

Article

Service Level Agreement Violations in Cloud Storage: Insurance and Compensation Sustainability

Loretta Mastroeni ^{1,†}, Alessandro Mazzocchi ^{2,†} and Maurizio Naldi ^{2,3,*,†} 

¹ Department of Economics, University of Roma Tre, 00145 Rome, Italy

² Department of Civil Engineering and Computer Science, University of Roma Tor Vergata, 00133 Rome, Italy

³ Department of Law, Economics, Politics and Modern languages, LUMSA University, 00192 Rome, Italy

* Correspondence: maurizio.naldi@uniroma2.it

† These authors contributed equally to this work.

Received: 22 May 2019; Accepted: 25 June 2019; Published: 30 June 2019



Abstract: Service Level Agreements are employed to set availability commitments in cloud services. When a violation occurs as in an outage, cloud providers may be called to compensate customers for the losses incurred. Such compensation may be so large as to erode cloud providers' profit margins. Insurance may be used to protect cloud providers against such a danger. In this paper, closed formulas are provided through the expected utility paradigm to set the insurance premium under different outage models and QoS metrics (no. of outages, no. of long outages, and unavailability). When the cloud service is paid through a fixed fee, we also provide the maximum unit compensation that a cloud provider can offer so as to meet constraints on its profit loss. The unit compensation is shown to vary approximately as the inverse square of the service fee.

Keywords: cloud storage; service level agreements; insurance; compensations; refunds

1. Introduction

Service Level Agreements (SLA) are typically employed by cloud providers to provide guarantees on the quality of their services to end customers [1–6]. Those agreements, which form a contract or a contract-like relationship between the parties, state the obligations imposed on the cloud provider, through a set of service quality metrics and constraints to be met.

The compliance of cloud providers with those contractual or contractual-like commitments has to be monitored, and several tools have been proposed in the literature for that purpose [7–9] or for acting on the basis of monitoring results to reach for compliance [10]. Any tool has to measure a set of parameters related to the quality of service and compare their values against SLAs' provisions. The foremost among them is availability [11], defined roughly as the percentage of time that the cloud is available.

As it happens, the obligations underwritten in SLAs may not be fulfilled. In that case, the cloud provider is expected to pay penalties, if the contract so prescribes [12], or anyway to compensate its customers for the loss incurred. If violations take place on a wide scale, penalties and compensations may endanger the economic balance of the cloud provider.

The cloud provider should therefore protect itself against such an occurrence. It can basically act in two ways:

- reduce the risk due to unavailability periods;
- be indemnified against possible losses.

The first countermeasure is technical in nature. Reducing the probability of occurrence and/or the amount of damages calls for investments in the infrastructure and better cloud management processes.

The second countermeasure instead relies on buying an insurance policy, whereby the insurer takes on the risk of paying all claims by customers in return for the premium to be paid by the cloud provider, which acts as the insured. The insurance approach to deal with possible excessive claims by cloud users in the framework of a Service Level Agreement has been proposed in [13]. A major problem lies in the definition of the right premium of an insurance policy. In that paper, the correct premium has been computed for the case where the outage phenomenon is described by a Poisson process for occurrence of outages and by the generalized Pareto distribution for their duration. However, an alternative model has been recently introduced in the literature, based on an extensive measurement campaign on a private cloud [14] belonging to a small/medium company. In that paper, the interarrival time between successive outages is modelled through a Pareto distribution, while the duration of an outage is modelled by a lognormal distribution.

In this paper, we again consider the problem of correctly determining the insurance premium against cloud outages. We adopt the expected utility paradigm as in [13], but we consider the outage model recently proposed in [14] in addition to the exponential-Pareto (a.k.a. Poisson–Pareto) examined in [13]. Both outage models are here examined under three QoS metrics. Our main new results are the following:

- we obtain closed formulas for the insurance premiums for all the combinations of outage model and QoS metric (rather than just for the exponential-Pareto model as in [13]);
- we obtain a closed formula to set the unit compensation offered to customers so as to avoid compensations (through the resulting premium) eroding the profit margins;
- we show that the maximum unit compensation is related to the service fee through an approximate inverse square law.

The paper is structured as follows. Section 2 is devoted to describing the two models for outages. The quality of service metrics are instead described in Section 3. The statistics of losses resulting from the models of outages are computed in Section 4. The fair premium is derived in Section 5. Finally, the conditions to make the refund process sustainable are derived in Section 6.

2. Models for Outages

Our analysis relies on the use of models that describe the behaviour of the cloud. Here, we rely on a simple ON-OFF description of the service offered by the cloud, which is well suitable for cloud storage. In the ON-OFF model, the cloud alternates between two states: a fully operating period (a.k.a. uptime or ON period) and a no-service period (a.k.a. downtime or OFF period), as shown in Figure 1. No partial performance degradation is considered in the ON-OFF model. The ON-OFF model can also serve as an approximation to more complex quality degradation processes. For example, a cloud storage provider could still be able to deliver the required files to its customers, but with a longer delivery time, i.e., with a delay with respect to normal operating conditions. In that case, a threshold could be imposed on the delivery time so that delays longer than that threshold are considered as service outages. In this section, we describe two models proposed in the literature for the statistics related to the alternance of ON and OFF states. We employ two variables to describe the duration of the two states: we use S for the duration of the ON period and D for that of the OFF period.

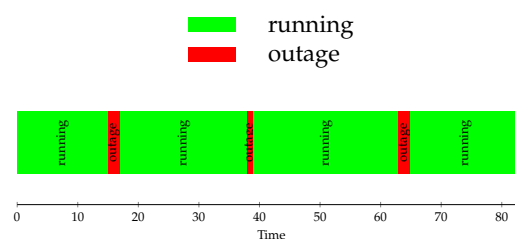


Figure 1. Cloud state sequence.

2.1. The Exponential-Pareto Model

The Exponential-Pareto model owes its name to the distributions proposed for the two alternating states of the general ON-OFF model. Namely, we have an exponential distribution for the duration of the ON period. The duration of the OFF period is instead governed by a generalized Pareto distribution (GPD). The cumulative distribution functions are therefore

$$F_S(x) = \mathbb{P}[S < x] = 1 - \lambda e^{-\lambda x}, \tag{1}$$

$$F_D(x) = \mathbb{P}[D < x] = \begin{cases} 1 - (1 + \xi x / \beta)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - e^{-x/\beta}, & \text{if } \xi = 0, \end{cases} \tag{2}$$

where β is the scale parameter and ξ is the shape parameter of the GPD.

Since the duration of the ON period is the time elapsing before a new failure occurs, the MTTF (mean-time-to-failure) is simply the reciprocal of the parameter of the exponential distribution

$$\text{MTTF} = \mathbb{E}[S] = \frac{1}{\lambda}. \tag{3}$$

Similarly, the duration of the OFF period is the time needed to repair the cloud, so that the MTTR (mean-time-to-repair) is the expected value of D

$$\text{MTTR} = \mathbb{E}[D] = \frac{\beta}{1 - \xi}. \tag{4}$$

The Exponential-Pareto model has first been proposed in [15] through a best-fit procedure on a dataset of customer-reported outages for five major cloud providers (Google, Amazon, Rackspace, Salesforce, Windows Azure). The sources of data were Cloutage (cloutage.org), founded by the Open Security Foundation in April 2010 but now discontinued, and the International Working Group on Cloud Computing Resiliency (IWGCR, a working group with a mission to monitor and analyze cloud computing resiliency).

This model has been employed in [13] to assess the sustainability of refunds linked to insurance contracts for cloud services, as well as in [16] to assess the contribution of the network to the overall unavailability. The values obtained for the parameters of the two distributions in that case are reported in Table 1.

Table 1. Parameters in the Exponential-Pareto model.

Provider	Exponential	Pareto	
	1/λ [Days]	β	ξ
Google	27.53	405.29	0.39
Amazon	85.6	276.43	−0.12
Rackspace	7.78	381.19	0.3
Salesforce	8.56	192.47	−0.64
Windows Azure	36.67	312.32	−0.35

2.2. The Pareto-Lognormal Model

As the name suggests, in the Pareto-Lognormal model the distribution of operating periods is modelled as a Pareto law, while the lognormal distribution describes the duration of outages.

The model has been proposed by Dunne and Malone [14], who analysed 300+ cloud outage events recorded for an enterprise cloud system (not better identified, but belonging to a small/medium enterprise) over an eighteen-month period. They tested seven models (Pareto, loglogistic, lognormal, gamma, exponential, logistic, and Weibull) through the Anderson–Darling test. In the end, they found the Pareto-Lognormal couple to be the best-fit solution.

The cumulative distribution function of operating periods is then

$$F_S(x) = \begin{cases} 1 - \left(\frac{h}{x}\right)^\alpha, & \text{if } x \geq h, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The distribution of outage duration is instead

$$F_D(x) = \begin{cases} G\left(\frac{\ln(w)-\mu}{\sqrt{2}\sigma}\right), & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The reliability indicators (MTTF and MTTR) are related to the parameters of the two distributions by the following formulas:

$$\begin{aligned} \text{MTTF} &= \frac{\alpha}{\alpha - 1}h, \\ \text{MTTR} &= e^{\mu + \frac{\sigma^2}{2}}. \end{aligned} \tag{7}$$

The values found by Dunne and Malone for the parameters of the two distributions are reported in Table 2 (where the abbreviation SME—Small/Medium Enterprise—stands for the company whose cloud has been monitored during the measurement campaign, but whose name has been undisclosed); they lead to

$$\begin{aligned} \text{MTTF} &= 2445 \text{ min,} \\ \text{MTTR} &= 227 \text{ min.} \end{aligned} \tag{8}$$

Table 2. Parameters in the Pareto-Lognormal model.

Provider	Pareto		Lognormal	
	<i>h</i>	α	μ	σ
SME	1834	4	4.58	1.3

3. Quality of Service Metrics

The assessment of damages and the ensuing compensation must be related to a metric that embodies the quality of service underwritten in the SLA. In this section, we provide three different QoS metrics with specific reference to the performance of the cloud. We adopt the same metrics already considered in [17] to derive the correct price for an insurance policy in the case of a Markovian system, and in [18] to assess the Value-at-Risk, again in the case of a Markovian system.

We consider the following metrics:

- Number of outages;
- Number of long outages;
- Unavailability.

The number of outages is computed as the number of no-service intervals during the observation period *T*. There are a number of technicalities involved in the detection of the actual instants of time when the cloud goes OFF and then goes back ON. We do not go into detail here and refer the reader interested to delve into the matter to [19], where the complexity of measurements is dealt with. An advantage of this metric is that even the smallest service interruptions are accounted for; on the other hand, the severity of outages is not considered.

The number of long outages is simply the number of outages lasting more than a pre-set threshold *w*. The rationale for this metric is to avoid considering very short interruptions of service (glitches). Such interruptions could add considerably to the number of outages (hence, the first metric) without

being significant for the user. This metric is a good choice if an outage must last some time to provoke a real damage to the company; this means that the company relies on some form of short-term resiliency against outages (this happens, e.g., with TCP where failures at the network layer can easily be recovered if they do not extend too much in space and time).

Finally, the unavailability is the percentage of time the service is OFF. According to this metric, a single long outage counts as a set of shorter outages if their cumulative duration is equal. In this case, damages simply accrue over time without any start effect.

4. Loss Statistics

In Section 3, we have introduced three outage-related metrics to describe the quality of the service delivered by the cloud provider. Any outage turns into an economic loss for the customer, which must be accounted for when setting the insurance premium. In this section, we evaluate the loss that a customer may claim under the QoS metric of choice for the two models at hand. In all cases, we assume a proportional relationship between the QoS metric and the claimed loss. The values depend on the statistical models assumed for the outages; we analyse the cases of the Exponential-Pareto model and the Pareto-Lognormal one in two separate subsections.

4.1. The Exponential-Pareto Model

In this section, we derive the loss statistics, as described by the first two moments of the economic loss, for the three QoS metrics defined in Section 3.

Number of Outages

If the number of outages is adopted as the QoS metric, and an economic loss k_f is claimed for each outage, the overall economic loss X_f is

$$X_f = k_f N_f, \tag{9}$$

which is a random variable, since the number N of outages occurring during a fixed amount of time is itself random. Over the observation period of duration T , its expected value and variance are respectively

$$\begin{aligned} \mathbb{E}[X_f] &= k_f \lambda T, \\ \mathbb{V}[X_f] &= k_f^2 \lambda T. \end{aligned} \tag{10}$$

4.2. Number of Long Outages

By long outages, we mean outages whose duration exceeds a given threshold w . If the economic loss X_{lf} for each long outage is k_{lf} , the overall loss is proportional to the number N_{lf} of long outages:

$$X_{lf} = k_{lf} N_{lf} = k_{lf} \sum_{i=1}^N I_{[D_i > w]}, \tag{11}$$

where $I_{[y]}$ is the indicator function, which takes the value 1 if the condition y is satisfied, and the value 0, otherwise. The indicator function is actually a random variable with a Bernoulli distribution, with its first two moments given by

$$\begin{aligned} \mathbb{E}[I_{[D_i > w]}] &= \mathbb{P} [I_{[D_i > w]} = 1] = 1 - \mathbb{P}[D_i < w] \\ &= 1 - \left[1 - \left(1 + \frac{\xi w}{\beta} \right)^{-\frac{1}{\xi}} \right] = \left(1 + \frac{\xi w}{\beta} \right)^{-\frac{1}{\xi}}, \end{aligned} \tag{12}$$

$$\begin{aligned}
 \mathbb{V}[I_{[D_i>w]}] &= \mathbb{P} \left[I_{[D_i>w]} = 1 \right] \left(1 - \mathbb{P} \left[I_{[D_i>w]} = 1 \right] \right) \\
 &= \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \left[1 - \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \right] \\
 &= \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} - \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{2}{\zeta}}.
 \end{aligned}
 \tag{13}$$

Since Equation (11) is a random sum of independent Bernoulli variables, we can obtain the probability distribution of the economic loss (which is now a discrete variable whose possible values are multiples of the unit loss per long outage k_{lf}):

$$\begin{aligned}
 \mathbb{P}[X_{lf} = j k_{lf}] &= \sum_{n=1}^{\infty} \mathbb{P} \left[\sum_{i=0}^n I_{[D_i>w]} = j \mid N_T = n \right] \mathbb{P}[N_T = n] \\
 &= \sum_{n=1}^{\infty} \binom{n}{j} \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{j}{\zeta}} \\
 &\quad \times \left[1 - \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \right]^{n-j} \times \frac{(\lambda T)^n}{n!} e^{-\lambda T}.
 \end{aligned}
 \tag{14}$$

We can now use Wald’s identities to obtain a closed form expression. We assume that there is only a weak correlation among the number of summands in the sum (11) and each of the summands: the duration of an outage (e.g., its exceeding w) is independent of the number of outages, while the number of outages is very weakly correlated with the outage duration, as long as we consider services with high availability.

By resorting to Wald’s identities (see, e.g., Section 34.14.2.11 of [20] or Section 1.7.3 of [21]), we can therefore compute the mean and the variance of the economic loss

$$\begin{aligned}
 \mathbb{E}[X_{lf}] &= k_{lf} \mathbb{E}[N_T] \mathbb{E} \left[I_{[D_i>w]} \right] = k_{lf} \lambda T \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}}, \\
 \mathbb{V}[X_{lf}] &= k_{lf}^2 \left\{ \mathbb{E}^2 \left[I_{[D_i>w]} \right] \mathbb{V}[N_T] + \mathbb{V} \left[I_{[D_i>w]} \right] \mathbb{E}[N_T] \right\} \\
 &= k_{lf}^2 \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{2}{\zeta}} \lambda T + k_{lf}^2 \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \left[1 - \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \right] \lambda T \\
 &= k_{lf}^2 \lambda T \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \left\{ \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} + \left[1 - \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \right] \right\} \\
 &= k_{lf}^2 \lambda T \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}}.
 \end{aligned}
 \tag{15}$$

4.3. Unavailability

Now, the economic loss X_u is proportional to the cumulative unavailability time U over the period T , with a loss k_u for each time unit during which the service is unavailable:

$$X_u = k_u U = k_u \sum_{i=1}^N D_i,
 \tag{16}$$

where k_u is the loss per unit time. The value of k_u depends of course on the actual company’s business. In Table 3, an estimate of the loss suffered for each minute of service unavailability is reported, as extracted from [22]. Those values can be taken as values for k_u when the durations D_i ’s are expressed in minutes. As can be seen, those values exhibit an extremely wide dispersion.

Table 3. Unit loss.

Company	Loss per Minute [US\$]
EdiActivity	1
GXS	913
Salesforce	7743
eBay	30,536
Southwest Airlines	35,407
Google	113,641
Amazon	141,647

We have again a random sum, for which we can apply Wald’s identities, following the same considerations introduced in Section 4.2. The mean economic loss and its variance are respectively

$$\begin{aligned} \mathbb{E}[X_u] &= k_u \mathbb{E}[U] = k_u \mathbb{E}[D_i] \mathbb{E}[N_T] \\ &= k_u \frac{\beta}{1 - \xi} \lambda T, \end{aligned} \tag{17}$$

$$\begin{aligned} \mathbb{V}[X_u] &= k_u^2 \mathbb{V}[U_T] = k_u^2 \left\{ \mathbb{E}^2[D_i] \mathbb{V}[N_T] + \mathbb{V}[D_i] \mathbb{E}[N_T] \right\} \\ &= k_u^2 \left[\frac{\beta^2}{(1 - \xi)^2} \lambda T + \lambda T \frac{\beta^2}{(1 - \xi)^2 (1 - 2\xi)} \right] \\ &= k_u^2 \frac{\beta^2}{(1 - \xi)^2} \lambda T \left(1 + \frac{1}{1 - 2\xi} \right) \\ &= k_u^2 \frac{\beta^2}{(1 - \xi)^2} \lambda T \frac{2 - 2\xi}{1 - 2\xi} \\ &= \frac{2k_u^2 \beta^2 \lambda T}{(1 - \xi)(1 - 2\xi)}. \end{aligned} \tag{18}$$

4.4. The Pareto-Lognormal Model

Similarly to what we have done for the Exponential-Pareto model, in this subsection, we derive the loss statistics for the Pareto-Lognormal model.

4.4.1. Number of Outages

In order to derive the expected value and the variance of the number of outages N_f , we first derive its probability distribution.

We describe the behaviour of the cloud during the observation interval T through the two sequence of alternating operating periods and outages:

$$\begin{aligned} \mathcal{S} &= \{S_1, S_2, \dots, S_i, \dots\}, \\ \mathcal{D} &= \{D_1, D_2, \dots, D_i, \dots\}, \end{aligned} \tag{19}$$

where all the S_i ’s are i.i.d.random variables following a Pareto distribution, while the D_i ’s are i.i.d. random variables following a lognormal distribution. The details of the two distribution are described in Section 2.2.

We start by considering the probability that there are no outages during the observation interval, which is tantamount to saying that the first operating period S_1 lasts more than the observation interval:

$$\mathbb{P}[N_f = 0] = \mathbb{P}(S_1 > T) = \left(\frac{h}{T} \right)^\alpha. \tag{20}$$

The probability that a single outage takes place over T is instead the probability that there is at least an outage minus the probability that there are at least two outages:

$$\begin{aligned} \mathbb{P}[N_f = 1] &= \mathbb{P}[N_f \geq 1] - \mathbb{P}[N_f \geq 2] \\ &= 1 - \mathbb{P}(N_f = 0) - \mathbb{P}[S_1 + D_1 + S_2 < T]. \end{aligned} \tag{21}$$

Similarly, we have

$$\begin{aligned} \mathbb{P}[N_f = 2] &= 1 - \mathbb{P}[N_f = 0] - \mathbb{P}(N_f = 1) \\ &\quad - \mathbb{P}[S_1 + D_1 + S_2 + D_2 + S_3 < T], \end{aligned} \tag{22}$$

so that we can write the general recursive expression

$$\begin{aligned} \mathbb{P}[N_f = i] &= 1 - \sum_{j=0}^{i-1} \mathbb{P}[N_f = j] - \mathbb{P}\left[\sum_{k=1}^{i+1} S_k + \sum_{k=1}^i D_k < T\right] \\ i &= 1, 2, \dots \end{aligned} \tag{23}$$

Unfortunately, Equation (23) involves the distribution of a sum of Pareto and lognormal random variables, for which no simple expression is known. However, since the third term in that equation involves a number of variables, we can define the sequence of variables

$$Z_i = \sum_{k=1}^{i+1} S_k + \sum_{k=1}^i D_k \quad i = 1, 2, \dots \tag{24}$$

and invoke the central limit theorem for it, so that we approximate the distribution of each Z_i by a normal distribution with moments

$$\begin{aligned} \mathbb{E}[Z_i] &= \sum_{k=1}^{i+1} \mathbb{E}[S_k] + \sum_{k=1}^i \mathbb{E}[D_k] = (i + 1)\mathbb{E}[S_1] + i\mathbb{E}[D_1] \\ &= (i + 1)\text{MTTF} + i \cdot \text{MTTR} \\ &= (i + 1) \frac{\alpha}{\alpha - 1} h + i \cdot e^{\mu + \frac{\sigma^2}{2}}, \end{aligned} \tag{25}$$

$$\begin{aligned} \mathbb{V}[Z_i] &= \sum_{k=1}^{i+1} \mathbb{V}[S_k] + \sum_{k=1}^i \mathbb{V}[D_k] = (i + 1)\mathbb{V}[S_1] + i\mathbb{V}[D_1] \\ &= (i + 1) \frac{\alpha}{(\alpha - 1)^2(\alpha - 2)} h^2 + i(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}, \end{aligned} \tag{26}$$

after recalling Equation (7).

The probability distribution of the number of outages over a period T can therefore be computed through the following general recursive expression, derived from Equation (23) with the starting point given by Equation (20):

$$\mathbb{P}[N_f = i] \simeq 1 - \sum_{j=0}^{i-1} \mathbb{P}[N_f = j] - G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \quad i = 1, 2, \dots \tag{27}$$

$G(\cdot)$ being the cumulative distribution function of the standard normal variable. Rolling out the recursive expression we get

$$\begin{aligned}
 \mathbb{P}[N_f = 1] &= 1 - \left(\frac{h}{T}\right)^\alpha - G\left(\frac{T - \mathbb{E}[Z_1]}{\sqrt{\mathbb{V}[Z_1]}}\right) \\
 \mathbb{P}[N_f = 2] &= G\left(\frac{T - \mathbb{E}[Z_1]}{\sqrt{\mathbb{V}[Z_1]}}\right) - G\left(\frac{T - \mathbb{E}[Z_2]}{\sqrt{\mathbb{V}[Z_2]}}\right) \\
 \mathbb{P}[N_f = 3] &= G\left(\frac{T - \mathbb{E}[Z_2]}{\sqrt{\mathbb{V}[Z_2]}}\right) - G\left(\frac{T - \mathbb{E}[Z_3]}{\sqrt{\mathbb{V}[Z_3]}}\right) \\
 &\dots
 \end{aligned}
 \tag{28}$$

so that we end up with the general expression

$$\begin{aligned}
 \mathbb{P}[N_f = i] &= G\left(\frac{T - \mathbb{E}[Z_{i-1}]}{\sqrt{\mathbb{V}[Z_{i-1}]}}\right) - G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \\
 &i = 2, 3, \dots
 \end{aligned}
 \tag{29}$$

We can now compute the first two moments of the number of outages over the interval T . The expected value is

$$\begin{aligned}
 \mathbb{E}[N_f] &= \sum_{i=0}^{\infty} i\mathbb{P}[N_f = i] \\
 &= 1 - \left(\frac{h}{T}\right)^\alpha - G\left(\frac{T - \mathbb{E}[Z_1]}{\sqrt{\mathbb{V}[Z_1]}}\right) + \sum_{i=2}^{\infty} i\mathbb{P}[N_f = i] \\
 &= 1 - \left(\frac{h}{T}\right)^\alpha - G\left(\frac{T - \mathbb{E}[Z_1]}{\sqrt{\mathbb{V}[Z_1]}}\right) + \\
 &\quad \sum_{i=2}^{\infty} i \left[G\left(\frac{T - \mathbb{E}[Z_{i-1}]}{\sqrt{\mathbb{V}[Z_{i-1}]}}\right) - G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \\
 &= 1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right),
 \end{aligned}
 \tag{30}$$

where $\mathbb{E}[Z_i]$ and $\mathbb{V}[Z_i]$ are provided respectively by Equations (25) and (26).

Similarly, the variance is

$$\begin{aligned}
 \mathbb{V}[N_f] &= \sum_{i=0}^{\infty} (i - \mathbb{E}[N_f])^2 \mathbb{P}[N_f = i] \\
 &= \sum_{i=0}^{\infty} i^2 \mathbb{P}[N_f = i] - 2\mathbb{E}[N_f] \sum_{i=0}^{\infty} i\mathbb{P}[N_f = i] + \mathbb{E}^2[N_f] \sum_{i=0}^{\infty} \mathbb{P}[N_f = i] \\
 &= \sum_{i=0}^{\infty} i^2 \mathbb{P}[N_f = i] - \mathbb{E}^2[N_f] \\
 &= 1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} (2i + 1)G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) - \mathbb{E}^2[N_f],
 \end{aligned}
 \tag{31}$$

where, again, $\mathbb{E}[Z_i]$ and $\mathbb{V}[Z_i]$ are provided respectively by Equations (25) and (26).

It is now fairly straightforward to get the first two moments of the economic loss X :

$$\mathbb{E}[X_f] = k_f \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right], \tag{32}$$

$$\mathbb{V}[X_f] = k_f^2 \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} (2i + 1)G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) - \mathbb{E}^2[N_f] \right]. \tag{33}$$

4.4.2. Number of Long Outages

For the expected number of long outages, we go back to the general expression in Equation (11). Since the number of outages is a random variable, we can again exploit the product of averages. After recalling Equation (30) for the expected number of outages and the tail of the lognormal distribution, we have

$$\begin{aligned} \mathbb{E}[N_{lf}] &= \mathbb{E}[N_f] \mathbb{E}[\mathbb{I}_{[D_i > w]}] = \mathbb{E}[N_f] \mathbb{P}(D_i > w) \\ &= \mathbb{E}[N_f] \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma}\right) \right] \\ &= \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \\ &\times \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma}\right) \right]. \end{aligned} \tag{34}$$

As to the variance, we adopt the same approach as in Section 4.2:

$$\begin{aligned} \mathbb{V}[N_{lf}] &= \mathbb{E}^2[\mathbb{I}_{[D_i > w]}] \mathbb{V}ar[N_f] + \mathbb{V}[\mathbb{I}_{[D_i > w]}] \mathbb{E}[N_f]^2 \\ &= \mathbb{P}^2(D_i > w) \mathbb{V}[N_f] + (1 - \mathbb{P}(D_i > w)) \mathbb{P}(D_i > w) \mathbb{E}[N_f]^2 \\ &= \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma}\right) \right]^2 \\ &\times \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} (2i + 1)G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) - \mathbb{E}[N_f]^2 \right] \\ &+ \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma}\right) \right] \left[G\left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma}\right) \right] \\ &\times \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right]. \end{aligned} \tag{35}$$

Again, we can now easily get the expected value and the variance of the economic loss X :

$$\mathbb{E}[X_{lf}] = k_{lf} \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma}\right) \right], \tag{36}$$

$$\begin{aligned}
 \mathbb{V}[X_{lf}] &= k_{lf}^2 \left[1 - G \left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma} \right) \right]^2 \\
 &\times \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} (2i + 1) G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) - \mathbb{E}[N_f]^2 \right] \\
 &+ k_{lf}^2 \left[1 - G \left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma} \right) \right] \left[G \left(\frac{\ln(w) - \mu}{\sqrt{2}\sigma} \right) \right] \\
 &\times \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) \right].
 \end{aligned} \tag{37}$$

4.4.3. Unavailability

We adopt the same approach as in Section 4.3. We obtain:

$$\mathbb{E}[U] = \mathbb{E}[N_f] \mathbb{E}[D_i] = \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) \right] e^{\mu + \frac{\sigma^2}{2}}, \tag{38}$$

$$\begin{aligned}
 \mathbb{V}[U] &= \mathbb{E}^2[D_i] \mathbb{V}[N_f] + \mathbb{V}[D_i] \mathbb{E}[N_f] = e^{2\mu + \sigma^2} \\
 &\times \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} (2i + 1) G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) - \mathbb{E}^2[N_f] \right] \\
 &+ \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) \right] e^{2\mu + \sigma^2} (e^{\sigma^2} - 1).
 \end{aligned} \tag{39}$$

The corresponding moments of the economic loss X are then:

$$\mathbb{E}[X_u] = k_u \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) \right] e^{\mu + \frac{\sigma^2}{2}}, \tag{40}$$

$$\begin{aligned}
 \mathbb{V}[X_u] &= k_u^2 e^{2\mu + \sigma^2} \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} (2i + 1) G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) - \mathbb{E}^2[N_f] \right] \\
 &+ k_u^2 \left[1 - \left(\frac{h}{T} \right)^\alpha + \sum_{i=1}^{\infty} G \left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}} \right) \right] e^{2\mu + \sigma^2} (e^{\sigma^2} - 1).
 \end{aligned} \tag{41}$$

5. Insurance Pricing

In Section 4, we have seen how outages result in economic losses for the cloud provider. These losses have been evaluated for three quality metrics (number of outages, number of long outages, and unavailability), which account for most of the cases that may be encountered in practice. Since these losses may endanger the economic balance of the cloud provider, the latter may wish to hedge against such losses. A natural way to hedge is through an insurance policy. In this section, we derive the fair price for such a policy. We first describe the general pricing methodology based on the expected utility approach as described in [23] and then apply that approach to the failure models described in Section 2 and the QoS metrics described in Section 3.

We consider a cloud provider whose assets are worth $\omega \in R^+$. This provider faces a possible monetary loss $X \in R^+$ and wishes to buy an insurance policy. In that case, it has to pay the insurance premium P to be fully indemnified against that loss.

We assume that the cloud provider's perception of the loss of the amount of money x is described by its utility function $u(x)$. The perception may be dramatically different, depending on the economic conditions of the insured cloud provider. For example, if even a modest loss changes the overall economic balance from a profit condition to a loss one, even that small amount of money leads to a

large decay of the utility. On the other hand, if the cloud provider is making a large profit, it will be able to withstand larger losses without a significant reduction of its utility.

Since the loss X is a random variable, we must consider the expected utility of the residual value of the company's assets when the loss X is suffered. We have to compare the provider's utility under two conditions: incurring the uncertain loss versus paying the certain insurance premium. The maximum tolerable premium P is the value that makes the two utilities equal. The equilibrium equation is then

$$\mathbb{E}[u(\omega - X)] = u(\omega - P). \tag{42}$$

The equilibrium equation may not be amenable to a closed form solution, depending on the nonlinear nature of the utility function. We can find an approximate solution by expanding both terms through a Taylor series in the neighbourhood of $\omega - \mathbb{E}[X]$:

$$\begin{aligned} u(\omega - P) &\simeq u(\omega - \mathbb{E}[X]) + (\mathbb{E}[X] - P)u'(\omega - \mathbb{E}[X]), \\ u(\omega - X) &\simeq u(\omega - \mathbb{E}[X]) + (\mathbb{E}[X] - X)u'(\omega - \mathbb{E}[X]), \\ &+ \frac{(\mathbb{E}[X] - X)^2}{2}u''(\omega - \mathbb{E}[X]). \end{aligned} \tag{43}$$

By employing the second of these Taylor expansions, we get

$$\begin{aligned} \mathbb{E}[u(\omega - X)] &\simeq u(\omega - \mathbb{E}[X]) + (\mathbb{E}[X] - \mathbb{E}[X])u'(\omega - \mathbb{E}[X]) \\ &+ \frac{\mathbb{V}[X]}{2}u''(\omega - \mathbb{E}[X]) \\ &= u(\omega - \mathbb{E}[X]) + \frac{\mathbb{V}[X]}{2}u''(\omega - \mathbb{E}[X]). \end{aligned} \tag{44}$$

Finally, by replacing those expressions in the equilibrium Equation (42), we obtain the maximum tolerable premium

$$P \simeq \mathbb{E}[X] + \frac{\mathbb{V}[X]}{2}r(\omega - \mathbb{E}[X]), \tag{45}$$

where we have introduced the risk aversion function r , which takes into account the effect of the utility function:

$$r(x) = -\frac{u''(x)}{u'(x)}. \tag{46}$$

Though several options are possible for the utility function (see, e.g., Section 1.3 in [23]), a typical assumption is to consider a constant risk aversion coefficient

$$r(x) = \delta > 0. \tag{47}$$

In that case, the utility function exhibits the Constant Absolute Risk Aversion (CARA) property [24]; the only function that satisfies the CARA property is the exponential function $u(x) = 1 - e^{-\delta x}$, which leads to the maximum tolerable premium

$$P \simeq \mathbb{E}[X] + \delta \frac{\mathbb{V}[X]}{2}. \tag{48}$$

The CARA property has been assumed, e.g., in [13,25].

Under the CARA property, the premium formulation is therefore of the mean-variance type, where the risk-aversion coefficient is to be defined to obtain the premium: higher values of δ are associated with a growing aversion for risk and then to the willingness to pay a higher premium: in [26], the risk-aversion coefficient is assumed to take values in the [0.5,4] range. However, its assessment is rather sensitive, since it can significantly contribute to the premium. In [27], a procedure is indicated to compute its value for a CARA utility function. Namely, it is set as

$$\delta = \frac{\ln \frac{1+2\eta}{1-2\eta}}{\mathbb{E}[X]}, \tag{49}$$

where η is the probability premium, i.e., the increase in probability above 0.5 that an individual requires to maintain a constant level of utility equal to the utility of status quo; it takes a value in the $[0, 0.5]$ range.

In the following subsections, we derive the pricing formula for the two models described in Section 2.

5.1. The Exponential-Pareto Model

Following the same approach adopted for the computation of the economic loss in Section 4, we evaluate the resulting insurance premium for the three QoS metrics. In order to get a numeric feeling for the outcome, we consider the parameter values shown in Table 4, which are fairly representative of the experiments leading to the models of Section 2.

Table 4. Sample parameter values for the Exponential-Pareto model.

Exponential	Pareto		Time
$\frac{1}{\lambda}$ [Days]	ζ	β	T [Days]
27.5	0.4	405	365

5.1.1. Number of Outages

Replacing the results of Equation (10) in Equation (48), we get the premium (actually the maximum tolerable premium) for the coverage period T when the loss is proportional to the number of outages:

$$P_f = k_f \lambda T + \frac{\delta k_f^2}{2} \lambda T = k_f \lambda T \left(1 + \frac{\delta k_f}{2} \right). \tag{50}$$

The resulting premium is proportional to the average number of outages and a quadratic function of the unit loss per outage.

5.1.2. Number of Long Outages

If the loss is proportional to the number of outages whose duration exceeds the threshold w , after recalling Equation (15), we obtain the premium

$$\begin{aligned} P_{lf} &= k_{lf} \lambda T \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} + \frac{\delta}{2} k_{lf}^2 \lambda T \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \\ &= k_{lf} \lambda T \left(1 + \frac{\zeta w}{\beta} \right)^{-\frac{1}{\zeta}} \left(1 + \frac{\delta k_{lf}}{2} \right). \end{aligned} \tag{51}$$

Similarly to what we found for the case where the QoS metric is the number of outages, here the premium is a linear function of the average number of outages and a quadratic function of the unit loss per long outage. Instead, in Figure 2, we see that the premium is a nearly linear function of the threshold w that defines long outages. The picture has been obtained for the values of Table 4 when the loss per minute is $k_U = 10,000\$$, which lies in the low range among the values reported in Table 3. The unit loss per long outage has been set according to the equation

$$k_{lf} \mathbb{E}[N_{lf}] = k_U \mathbb{E}[U]. \tag{52}$$

Aside from its absolute value, we can get a feeling for the size of the premium if we compare it with the expected loss, i.e., the excess premium defined as

$$EP_{lf} = \frac{P_{lf} - \mathbb{E}[X_{lf}]}{\mathbb{E}[X_{lf}]} = \frac{\delta k_{lf}}{2}. \tag{53}$$

The excess premium is shown in Figure 3 for the same case of Figure 2. We can see that the coverage of risk requires an excess premium of the order of some percentage points; its dependence on the outage duration threshold w is again roughly linear.

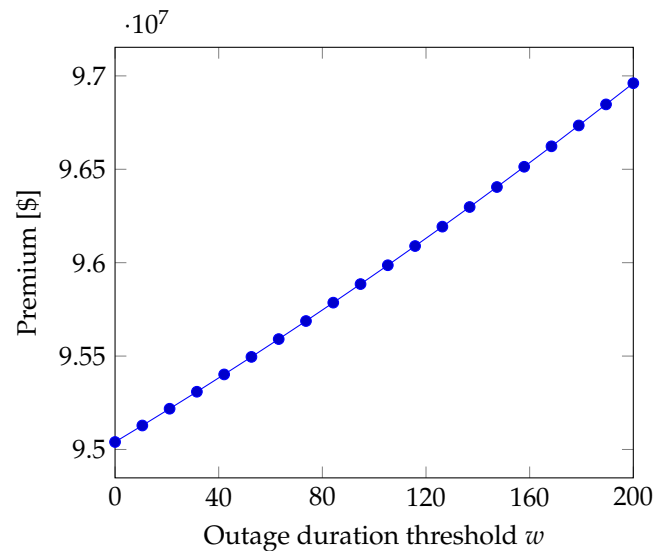


Figure 2. Impact of the outage duration threshold on the premium in the case of long outages (Exponential-Pareto model).

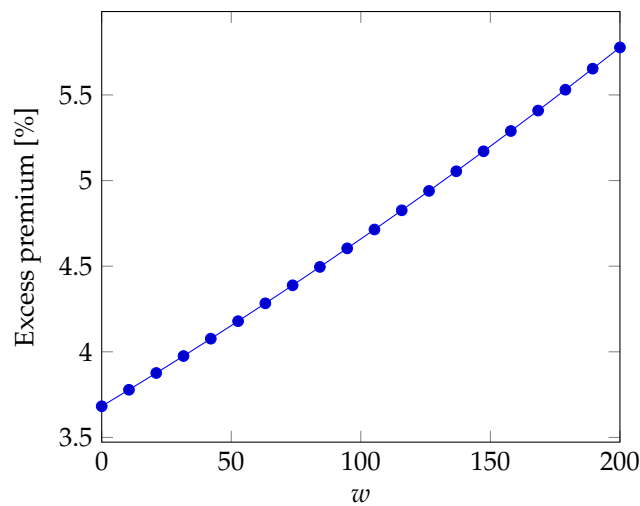


Figure 3. Relative excess premium for long outages (Exponential-Pareto model).

5.1.3. Unavailability

If a loss is experienced for each minute of no service, after recalling Equations (17) and (18), we have the premium

$$\begin{aligned}
 P_u &= k_u \frac{\beta}{1-\xi} \lambda T + \frac{\delta}{2} \frac{2k_u^2 \beta^2 \lambda T}{(1-\xi)(1-2\xi)} \\
 &= \frac{k_u \beta \lambda T}{1-\xi} \left(1 + \frac{\delta k_u \beta}{1-2\xi} \right).
 \end{aligned}
 \tag{54}$$

The excess premium is then

$$EP_u = \frac{P_u - \mathbb{E}[X_u]}{\mathbb{E}[X_u]} = \frac{\delta k_u \beta}{1-2\xi}.
 \tag{55}$$

For the values in Table 4, we have the excess premium $EP_u \simeq 0.226$, which represents a fairly large premium for risk.

5.2. The Pareto-Lognormal Model

Similarly to what we have done in Section 5.1, in this section, we compute the premium when outages are described by the Pareto-Lognormal model. Again, we proceed separately for the three QoS metrics. As for the Exponential-Pareto model, we report the parameter values used for numerical examples in Table 5.

Table 5. Sample parameter values for the Pareto-Lognormal model.

Pareto		Lognormal		Time
h	α	μ	σ	T [Days]
1834	4	4.6	1.3	365

5.2.1. Number of Outages

By recalling the general expression of the premium in Equation (48) and the first two moments of the economic loss in Equations (32) and (33), we get the following premium:

$$\begin{aligned}
 P_f &= k_f \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \\
 &+ \delta \frac{k_f^2}{2} \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} (2i+1) G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) - \mathbb{E}^2[N_f] \right] \\
 &= \left[k_f + \delta \frac{k_f^2}{2} \left[\left(\frac{h}{T}\right)^\alpha - \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \right] \\
 &\times \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \\
 &+ \delta \frac{k_f^2}{2} \sum_{i=1}^{\infty} 2i G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right).
 \end{aligned}
 \tag{56}$$

5.2.2. Number of Long Outages

Here, we combine the general expression of the premium in Equation (48) and the first two moments of the economic loss provided by Equations (36) and (37), which gives us the premium

$$\begin{aligned}
 P_{lf} &= k_{lf} \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right) \right] \\
 &+ \delta \frac{k_{lf}^2}{2} \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right) \right]^2 \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} (2i+1)G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) - \mathbb{E}[N_f]^2 \right] \\
 &+ \delta \frac{k_{lf}^2}{2} \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right) \right] \left[G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right) \right] \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \\
 &= \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right) \right] \left(k_{lf} + \delta \frac{k_{lf}^2}{2} \right) \\
 &- \delta \frac{k_{lf}^2}{2} \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right]^2 \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right) \right]^2 \\
 &+ \delta \frac{k_{lf}^2}{2} \left[1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right) \right]^2 \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right).
 \end{aligned} \tag{57}$$

The excess premium can be computed as in Equation (54). Its behaviour as the outage duration threshold grows is however shown in Figure 4.

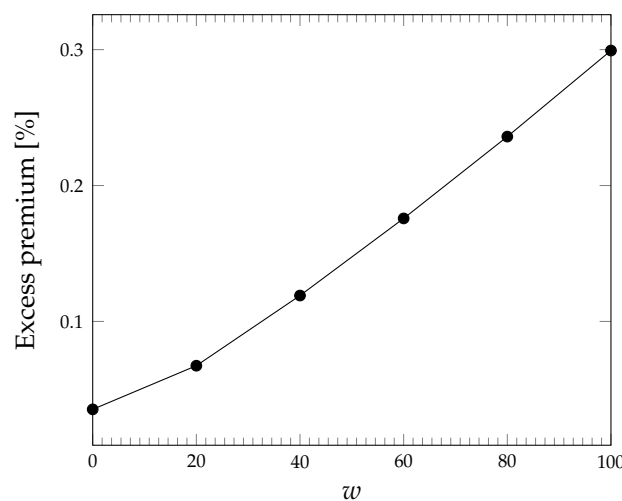


Figure 4. Relative excess premium for long outages (Pareto-Lognormal model).

5.2.3. Unavailability

Again, we employ here the general expression of the premium in Equation (48) and the first two moments of the economic loss provided by Equations (40) and (41) to get

$$\begin{aligned}
 P_u &= k_u \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] e^{\mu + \frac{\sigma^2}{2}} \\
 &+ \delta \frac{k_u^2}{2} e^{2\mu + \sigma^2} \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} (2i + 1) G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) - \mathbb{E}^2[N_f] \right] \\
 &+ \delta \frac{k_u^2}{2} \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \\
 &= k_u e^{\mu + \frac{\sigma^2}{2}} \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \\
 &+ \left[1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) \right] \\
 &\times \left[\delta \frac{k_u^2}{2} e^{\mu + \frac{\sigma^2}{2}} \left[\left(\frac{h}{T}\right)^\alpha - \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) + e^{\sigma^2} - 1 \right] \right] \\
 &+ \delta \frac{k_u^2}{2} e^{2\mu + \sigma^2} \sum_{i=1}^{\infty} 2i G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right).
 \end{aligned} \tag{58}$$

For the values in Table 5, we have an excess premium equal to $EP_u \simeq 0.117$, which is roughly half that obtained under the Exponential-Pareto model.

6. Refund Sustainability

In Section 5, we have seen that the insurance premium is a function of the unit compensation (i.e., k_f , k_{If} or k_u , depending on the QoS metric adopted). The relationship is typically of the quadratic kind. This means that the premium will go up with the unit compensation (and faster than that) that the cloud provider is bound to pay its customer. As a consequence, large unit compensations lead to high premiums, which may become unsustainable for the cloud provider: premiums represent a fixed (deterministic) outcome that erodes the profit margin of the cloud provider and, if large, may even turn a potential profit into a sure loss. In Table 3, we have seen how large the unit loss per minute can be; though those values may represent a basis for setting the unit compensation (in that case for k_u), this way of acting may lead to an exceedingly high liability for the cloud provider. We suggest here a possible way to limit the liability of the cloud provider so as to make it sustainable. This involves setting a limit for the unit compensation that the cloud provider may pay. In this section, we derive such a limit for the two failure models and the three QoS metrics that we have considered so far. After obtaining the general expression for the maximum unit refund, we follow therefore the same approach as in Sections 3 and 5, dealing separately with each subcase.

6.1. Unit Refund Limit

As hinted above, the compensation that cloud providers may pay their customers should be bounded, so that it does not pose a significant threat to the overall economic balance of the company. In many cases, the compensation is set as a percentage (well below 100%) of the service fee (the compensation is therefore a partial refund); this limit guarantees that the outflow related to refunds cannot exceed that same percentage of the revenues. In Table 6, the percentage values set by some providers are reported for the case where the QoS metric is the availability. All providers allow for a larger refund percentage as the QoS degrades more and more. However, for three out of four providers, the refund can never be full.

Table 6. Refunds as a percentage of the service fee.

Cloud Provider	Monthly Uptime [%]	Service Credit [%]
Amazon ¹	99–99.99	10
	< 99	30
Google ²	99–99.99	10
	95–99	25
	<95	50
Microsoft Azure ³	99–99.9	10
	< 99	25
Rackspace ⁴	99.5–99.9	10
	99–99.49	25
	98–98.99	40
	97.5–97.99	55
	97–97.49	70
	96.5–96.99	85
	<96.5	100

¹ <https://aws.amazon.com/it/compute/sla/>; ² <https://cloud.google.com/functions/sla/>; ³ <https://azure.microsoft.com/en-us/support/legal/sla/summary/>; ⁴ https://www.rackspace.com/sites/default/files/legal/cloud_slas_global_08_feb_2016.pdf.

In order to limit the liability of cloud providers within safe bounds, we consider here a similar approach by limiting the unit refund. We set therefore

$$k_* = \gamma F \quad \gamma \in (0, 1), \tag{59}$$

where the asterisk is a jolly character so that the expression may be applied to any of the three performance parameters, and γ defines the fraction of the fee that the cloud provider may accept to lose due to compensation for each QoS violation.

On the other hand, the cloud provider cannot be led to pay too high a premium, since the premium is a fixed periodic payment that erodes the provider’s profit. In order to guarantee that the revenues represented by the service fees are not overly reduced, we can set the premium so as not to be larger than a predefined fraction of the service fee itself:

$$P_* < \rho F \quad \rho \in (0, 1). \tag{60}$$

Since the economic loss X_* is proportional to the QoS metric (which is either N_f , N_{if} or U), as embodied by Equations (9), (11) and (16), we introduce a generic variable L to derive a general expression that may then be tailored for the case at hand by replacing L with N_f , N_{if} or U , respectively.

The general constraint on the premium can then be expressed as

$$\begin{aligned} P_* &= \mathbb{E}[X_*] + \delta \frac{\mathbb{V}[X_*]}{2} = k_* \mathbb{E}[L] + k_*^2 \delta \frac{\mathbb{V}[L]}{2} \\ &= \gamma F \mathbb{E}[L] + \gamma^2 F^2 \delta \frac{\mathbb{V}[L]}{2} < \rho F. \end{aligned} \tag{61}$$

This constraint leads to a quadratic inequality in the unit refund γ :

$$F \delta \frac{\mathbb{V}[L]}{2} \gamma^2 + \mathbb{E}[L] \gamma - \rho < 0. \tag{62}$$

Since the discriminant associated with the quadratic form is

$$\Delta = \left(\frac{\mathbb{E}[L]}{2} \right)^2 + \frac{\delta}{2} F \rho \mathbb{V}[L] \tag{63}$$

and $\gamma > 0$, the solution of the quadratic inequality is

$$\gamma < \frac{2\sqrt{\Delta} - \mathbb{E}[L]}{\delta F \mathbb{V}[L]} = \frac{\sqrt{1 + 2\delta F \rho \frac{\mathbb{V}[L]}{\mathbb{E}^2[L]} - 1}}{\delta F \frac{\mathbb{V}[L]}{\mathbb{E}[L]}}, \tag{64}$$

which gives us an upper bound for the unit refund γ to be sustainable.

The premium-setting procedure therefore goes through the following steps:

1. Set the tolerable fraction ρ of the service fee as in Equation (60);
2. Compute the upperbound γ on the refund (Sections 6.2 and 6.3);
3. Compute the limit unit refund as per Equation (59);
4. Compute the premium (Sections 5.1 and 5.2).

In the following subsections, we derive the expression for γ when either outage model applies and for the three QoS metrics. In order to show some numerical results, we consider the parameter values shown Tables 4 and 5 for the failure models, and in Table 7 for the refund sustainability computation. The risk aversion factor δ is computed according to Equation (49).

Table 7. Parameter values used in the refund sustainability examples.

Parameter	Value
Long outage threshold w	2 h
Probability premium η	0.25
Allowable fee percentage ρ	0.05 (5%)

6.2. The Exponential-Pareto Model

6.2.1. Number of Outages

In this case, the loss statistics is the number of outages N_f , so that $\mathbb{E}[L] = \lambda T$ and $\mathbb{V}[L] = \lambda T$ and

$$\frac{\mathbb{V}[L]}{\mathbb{E}[L]} = 1 \quad \frac{\mathbb{V}[L]}{\mathbb{E}^2[L]} = \frac{1}{\lambda T}. \tag{65}$$

From Equation (64), the constraint on the refund factor then becomes

$$\gamma_f < \frac{\sqrt{1 + \frac{2\delta F \rho}{\lambda T}} - 1}{\delta F}. \tag{66}$$

As can be seen in Figure 5a, the maximum value for γ_f diminishes as the fee grows with a roughly inverse square root dependence.

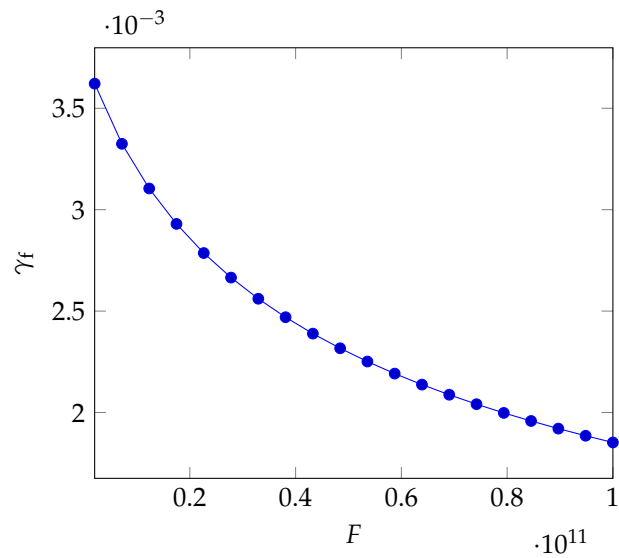
6.2.2. Number of Long Outages

In this case $L = N_{lf}$, and

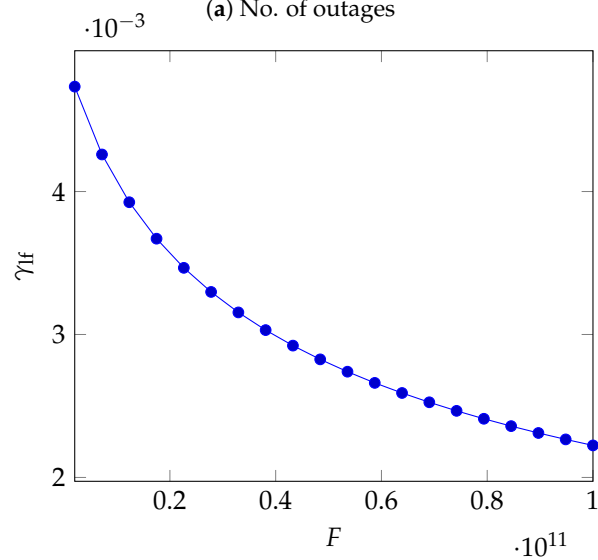
$$\frac{\mathbb{V}[L]}{\mathbb{E}[L]} = 1 \quad \frac{\mathbb{V}[L]}{\mathbb{E}^2[L]} = \frac{1}{\lambda T} \left(1 + \frac{\xi W}{\beta} \right)^{\frac{1}{\xi}} \tag{67}$$

so that Equation (64) becomes

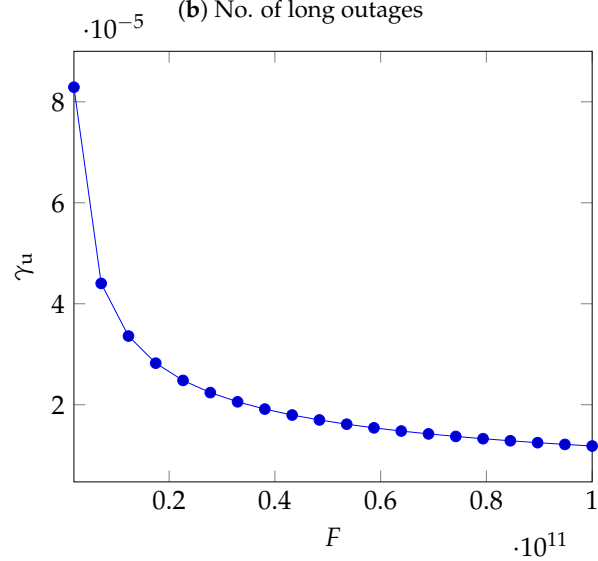
$$\gamma_{lf} < \frac{\sqrt{1 + \frac{2\delta F \rho}{\lambda T} \left(1 + \frac{\xi W}{\beta} \right)^{\frac{1}{\xi}} - 1}}{\delta F}. \tag{68}$$



(a) No. of outages



(b) No. of long outages



(c) Unavailability

Figure 5. Maximum allowable fraction of fee for unit refund under the Exponential-Pareto model.

In Figure 5b again, we see that, as the fee grows, the cloud provider has to lower the fraction of the fee that it can return to its customers for each long outage.

6.2.3. Unavailability

In this case $L = U$, and

$$\frac{\mathbb{V}[L]}{\mathbb{E}[L]} = \frac{2\beta}{1 - 2\zeta} \quad \frac{\mathbb{V}[L]}{\mathbb{E}^2[L]} = \frac{2}{\lambda T} \frac{1 - \zeta}{1 - 2\zeta}, \tag{69}$$

so that Equation (64) becomes

$$\gamma_u < \frac{1 - 2\zeta}{2\delta\beta F} \left[\sqrt{1 + \frac{4\delta F\rho}{\lambda T} \frac{1 - \zeta}{1 - 2\zeta}} - 1 \right]. \tag{70}$$

The upper bound on the unit refund is now shown in Figure 5c, where the decline as the fee grows is steeper in the lower domain of the fee values.

6.3. The Pareto-Lognormal Model

6.3.1. Number of Outages

From Section 4.4.1, we recall that the first two moments of the number of outages are respectively

$$\mathbb{E}[N_f] = 1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right), \tag{71}$$

$$\begin{aligned} \mathbb{V}[N_f] &= 1 - \left(\frac{h}{T}\right)^\alpha + \sum_{i=1}^{\infty} (2i + 1)G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right) - \mathbb{E}^2[L_f], \\ &= \mathbb{E}[L_f] - \mathbb{E}^2[L_f] + \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right). \end{aligned} \tag{72}$$

The two ratios involving those moments and employed in the computation of the maximum refund are respectively:

$$\frac{\mathbb{V}[N_f]}{\mathbb{E}[N_f]} = \frac{\mathbb{E}[N_f] - \mathbb{E}^2[N_f] + \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right)}{\mathbb{E}[N_f]}, \tag{73}$$

$$\frac{\mathbb{V}[N_f]}{\mathbb{E}^2[N_f]} = \frac{\mathbb{E}[N_f] - \mathbb{E}^2[N_f] + \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right)}{\mathbb{E}^2[N_f]}. \tag{74}$$

By replacing in the general expression (64) the expressions just obtained and using the abbreviation $G_i := G\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right)$, and adopting the two shortened forms $G_s = \sum_{i=1}^{\infty} G_i$ and $G_w = \sum_{i=1}^{\infty} 2iG_i$, we get the final expression of the maximum unit refund when the number of outages is used as the QoS metric:

$$\gamma_f = \frac{\sqrt{1 + 2\delta F\rho \left[\frac{1}{1 - \left(\frac{h}{T}\right)^\alpha + G_s} + \frac{G_w}{\left(1 - \left(\frac{h}{T}\right)^\alpha + G_s\right)^2} - 1 \right]} - 1}{\delta F \left[\frac{1}{1 - \left(\frac{h}{T}\right)^\alpha + G_s} + \frac{G_w}{\left(1 - \left(\frac{h}{T}\right)^\alpha + G_s\right)^2} - 1 \right]}. \tag{75}$$

This equation shows again that the maximum unit refund has to lower according to an inverse square law as the fee grows. In Figure 6a, we show the value of γ_f for the parameters listed in Table 7. The dependence here appears to be roughly linear because of the windows chosen for the fee values.

6.3.2. Number of Long Outages

Similarly to what we have done for the number of outages, we compute the ratios of the two moments for the case of long outages, by exploiting the results of Section 4.4.2:

$$\frac{\mathbb{V}[N_{lf}]}{\mathbb{E}[N_{lf}]} = 1 - \frac{\left(1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right)\right)}{\mathbb{E}[N_f]} \left(\mathbb{E}^2[N_f] - \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right)\right), \tag{76}$$

$$\frac{\mathbb{V}[N_{lf}]}{\mathbb{E}^2[N_{lf}]} = \frac{1}{\mathbb{E}[N_f]} - \frac{\left(1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right)\right)}{\mathbb{E}^2[N_f]} \left(\mathbb{E}^2[N_f] - \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right)\right). \tag{77}$$

We can now replace those expressions into the unit refund limit general expression, adopting the abbreviation $B(w) := 1 - G\left(\frac{\ln(w) - \mu}{\sqrt{2\sigma}}\right)$. The maximum unit refund for the case of long outages is then

$$\gamma_{lf} = \frac{\sqrt{1 + 2\delta F\rho \left[\frac{1}{1 - \left(\frac{h}{T}\right)^\alpha + G_s} + \frac{B(w)G_w}{\left(1 - \left(\frac{h}{T}\right)^\alpha + G_s\right)^2} - B(w)\right]} - 1}{\delta F \left[\frac{1}{1 - \left(\frac{h}{T}\right)^\alpha + G_s} + \frac{B(w)G_w}{\left(1 - \left(\frac{h}{T}\right)^\alpha + G_s\right)^2} - B(w)\right]}. \tag{78}$$

The same inverse square relationship appears to hold for the number of long outages as well, which is shown in Figure 6b.

6.3.3. Unavailability

Again, for the case of unavailability, we compute the ratio of the moments of the Unavailability, employing the expressions shown in Equations (40) and (41):

$$\frac{\mathbb{V}[L_U]}{\mathbb{E}[L_U]} = \frac{\left[\mathbb{E}[L_f]e^{\sigma^2} - \mathbb{E}^2[L_f] + \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right)\right] e^{\mu + \frac{\sigma^2}{2}}}{\mathbb{E}[L_f]}, \tag{79}$$

$$\frac{\mathbb{V}[L_U]}{\mathbb{E}^2[L_U]} = \frac{\mathbb{E}[L_f]e^{\sigma^2} - \mathbb{E}^2[L_f] + \sum_{i=1}^{\infty} 2iG\left(\frac{T - \mathbb{E}[Z_i]}{\sqrt{\mathbb{V}[Z_i]}}\right)}{\mathbb{E}^2[L_f]}. \tag{80}$$

We then replace these ratios in the general expression of the refund limit given by Equation (64) to get the maximum unit refund for the case of unavailability:

$$\gamma_u = \frac{\sqrt{1 + 2\delta F\rho \left[\frac{e^{\sigma^2}}{1 - \left(\frac{h}{T}\right)^\alpha + G_s} + \frac{G_w}{\left(1 - \left(\frac{h}{T}\right)^\alpha + G_s\right)^2} - 1\right]} - 1}{\delta F \left[\frac{e^{\sigma^2}}{1 - \left(\frac{h}{T}\right)^\alpha + G_s} + \frac{G_w}{\left(1 - \left(\frac{h}{T}\right)^\alpha + G_s\right)^2} - 1\right] e^{\mu + \frac{\sigma^2}{2}}}. \tag{81}$$

Finally, in Figure 6c, we show the same diminishing trend for γ_u .

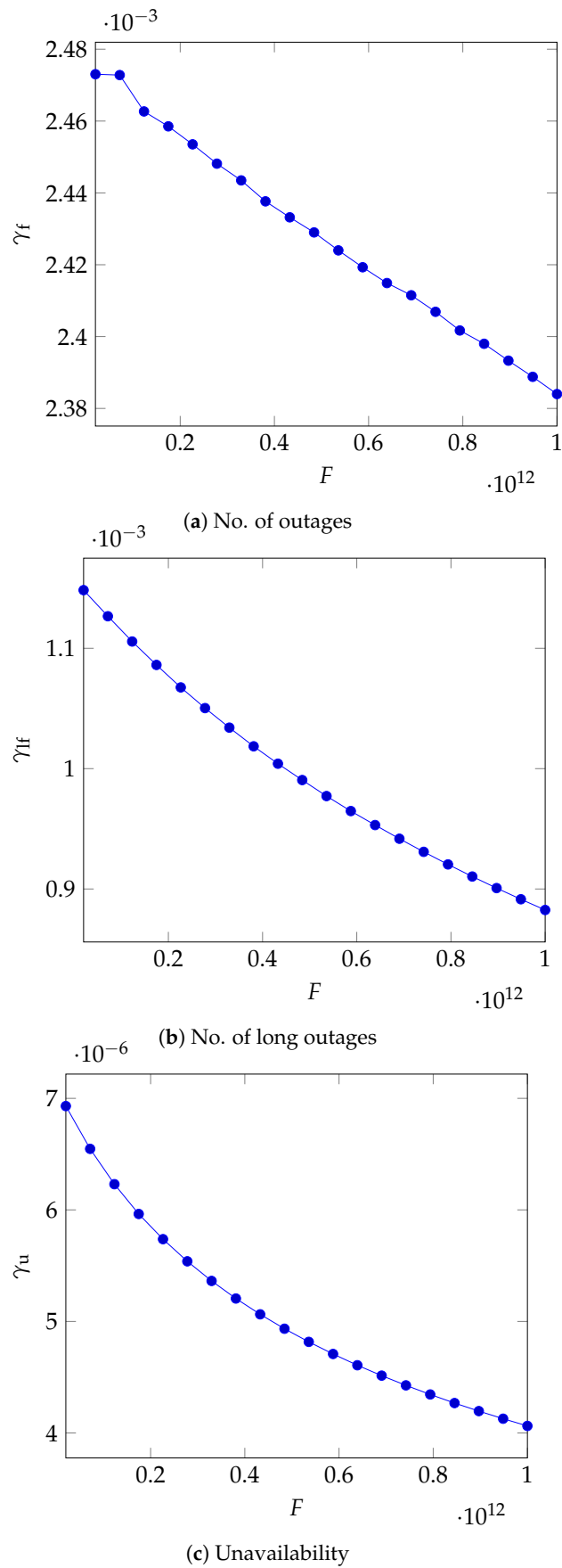


Figure 6. Maximum allowable fraction of fee for unit refund under the Pareto-Lognormal model.

7. Conclusions

We have provided formulas for the insurance premium that cloud providers have to pay to protect themselves against the danger of excessive claims by their customers in the case of cloud outages. The formulas can be applied quite straightforwardly as a function of the parameters describing the cloud outage phenomenon. The formulas depend however on how much the cloud provider is willing to pay for each disruption unit (i.e, for each outage, or for each outage exceeding a prescribed duration, or for each minute of unavailability, depending on the QoS metric chosen). In order to keep the insurance premium within acceptable limit and avoid endangering the economic balance of the cloud provider, we have provided formulas for the maximum unit refund as well. This upper bound lowers as the service fee grows, roughly according to an inverse square law.

Our set of formulas allows cloud providers to consider insurance as an additional, non network-centric, means of protection against the consequences of poor cloud performance. A possible extension would lead us to consider a non-binary state space, where several levels of QoS degradation are possible as an approximation to the case of graceful degradation of service quality.

Author Contributions: Conceptualization, L.M., A.M., and M.N.; methodology, L.M., A.M., and M.N.; formal analysis, L.M., A.M., and M.N.; investigation, L.M., A.M., and M.N.; software, A.M.; writing—original draft preparation, L.M., A.M., and M.N.; writing—review and editing, L.M., and M.N.; supervision, L.M. and M.N.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following symbols and abbreviations are used in this manuscript:

$\gamma_f, \gamma_{lf}, \gamma_u$	unit refund as a fraction of the fee
D	Duration of the OFF period
δ	Risk aversion coefficient
F	Service fee
k_f, k_{lf}, k_u	Loss for each outage, long outage, and per unit time
L	Generic variable for QoS metric
λ	Inverse mean value of the exponential distribution
MTTF	Mean Time To Failure
MTTR	Mean Time To Repair
μ, σ	Parameters of the lognormal distribution
N_f, N_{lf}	Number of outages, Number of long outages
r	Risk aversion function
ρ	Allowable fee percentage
S	Duration of the ON period
T	Duration of contract
U	Cumulative unavailability time
w	Threshold for long outages
X_f, X_{lf}, X_u	Economic loss for number of outages, number of long outages and unavailability time
ξ, β	Shape and scale parameter for GPD
ω	Value of insured's assets

References

1. Alhamad, M.; Dillon, T.; Chang, E. Conceptual SLA framework for cloud computing. In Proceedings of the 2010 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST), Dubai, UAE, 13–16 April 2010; pp. 606–610.
2. Baset, S.A. Cloud SLAs: Present and future. *ACM SIGOPS Oper. Syst. Rev.* **2012**, *46*, 57–66. [[CrossRef](#)]

3. Hussain, W.; Hussain, F.K.; Hussain, O.K.; Damiani, E.; Chang, E. Formulating and managing viable SLAs in cloud computing from a small to medium service provider's viewpoint: A state-of-the-art review. *Inf. Syst.* **2017**, *71*, 240–259. [[CrossRef](#)]
4. Mubeen, S.; Asadollah, S.A.; Papadopoulos, A.V.; Ashjaei, M.; Pei-Breivold, H.; Behnam, M. Management of service level agreements for cloud services in IoT: A systematic mapping study. *IEEE Access* **2018**, *6*, 30184–30207. [[CrossRef](#)]
5. Qiu, M.M.; Zhou, Y.; Wang, C. Systematic analysis of public cloud service level agreements and related business values. In Proceedings of the 2013 IEEE International Conference on Services Computing (SCC), Santa Clara, CA, USA, 28 June–3 July 2013; pp. 729–736.
6. Serrano, D.; Bouchenak, S.; Kouki, Y.; de Oliveira, F.A., Jr.; Ledoux, T.; Lejeune, J.; Sopena, J.; Arantes, L.; Sens, P. SLA guarantees for cloud services. *Future Gener. Comput. Syst.* **2016**, *54*, 233–246. [[CrossRef](#)]
7. Alboghdady, S.; Winter, S.; Taha, A.; Zhang, H.; Suri, N. C'mon: Monitoring the Compliance of Cloud Services to Contracted Properties. In Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy, 29 August–1 September 2017; p. 36.
8. Stephen, A.; Benedict, S.; Kumar, R.A. Monitoring IaaS using various cloud monitors. *Cluster Comput.* **2018**, 1–13. [[CrossRef](#)]
9. Syed, H.J.; Gani, A.; Ahmad, R.W.; Khan, M.K.; Ahmed, A.I.A. Cloud Monitoring: A Review, Taxonomy, and Open Research Issues. *J. Netw. Comput. Appl.* **2017**, *98*, 11–26. [[CrossRef](#)]
10. Nawaz, F.; Hussain, O.K.; Janjua, N.; Chang, E. A proactive event-driven approach for dynamic QoS compliance in cloud of things. In Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, 23–26 August 2017; ACM: New York, NY, USA, 2017; pp. 971–975.
11. Yuan, X.; Li, Y.; Jia, T.; Liu, T.; Wu, Z. An analysis on availability commitment and penalty in cloud sla. In Proceedings of the 2015 IEEE 39th Annual Computer Software and Applications Conference (COMPSAC), Taichung, Taiwan, 1–5 July 2015; Volume 2, pp. 914–919.
12. Yuan, X.; Tang, H.; Li, Y.; Jia, T.; Liu, T.; Wu, Z. A competitive penalty model for availability based cloud SLA. In Proceedings of the 2015 IEEE 8th International Conference on Cloud Computing (CLOUD), New York, NY, USA, 27 June–2 July 2015; pp. 964–970.
13. Mastroeni, L.; Naldi, M. Insurance Pricing and Refund Sustainability for Cloud Outages. In *Economics of Grids, Clouds, Systems, and Services*; Pham, C., Altmann, J., Bañares, J.Á., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 3–17.
14. Dunne, J.; Malone, D. Obscured by the cloud: A resource allocation framework to model cloud outage events. *J. Syst. Softw.* **2017**, *131*, 218–229. [[CrossRef](#)]
15. Naldi, M. The availability of cloud-based services: Is it living up to its promise? In Proceedings of the 2013 9th International Conference on the Design of Reliable Communication Networks (DRCN), Budapest, Hungary, 4–7 March 2013; pp. 282–289.
16. Naldi, M. ICMP-based third-party estimation of cloud availability. *Int. J. Adv. Telecommun. Electrotech. Signals Syst.* **2017**, *6*, 11–18. [[CrossRef](#)]
17. Mastroeni, L.; Naldi, M. Network protection through insurance: Premium computation for the ON-OFF service model. In Proceedings of the 8th International Workshop on the Design of Reliable Communication Networks DRCN, Krakow, Poland, 10–12 October 2011; pp. 46–53.
18. Mastroeni, L.; Naldi, M. Compensation Policies and Risk in Service Level Agreements: A Value-at-Risk Approach under the ON-OFF Service Model. In *Lecture Notes in Computer Science, Proceedings of the Economics of Converged, Internet-Based Networks—7th International Workshop on Internet Charging and QoS Technologies, ICQT 2011, Paris, France, 24 October 2011*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6995, pp. 2–13.
19. Hogben, G.; Pannetrat, A. Mutant apples: A critical examination of cloud SLA availability definitions. In Proceedings of the 2013 IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom), Bristol, UK, 2–5 December 2013; Volume 1, pp. 379–386.
20. Poularikas, A.D. Probability and Stochastic Processes. In *The Handbook of Formulas and Tables for Signal Processing*; Poularikas, A.D., Ed.; CRC Press: Boca Raton, FL, USA, 1999.
21. Beichelt, F. *Stochastic Processes in Science, Engineering and Finance*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.
22. Gaurav, S.; Machiraju, S. *Hardening Azure Applications*; Apress: New York, NY, USA, 2015.

23. Kaas, R.; Goovaerts, M.; Dhaene, J.; Denuit, M. *Modern Actuarial Risk Theory*; Springer: Berlin/Heidelberg, Germany, 2004.
24. Pratt, J.W. Risk Aversion in the Small and in the Large. *Econometrica* **1964**, *32*, 122–136. [[CrossRef](#)]
25. Morshedlou, H.; Meybodi, M.R. Decreasing impact of SLA violations: A proactive resource allocation approach for cloud computing environments. *IEEE Trans. Cloud Comput.* **2014**, *2*, 156–167. [[CrossRef](#)]
26. Böhme, R.; Kataria, G. Models and Measures for Correlation in Cyber-Insurance. In Proceedings of the DIMACS Workshop on Information Security Economics, Piscataway, NJ, USA, 18–19 January 2007.
27. Babcock, B.A.; Choi, E.K.; Feinerman, E. Risk and probability premiums for CARA utility functions. *J. Agric. Resour. Econ.* **1993**, *18*, 17–24.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).