2018 UKSim-AMSS 20th International Conference on Modelling & Simulation

Bagging Estimation of Availability in Public Cloud Storage

Maurizio Naldi
Dpt. of Civil Engineering and Computer Science
University of Rome Tor Vergata
Rome, Italy
Email: maurizio.naldi@uniroma2.it

***Abstract*—The availability of cloud storage is a key performance parameter. Though its measurement is included in any cloud monitoring system, the statistical accuracy of such measurements is often neglected. Such measurements may therefore be useless in SLA-related disputed between cloud providers and customers. In this paper Bootstrap Averaging (Bagging) is employed to assess the statistical accuracy when we measure the availability and the number of outages over a single period of time. The distribution and the confidence intervals are reported for a set of major cloud providers, using datasets built from customer reports. By using proper confidence intervals, the compliance of cloud providers with respect to target availability and outage figures can be checked.**

Keywords*—Cloud storage, Availability, Service Level Agreement, Bagging.

I. Introduction

Cloud storage services represent one of the most successful stories since the advent of cloud computing [1], [2], [3], [4]. Platform such as Dropbox [5] or Google Drive now provide storage services used daily by both corporate and private customers around the world.

Their usage, favoured by very low prices, or even free offers [6], spurs however several concerns about their security [7] and availability. In particular, availability has been unanimously listed among the performance objectives that any cloud storage services should be tested for [8], [9]. Unfortunately, deciding that availability is a key performance parameter and that it must be measured is not the end of the story, since it may not be easy to measure the availability accurately enough to exploit the results in a possible dispute between the cloud service provider and the customer. For example, the network connecting the customer with the cloud may affect the measurement, so that the measured availability may reflect the network availability rather than that of the cloud [10], [11], [12].

A further problem is represented by the snapshot provided by the measurement system: since the availability is measured on the basis of the duration of working and repair periods, which are intrinsically stochastic quantities, what we measure is actually an instance of a stochastic process, and we estimate availability rather than measuring it. Relying on a single snapshot, i.e. a single observation of the cloud behaviour over a single period of time, however, does not allow us to obtain an indication of the accuracy of our estimate.

In this paper, we deal with the problem of providing an indication of the accuracy of the estimate of the availability of a cloud storage platform when we have observations over a single period. We adopt a Bootstrap Averaging (Bagging) technique and demonstrate its results by applying it to a dataset containing outage data of five major cloud providers.

After briefly reviewing Service Level Agreements (SLA) for cloud storage in Section II, and describing the bagging technique in Section III and the dataset in Section IV, we show the results of the technique for two major performance parameters: the availability in Section V and the number of outages in Section VI.

II. Cloud Storage Service Level Agreements

Similarly to any service provisioning, cloud services are typically sold with some guarantees on the service itself. In this section, we go through a brief review of the description and management of such service guarantees.

The quality of service to be provided when using a cloud can be defined through a Service Level Agreement between the parties, i.e., the cloud provider and the customer. Such SLAs include both functional and performance requirements, which the provider must comply with. Mechanisms for the cloud provider to be able to meet those requirements by a suitable resource provisioning process are described, e.g., in [13]. Any SLA management system must anyway include a cloud monitoring system, so as to measure the quantities prescribed by the SLA and react accordingly [14], [15], [16], [17], [18], [19].

The goal of such monitoring systems is dual: they may be used to act on the cloud itself to bring it back to an acceptable operating condition, but they may also be used to detect and report SLA violations. In the latter case, a penalty or compensation mechanism is incorporated in the SLA to account for the service disruption suffered by customers [20].

Availability is therefore included among Service Level Objectives (SLO), with ambitious targets: in Table I (excerpted from [13]) we see that the majority of cloud providers claim to provide 100% availability, a hard-to-believe goal even if we do not include shutdown periods due to preventive maintenance in the total downtime.

We introduce now the established notion of availability in the case of cloud storage. We consider an ON-OFF

0-7695-6405-4/18/$31.00 ©2018 IEEE
DOI 10.1109/UKSim.2018.00040

Cloud provider	Availability SLO [%]
Amazon Web Service	99.95
AT&T Synaptic	99.9
CloudSigma	100
ElasticHosts	100
FlexiScale	100
GoGrid	100
JoyentCloud	100
layeredtech	100
Locaweb	99.9
Opsource	100
Rackspace	100
ReliaCloud	100
RSAWEB Cloud servers	ND
SliceHost	ND
Storm On demand	100
Terremark vCloud express	100
VPSNET	100

TABLE I
SERVICE LEVEL AGREEMENT COMMITMENTS FOR AVAILABILITY

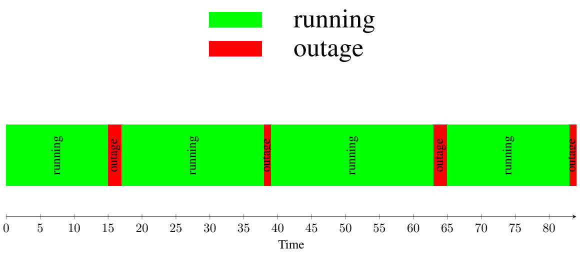

Fig. 1. Cloud state sequence

service model, i.e., the data stored on the cloud are either fully available or not [21]. The service timeline is therefore a sequence of ON and OFF states as shown in Fig. 1, where the duration of the OFF state is the time needed to repair the cloud. In the alternation of ON and OFF states, the durations of both can be considered as the instances of two random variables, respectively S and D, so that the availability is

$$A = \frac{\mathbb{E}[S]}{\mathbb{E}[S] + \mathbb{E}[D]} \tag{1}$$

For the time being, we assume that we are able to precisely measure the duration of working and outage periods of the cloud, i.e., we are able to detect and leave out the cases where the data are unreachable because of the network rather than the cloud [10], [11], [12].

III. BAGGING ESTIMATION

In this section, we describe the estimation procedure based on Bagging, i.e., Bootstrap Averaging. A good introduction to Bagging can be found in [22]. We apply the procedure to two performance parameters: availability and number of outages.

We assume that we have observed the cloud over a period of time. As the result of that observation, we have two sets of data, including respectively the observed times between subsequent outages and the outage durations. We indicate the first dataset as $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ and the second as $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$. The cloud undergoes therefore a sequence of ON (running) and OFF (outage) states, as depicted in Fig. 1, where green and red strips have lengths drawn from $\mathcal{S}$ and $\mathcal{D}$ respectively.

The observation of such a sequence of alternating states provides us with the following estimate of the availability as defined in Equation (1):

$$a = \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} s_i + \sum_{i=1}^{n} d_i}, \tag{2}$$

which can be considered as an instance of the random variable

$$A_0 = \frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} S_i + \sum_{i=1}^{n} D_i}, \tag{3}$$

where the S_i's and D_i's are i.i.d. random variables distributed as S and D respectively.

However, since we have a single sample of observations, we cannot derive any statistical properties of that estimator. We resort therefore to the bootstrap estimator

By sampling the original datasets with replacement, we generate $2B$ bootstrap datasets, respectively made of n outage interarrival times (drawn from the dataset $\mathcal{S}$) and n outage durations (drawn from the dataset $\mathcal{D}$). We indicate the i-th couple of such dataset as $\{\hat{\mathcal{S}}_i, \hat{\mathcal{D}}_i\}$, where $\hat{\mathcal{S}}_i = \{\hat{s}_{i1}, \hat{s}_{i2}, \ldots, \hat{s}_{1n}\}$ and $\hat{\mathcal{D}}_i = \{\hat{d}_{i1}, \hat{d}_{i2}, \ldots, \hat{d}_{1n}\}$, with the hatted quantities representing the output of the sampling with replacement procedure and $i = 1, 2, \ldots, B$.

For each bootstrap replica we can now build a chain of intervals of availability and unavailability, resembling that shown in Fig. 1. For each couple of bootstrap datasets we can therefore estimate the availability as

$$A_i = \frac{\sum_{j=1}^{n} \hat{s}_{ij}}{\sum_{j=1}^{n} \hat{s}_{ij} + \sum_{j=1}^{n} \hat{d}_{ij}} \qquad i = 1, \ldots, B. \tag{4}$$

The bagging estimate of the availability is finally the arithmetic average

$$\hat{A} = \frac{\sum_{j=1}^{B} A_i}{B}. \tag{5}$$

We can now obtain a statistical evaluation of the estimator of Equation (3), by considering the properties of $\hat{A} - a$ in place of $A_0 - A$ (see Section 2.3 of [23]). We can also build confidence interval for this estimate by using the quantiles of the distribution of $\hat{A}$. In particular the α quantile of $\hat{A} - a$ shall be estimated as $A_{((B+1)\alpha)} - a$, where $A_{(j)}$ represents the j-th order statistics in the set $\{A_1, A_2, \ldots, A_B\}$.

An additional performance metric is the number of outages over a given period of time T (e.g., a month or a year), as considered in [21], which can be defined as

$$\text{NO} = M : M \cdot \mathbb{E}[S+D] < T \wedge (M+1)\mathbb{E}[S+D] > T. \tag{6}$$

Similarly to what we have done for the availability, we can build two sets of bootstrap replications and derive B estimates of the number of outages

$$\text{NO}_i = m_i : \sum_{j=1}^{m_i} \left(\hat{s}_{ij} + \hat{d}_{ij}\right) < T \wedge \sum_{j=1}^{m_i+1} \left(\hat{s}_{ij} + \hat{d}_{ij}\right) > T \qquad i = 1, 2, \ldots B \tag{7}$$

and finally get the bagging estimate as

$$\hat{\text{NO}} = \frac{\sum_{i=1}^{B} \text{NO}_i}{B}. \tag{8}$$

IV. Outage Datasets

In order to demonstrate the effectiveness of bagging estimation, we wish to apply it to a real-world dataset. As remarked in the Introduction, measurements on clouds are quite rare in the literature. Though we start to see some [24], data are typically reported in an aggregate way, and the original dataset is not available. We have therefore resorted to a dataset that we have already employed in [25]. That dataset is composed of customer-reported data rather than instrumental measurements. In this section we describe its characteristics.

We consider data provided by two major providers of data about cloud outages:

- Cloutage;
- IWGCR.

Cloutage was founded by the Open Security Foundation in April 2010 and documents known and reported incidents with cloud services, reporting type (hacking, outage, or vulnerability), date, duration, and a reference to the web page containing more info on each incident. The website (cloutage.org) now appears to have been discontinued, so that we rely on the data collected in our previous measurement campaign [25].

The International Working Group on Cloud Computing Resiliency (IWGCR, hosted on http://iwgcr.org/) is a working group with the mission to monitor and analyze cloud computing resiliency, composed of IT executives, academic researchers, and industry representatives. It started its activities in March 2012, though reporting data as back in time as October 2011.

Though the two data sources report data on several providers, we have focussed on the following five providers:

- Google;
- Amazon;
- Rackspace;
- Salesforce;
- Windows Azure.

Company	No. of outages	No. of durations
Google	59	37
Amazon	21	16
Rackspace	76	62
Salesforce	10	10
Windows Azure	10	10

TABLE II
SIZE OF DATASETS

In Table II, we report the number of observations collected during the campaign.

For each event, the date of occurrence and its duration have been recorded. The latter quantity is sometimes unreported. We are aware of the lack of absolute guarantees about the dataset coverage and accuracy: the ending time is reported by the provider as that corresponding to the service having been restored for a large fraction of the people affected, so that for some people the duration may be longer than the reported one. The reported times are therefore to be taken as underestimates, erring on the low side. But we must also consider that there are no widely available third-party measurements of cloud performance encompassing a significant number of providers [10], [26]. In addition, even if the results of extensive third-party measurements were available, their accuracy should be subject to scrutiny, since it has been shown that the influence of the network on the measuring probes may lead to inaccurate estimates [11], [12]

V. Availability Estimation in a Customer-Reported Context

As recalled in Section IV, there is not a public repository of data concerning cloud outages. Though for any customer it is always possible to monitor its own cloud provider, a third-party reliable and accessible monitoring platform would be strongly needed. However, the platforms described in Section IV may provide data suitable to perform an availability estimation. In this section, we see how the availability estimation procedure described in Section III may be applied by using the data provided by the sources indicated in Section IV, where all the information are supplied by customers themselves.

The aim is to achieve a statistically robust estimation of the availability, rather than the single instance measurement employed in most monitoring systems. The method described in Section III provides us with an indication of the confidence to be attached to the observed availability value, i.e., that computed by the ratio of the observed uptime to the total length of the monitoring window. We consider a number of tools and parameters that may help implement a statistically robust violation detection procedure, namely, the boxplot, the 95% confidence interval, and the kernel density estimator.

We can first describe how the observed availability is distributed around its actual value, i.e. that defined in

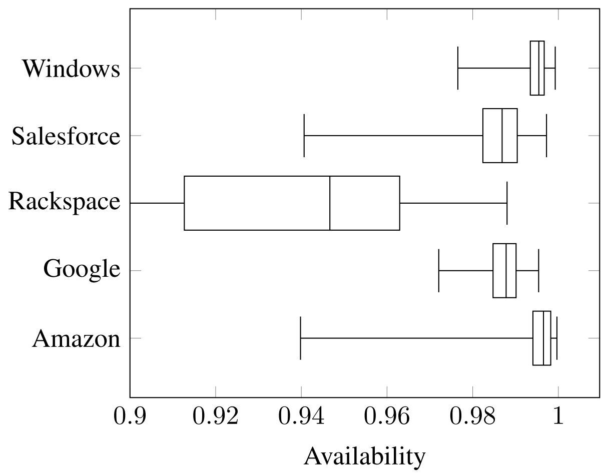

Fig. 2. Boxplot of availability

Equation (1) where the expected values of the uptime and downtime are considered. In Fig. 2 we report the boxplot showing the main statistics for the availability, as estimated through bagging. The boxplot's whiskers mark the minimum and maximum value among the B values obtained through bootstrap, while the box boundaries mark the 1st and 3rd quartile, and the segment inside the box represents the median availability estimate. Such boxplots may be used to see how disperse the observed availability values may be. If the boxplot extends way below the availability target indicated in the SLA, we may observe a violation when the system is actually complying with the SLA specifications, triggering, e.g., a compensation claim. We also see that the uncertainty surrounding the actual availability may be quite different among operators. In the case built with our measurements, we see that the estimates for Windows and Google appear quite tight, while that for Rackspace has a far larger dispersion. We also see that for three out of five operators the main body of the distribution (i.e., from the left tail to the 75% percentile) lies below the two-nines value (0.99), which can be considered as a minimum target.

A measure of the range of the availability estimates may be obtained by computing the two-sided 95% confidence interval for A. If the availability target lies outside the range provided by the confidence interval, we can be pretty sure that the cloud provider has missed the performance target and may therefore be considered liable to compensation claims. In Table III we report what is obtained with the dataset described in Section IV. On the upper side, just Amazon's confidence interval encloses the three-nines value, while three additional providers cross the two-nines target. If the availability target were 0.99, we could strike out one operator as non-compliant. On the lower side, all the providers have confidence limits that go below the two-nines value.

In order to have a complete view of the distribution of the availability, we estimate its probability density function

Provider	Lower limit	Upper limit
Windows	0.9855082	0.9982058
Salesforce	0.9674701	0.9945194
Rackspace	0.7981062	0.9799475
Google	0.9784518	0.9934143
Amazon	0.9820859	0.9993779

TABLE III
95% CONFIDENCE INTERVALS FOR THE AVAILABILITY

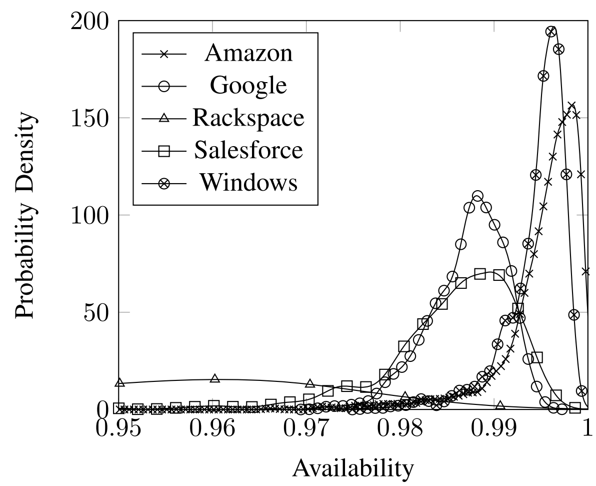

Fig. 3. Kernel Density Estimate of Availability

(pdf) through the kernel method, which places a kernel function around each data point and estimates the pdf as the sum of all the kernel components [27], [28]. In our case we use the standard Gaussian probability density function as a kernel, so that the estimate is

$$\hat{p}(x) = \frac{1}{B} \sum_{i=1}^{B} g(x - A_i), \tag{9}$$

where

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}. \tag{10}$$

In Fig. 3 we show the kernel density estimates for the five providers. All exhibit a significant skewness to the left. Only three providers, however, exhibit a peak beyond the two-nine availability value. The pdf for Rackspace encompasses a much wider range.

VI. NUMBER OF OUTAGES

After estimating the availability, we can conduct a similar analysis for the number of outages, which is another performance measure that may be employed in a SLA [21], [29]. In this section, we provide estimates for that metric, again for the dataset described in Section IV.

The boxplot distribution for the bagging estimator of the number of outages is shown in Fig. 4 and Fig. 5. Contrarily to the case of availability, where all providers exhibit values

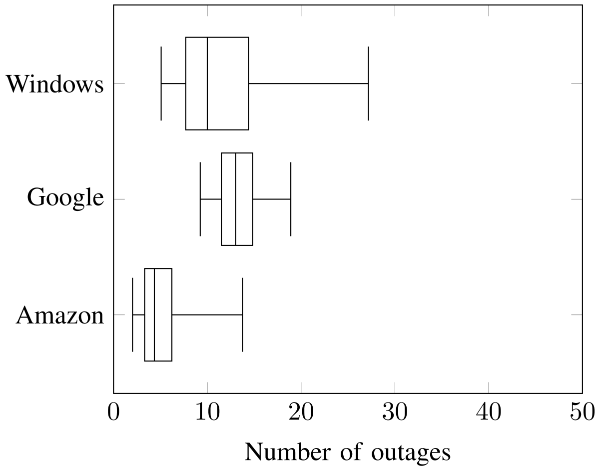

Fig. 4. Boxplot of the number of outages

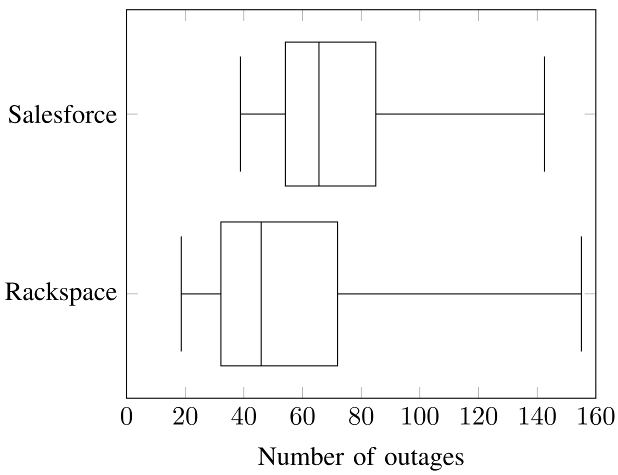

Fig. 5. Boxplot of the number of outages

roughly in the same range, significant differences emerge for the case of the number of outages: some providers are marred by many outages, though of short duration, while others are instead hit by fewer, but longer, outages. This forces us to show the results on two plots, since one scale could not accommodate all the results easily. We observe very tight confidence intervals for Google and Amazon. In particular for Amazon, the whole distribution entails less than an outage per month, while for Rackspace and Salesforce we have roughly between 4 and 6 outages per month.

The 95% confidence intervals are reported in Table IV. We observe a very large confidence interval for Rackspace and Salesforce.

Finally, the kernel density estimates are shown in Fig. 6 and Fig. 7. All curves are quite skewed to the right and exhibit a rather long tail.

Provider	Lower limit	Upper limit
Windows	5.075332	27.175377
Salesforce	38.81662	142.48247
Rackspace	18.64925	155.06943
Google	9.242874	18.900919
Amazon	2.024132	13.747629

TABLE IV
95% CONFIDENCE INTERVALS FOR THE NUMBER OF OUTAGES

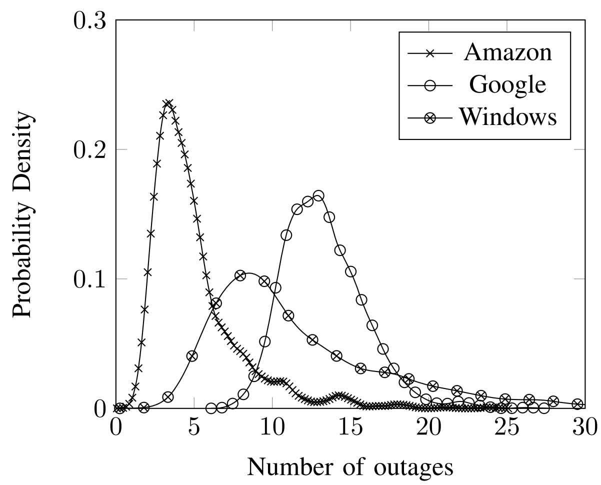

Fig. 6. Kernel Density Estimate of the Number of Outages

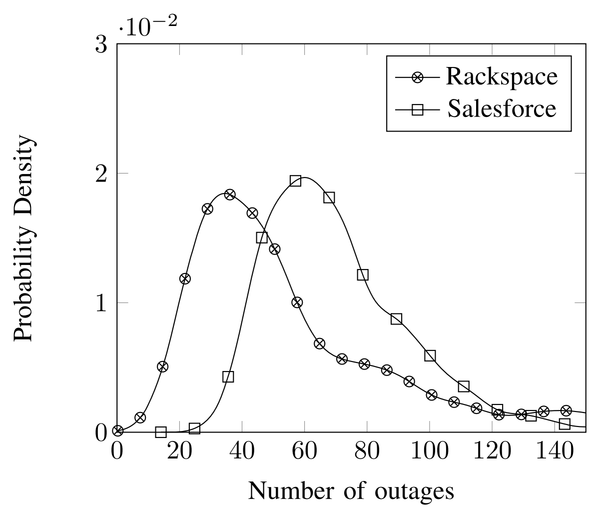

Fig. 7. Kernel Density Estimate of the Number of Outages

VII. Conclusion

Estimates of the availability of cloud storage services to be measured by monitoring systems are typically provided without indications of their statistical accuracy. This shortcoming makes them nearly useless to assess whether the cloud provider meets SLA requirements. A method to estimate the confidence intervals for the estimate of the number of outages and the availability is proposed, based on the Bootstrap Averaging (Bagging) technique.

The application of the technique to an experimental dataset shows that we are now able to compare the actual performances of cloud providers with the target ones on a statistically sound basis, and output a compliance/no compliance statement with the desired confidence level.

References

[1] W. Zeng, Y. Zhao, K. Ou, and W. Song, "Research on cloud storage architecture and key technologies," in *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*. ACM, 2009, pp. 1044–1048.

[2] J. Wu, L. Ping, X. Ge, Y. Wang, and J. Fu, "Cloud storage as the infrastructure of cloud computing," in *Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on*. IEEE, 2010, pp. 380–383.

[3] M. Naldi and L. Mastroeni, "Economic decision criteria for the migration to cloud storage," *European Journal of Information Systems*, vol. 25, no. 1, pp. 16–28, Jan 2016.

[4] X. Cheng and A. Bounfour, "The determinants of cloud computing adoption by large european firms," *INFORMATION SYSTEMS IN A CHANGING ECONOMY AND SOCIETY*, p. 50, 2015.

[5] I. Drago, M. Mellia, M. M Munafo, A. Sperotto, R. Sadre, and A. Pras, "Inside dropbox: understanding personal cloud storage services," in *Proceedings of the 2012 Internet Measurement Conference*. ACM, 2012, pp. 481–494.

[6] M. Naldi and L. Mastroeni, "Cloud storage pricing: A comparison of current practices," in *Proceedings of the 2013 International Workshop on Hot Topics in Cloud Services*, ser. HotTopiCS '13. New York, NY, USA: ACM, 2013, pp. 27–34.

[7] B. R. Kandukuri, A. Rakshit *et al.*, "Cloud security issues," in *Services Computing, 2009. SCC'09. IEEE International Conference on*. IEEE, 2009, pp. 517–520.

[8] S. A. Baset, "Cloud service level agreement," *Encyclopedia of Cloud Computing*, p. 433, 2016.

[9] F. Faniyi and R. Bahsoon, "A systematic review of service level management in the cloud," *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, p. 43, 2016.

[10] Z. Hu, L. Zhu, C. Ardi, E. Katz-Bassett, H. V. Madhyastha, J. Heidemann, and M. Yu, "The need for end-to-end evaluation of cloud availability," in *International Conference on Passive and Active Network Measurement*. Springer, 2014, pp. 119–130.

[11] M. Naldi, "Accuracy of third-party cloud availability estimation through icmp," in *Telecommunications and Signal Processing (TSP), 2016 39th International Conference on*. IEEE, 2016, pp. 40–43.

[12] ——, "ICMP-based third-party estimation of cloud availability," *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 6, no. 1, pp. 11–18, 2017.

[13] E. Casalicchio and L. Silvestri, "Mechanisms for SLA provisioning in cloud-based service providers," *Computer Networks*, vol. 57, no. 3, pp. 795–810, 2013.

[14] G. Aceto, A. Botta, W. De Donato, and A. Pescapè, "Cloud monitoring: A survey," *Computer Networks*, vol. 57, no. 9, pp. 2093–2115, 2013.

[15] H. J. Syed, A. Gani, R. W. Ahmad, M. K. Khan, and A. I. A. Ahmed, "Cloud monitoring: A review, taxonomy, and open research issues," *Journal of Network and Computer Applications*, 2017.

[16] A. Stephen, S. Benedict, and R. A. Kumar, "Monitoring IaaS using various cloud monitors," *Cluster Computing*, pp. 1–13, 2018.

[17] A. Mdhaffar, R. B. Halima, M. Jmaiel, and B. Freisleben, "Reactive performance monitoring of cloud computing environments," *Cluster Computing*, vol. 20, no. 3, pp. 2465–2477, 2017.

[18] S. Alboghdady, S. Winter, A. Taha, H. Zhang, and N. Suri, "C'mon: Monitoring the compliance of cloud services to contracted properties," in *Proceedings of the 12th International Conference on Availability, Reliability and Security*. ACM, 2017, p. 36.

[19] K. Alhamazani, R. Ranjan, K. Mitra, F. Rabhi, P. P. Jayaraman, S. U. Khan, A. Guabtni, and V. Bhatnagar, "An overview of the commercial cloud monitoring tools: research dimensions, design issues, and state-of-the-art," *Computing*, vol. 97, no. 4, pp. 357–377, Apr 2015.

[20] 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.

[21] ——, "Network protection through insurance: Premium computation for the on-off service model," in *8th International Workshop on the Design of Reliable Communication Networks, DRCN 2011, Krakow, Poland, 10-12 October, 2011*, 2011, pp. 46–53.

[22] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1.

[23] A. C. Davison and D. V. Hinkley, *Bootstrap methods and their application*. Cambridge university press, 1997, vol. 1.

[24] J. Dunne and D. Malone, "Obscured by the cloud: A resource allocation framework to model cloud outage events," *Journal of Systems and Software*, vol. 131, pp. 218–229, 2017.

[25] M. Naldi, "The availability of cloud-based services: Is it living up to its promise?" in *Design of Reliable Communication Networks (DRCN), 2013 9th International Conference on the*. IEEE, 2013, pp. 282–289.

[26] ——, "A note on "The Need for End-to-End Evaluation of Cloud Availability"," *arXiv preprint arXiv:1408.0510*, 2014.

[27] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.

[28] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[29] L. Mastroeni and M. Naldi, "Insurance pricing and refund sustainability for cloud outages," in *Economics of Grids, Clouds, Systems, and Services*, C. Pham, J. Altmann, and J. Á. Bañares, Eds. Cham: Springer International Publishing, 2017, pp. 3–17.