts has been chosen high enough to guarantee that the widths of the 95 % confidence intervals do not exceed 2 % of the corresponding mean (therefore we omit the error bars in the following figures).

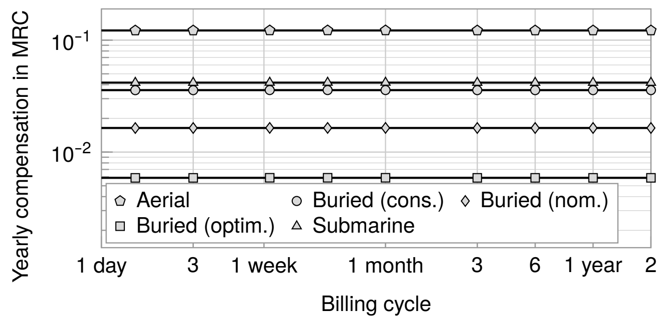
Figure 4 shows the results of both the simulation and the numerical evaluation for the expected compensation amount. To make the different billing cycles comparable, the compensation is given for the period of one year, i.e., in the simulation the sum of all compensations has been divided by the number of years simulated. The relative difference between the numerical evaluation and the simulation stays below 0.5 %, which shows that the numerical approach yields accurate results. Further evaluations show the same level of accuracy also for systems with much lower availability. It is visible that the compensation policies *Binary* and *Cropped Linear* in Figure 4a and Figure 4c have similar behavior. Of course, the expected compensation depends on the fiber deployment because different deployments have different failure and repair processes and availabilities. Moreover, the compensation also depends on the billing cycle. For the buried fibers, the figures show a maximum compensation between 2.5 and 10.9 months. For shorter or longer billing cycles, the compensation is lower which shows that the choice of the billing cycle can have a significant impact on business models. Although not visible in the graph, the same holds for the submarine fiber. Here, the maximizing billing cycle is close to 290 months due to the large MTTR. This makes clear that the choice of the billing cycle has to be considered jointly with the failure and repair characteristics of the network components. For the aerial fiber there is no maximum because its steady-state availability a_{f} is worse than the required availability a_{req}. With the billing cycle approaching infinity, the RV X converges to the constant $x_{\mathrm{f}} = T(1-a_{\mathrm{f}}) > x_{\mathrm{req}}$ and $E[C]$ converges to $g_p(x_{\mathrm{f}})$. Therefore, the expected yearly compensation approaches 12 MRC for *Binary* and $12\,\mathrm{MRC} \cdot (a_{\mathrm{req}} - a_{\mathrm{f}} + 0.5) = 6.062\,\mathrm{MRC}$ for *Cropped Linear*. As can be seen in Figure 4b, the billing cycle has no impact on the policy *Linear* because it considers outages of any length proportionally. The differences between the deployments arise solely from their different steady-state availabilities.

V. Network Study

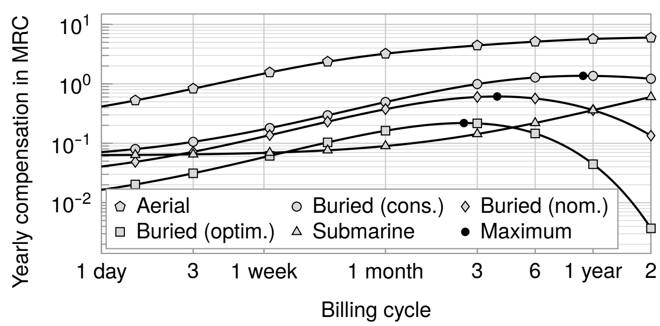
In this section, we extend the previous evaluation to a network-level study. We consider the European topology "Cost266" from [18] depicted in Figure 5. The topology has 37 nodes and an average fiber link length of 657 km. We assume uniform traffic, i.e., one connection between each node pair. A



(a) *Binary* compensation policy



(b) *Linear* compensation policy



(c) *Cropped Linear* compensation policy

Fig. 4. Yearly amount of compensation for different deployments. The lines correspond to the numerical calculation; the marks (except "Maximum") show the simulation results. Both axes are in log scale.

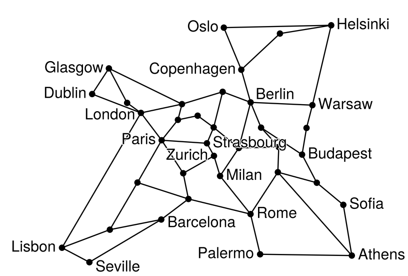


Fig. 5. European fiber topology "Cost266" (based on [18]).

connection's working path is protected by a dedicated backup path. Both are routed on link-disjoint shortest paths w.r.t. the fiber length. We employ the findings from Section III to compute the expected compensation for these protected connections. We assume that of all involved network components only fibers fail because they have the biggest impact on the downtime. We model all links as buried fibers and use the conservative case

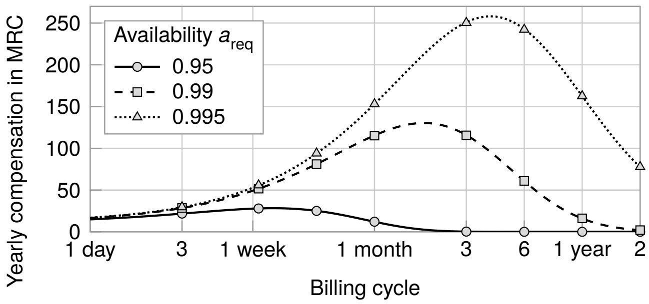


Fig. 6. Yearly amount of compensation for the European network and the *Cropped Linear* compensation policy. The lines correspond to the numerical calculation; the marks show the simulation results. The x-axis is in log scale.

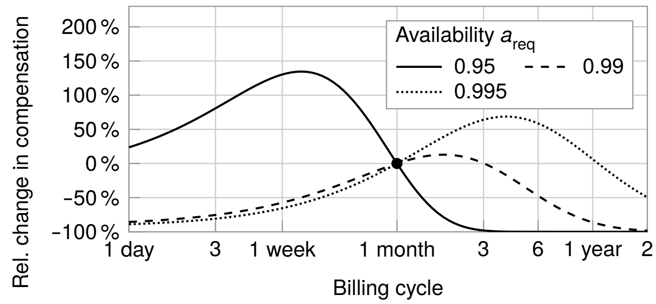


Fig. 7. Relative change in the expected amount of compensation w.r.t. a billing cycle of one month for the European network and the *Cropped Linear* compensation policy. The x-axis is in log scale.

of Table I, i.e., $CC = 275$ km and $MTTR = 24$ h. The resulting average connection availability over all connections is 0.9991, with a minimum of 0.996 between Seville and Oslo and a maximum of 0.99997 between Strasbourg and Zurich.

As before, we also confirmed the numerical computations with a simulation. Since the compensation policy *Linear* is insensitive to the billing cycle and *Binary* and *Cropped Linear* behave similarly, we only consider *Cropped Linear* here.

Figure 6 shows the total expected yearly compensation over all connections for different required availabilities. As can be seen, the amount of compensation increases with a stricter required availability. Also, the billing cycle that maximizes the compensation increases with the required availability. For $a_{\text{req}} = 0.95$, this means that by increasing the typical billing cycle of one month to a higher value, the expected compensation can be reduced. Contrary to this, for $a_{\text{req}} = 0.99$ and 0.995, stretching the billing cycle first increases the expected compensation until the maximum is reached. This behavior is also illustrated in Figure 7, where the relative change in expected compensation w.r.t. a billing cycle of one month is shown. As an example, for $a_{\text{req}} = 0.99$ the billing cycle must be increased to over three months to achieve a reduction in compensations. For the stricter $a_{\text{req}} = 0.995$, the cycle needs to be increased to more than one year to yield a reduction. Alternatively, the cycle can be reduced to any value below one month in both cases.

VI. Conclusion

In this paper we have studied the distribution of the downtime and the resulting compensations for dedicated path-protected network connections that are billed in regular cycles. The derived analytical model provides a means for network operators to estimate expected compensations numerically. It was demonstrated in different scenarios that there is a billing cycle which maximizes the expected compensation, and that the compensation obligations can be reduced if billing cycles are selected that are longer or shorter than this maximizing cycle. However, it is important to consider the billing cycle and the failure and repair processes of the network components jointly. Together, they should play an important role in the SLA design and the development of business models. Future work will extend the presented methods to more sophisticated protection mechanisms and additional repair distributions.

References

[1] M. Xia, M. Tornatore, C. U. Martel, and B. Mukherjee, "Risk-Aware Provisioning for Optical WDM Mesh Networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 921–931, 2010.

[2] C. Meusburger and D. A. Schupke, "Method to Estimate the Break-Even Point Between SLA Penalty Expenses and Protection Costs," in *Optical Fiber Communication Conference*. Optical Society of America, 2008.

[3] M. Xia, J. ho Choi, and T. Wang, "Risk Assessment in SLA-Based WDM Backbone Networks," in *Optical Fiber Communication Conference*. Optical Society of America, 2009.

[4] A. J. Gonzalez and B. E. Helvik, "Guaranteeing Service Availability in SLAs; a Study of the Risk Associated with Contract Period and Failure Process," *IEEE LATAM*, vol. 8, no. 4, pp. 410–416, 2010.

[5] ——, "A Study of the Interval Availability and Its Impact on SLAs Risk," in *Advances in Computer Science, Engineering & Applications*. Springer, 2012, pp. 879–890.

[6] K. Rusek, P. Guzik, and P. Chołda, "Effective Risk Assessment in Resilient Communication Networks," *Journal of Network and Systems Management*, vol. 24, no. 3, pp. 491–515, 2016.

[7] L. Mastroeni and M. Naldi, "Compensation Policies and Risk in Service Level Agreements: A Value-at-Risk approach under the ON-OFF Service Model," in *International Workshop on Internet Charging and QoS Technologies*. Springer, 2011, pp. 2–13.

[8] ——, "Violation of Service Availability Targets in Service Level Agreements," in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2011, pp. 537–540.

[9] D. A. Mello, H. Waldman, and G. S. Quitério, "Interval availability estimation for protected connections in optical networks," *Computer Networks*, vol. 55, no. 1, pp. 193–204, 2011.

[10] W. Ni, J. Wu, C. Huang, and M. Savoie, "Analytical models of flow availability in two-layer networks with dedicated path protection," *Optical Switching and Networking*, vol. 10, no. 1, pp. 62–76, 2013.

[11] L. Takács, "On certain sojourn time problems in the theory of stochastic processes," *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 8, pp. 169–191, 1957.

[12] A. Mykkeltveit and B. E. Helvik, "Adaptive Management of Connections to Meet Availability Guarantees in SLAs," in *2009 IFIP/IEEE International Symposium on Integrated Network Management*. IEEE, 2009, pp. 545–552.

[13] S. Verbrugge, D. Colle, P. Demeester, R. Huelsermann, and M. Jaeger, "General Availability Model for Multilayer Transport Networks," in *Proceedings of the 5th International Workshop on Design of Reliable Communication Networks (DRCN 2005)*. IEEE, 2005, pp. 85–92.

[14] A. Birolini, *Reliability Engineering: Theory and Practice*, 8th ed. Berlin, Germany: Springer-Verlag, 2017.

[15] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*, ser. American Mathematical Society Colloquium Publications. Providence, RI, USA: American Mathematical Society, 1996, vol. 31.

[16] P. Virtanen *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[17] F. Johansson *et al.*, *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.1.0)*, December 2018.

[18] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski, "SNDlib 1.0–Survivable Network Design Library," *Networks: An International Journal*, vol. 55, no. 3, pp. 276–286, 2010.