End-to-End Availability of Cloud Services

Victor Netes
Moscow Technical University of Communications and Informatics
Moscow, Russia
vicnet@yandex.ru

***Abstract*—Currently, cloud services are rapidly developing and are becoming more widely used. One of their most important attributes is availability that determines the need for research in this area. In the past, dozens of publications focused on the availability of cloud services. However, some of them provide only qualitative consideration, others do not consider whole infrastructure of cloud services from end to end. This paper clarifies the definition of service availability and gives quantitative evaluation of end-to-end availability of cloud services for typical situations taking into account all components of the service infrastructure. It shows that achieving high availability requires redundancy for such components as network connections and data centers.**

I. Introduction

In the past decade we observed the rapid evolution of the cloud market. Cloud computing leads to the emergence of new services, and it is essentially changing the way services are built, provided and consumed. Along with this, as applications are moved from dedicated hardware into the cloud, they need to achieve the same or even more demanding levels of service than with the traditional approach.

One of the general requirements for cloud computing is service availability, service reliability and quality assurance [1]. Namely, it is recommended that the cloud service provider provides end-to-end quality of service assurance, high levels of reliability and continued availability of cloud services according to the service level agreement (SLA) with the cloud service customer. Therefore, availability is an important service attribute and metric included in SLAs [2].

There are many publications focused on the availability of cloud services. A comprehensive overview and analysis of them was made in [3]. However, some of them provide only qualitative consideration, others do not consider whole infrastructure of cloud services from end to end. In particular, many papers consider only computational resources of data centers, enabling the provision of cloud services. A good example is [4], which proposes an analytical model for availability evaluation of cloud service provisioning system.

The main purposes of this paper are: to draw more attention to this topic, to clarify the definition of service availability, and to evaluate end-to-end availability of cloud services for typical situations taking into account all components of the service infrastructure. Herewith, general considerations will be detailed for the cloud service “Desktop as a Service” (DaaS).

Some of the key ideas of this paper previously were briefly presented in [5] and [6].

The rest of the paper is organized as follows. Section II gives definitions of availability, availability related measures and basic ways of their calculation. Section III presents work path for cloud services and its components. In Section IV typical availability target values are discussed. Section V contains calculations of end-to-end availability for different variants of cloud service infrastructure. An example of assessment requirements for availability of DaaS is given in Section VI. Concluding Section VII gives main findings and directions for future work.

II. Availability and its measures

A. Standard definitions

In the beginning there will be given definitions of basic concepts. This will be used by the official international standards of International Electrotechnical Commission (IEC), International Organization for Standardization (ISO) and International Telecommunication Union (ITU). These three organizations constitute the World Standards Cooperation.

According to the terminological International Standard [7], availability is one of the dependability attributes (characteristics). Availability is defined for an item as its ability to be in a state to perform as required. Availability depends upon the combined characteristics of the reliability, recoverability, and maintainability of the item, and the maintenance support performance.

An item in this standard is defined as a subject being considered. It may be an individual part, component, device, functional unit, equipment, subsystem, or system. The item may consist of hardware, software, people or any combination thereof.

The International Standard [8] deals with services and defines availability as the ability of a service or service component to perform its required function at an agreed instant or over an agreed period of time.

The International Standard [9] devoted to security and gives another definition of availability: property of being accessible and usable upon demand by an authorized entity. The basic International Standard on cloud computing [10] uses this definition and clarifies that in cloud computing the "authorized entity" is typically a cloud service customer.

Comparing these definitions, one can conclude that the definition from [7] is the most universal and clear. It is not a coincidence, as the standard [7] gives the general terminology used in the field of dependability; its terms are generic and are applicable to all fields of dependability methodology. Besides that, [7] gives not only the definition, but also specifies the relationship of availability with other dependability attributes and provides availability related measures, which are necessary for quantitative evaluation of availability. They will be considered and used below.

However, strictly speaking, the definitions of availability from [7] cannot be directly applied for a service, because a service is not an item. In order to overcome this formal obstacle, the following way is proposed. For each service an appropriate work path can be identified. It is a set of components or subsystems that should perform as required in order to the service be provided. It can be regarded as an item and its availability considered as availability of the service. Anyway, this work path has to be specified for service availability calculation.

B. Availability measures and their calculation

In [7] several availability related measures are defined. The most commonly used of them is steady state availability. It is the limit of the probability that an item is in a state to perform as required at a given instant when the time tends to infinity. As a rule, it may be expressed as the ratio of the mean up time to the sum of the mean up time and mean down time. Usually, steady state availability is called merely availability and denoted by A.

By the way, the definitions of service availability in some ITU-T Recommendations actually define this measure. Namely, in X.140 [11] service availability is defined as the ratio of the aggregate time during which satisfactory or tolerable service is, or could be, provided to the total observation period; in E.860 [12] it is the percentage of time during which the contracted service is operational at the respective service access points.

The note to the definition of availability in [8] also states that availability is normally expressed as a ratio or a percentage of the time that the service or service component is actually available for use by the customer to the agreed time that the service should be available.

Availability is often expressed in percent, but this form is unsuitable for calculations, so it will not be used here. Besides that, availability can be expressed through a downtime per year or other time period. For example, SLAs often refer to monthly downtime in order to calculate service credits to match monthly billing cycles. Downtime for the period T corresponding to the availability A is calculated by the formula

$$D_T = (1 - A)T.$$

Popular values of availability and the corresponding downtimes are shown in the Table I. Availability values 0.99 (99 %), 0.999 (99.9 %), etc. are often called “two nines”, “three nines”, etc.

TABLE I. POPULAR VALUES OF AVAILABILITY AND CORRESPONDING DOWNTIMES

Availability	Downtime per year	Downtime per month
0.99	3.65 days	7.20 hours
0.995	1.83 days	3.60 hours
0.999	8.76 hours	43.8 minutes
0.9995	4.38 hours	21.56 minutes
0.9999	52.56 minutes	4.38 minutes
0.99995	26.28 minutes	2.16 minutes
0.99999	5.26 minutes	25.9 seconds
0.999999	31.5 seconds	2.59 seconds

Availability of an item can be calculated on the base of its mean operating time between failures (MTBF) and mean time to restoration (MTTR):

$$A = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}. \quad (1)$$

Availability of a system usually can be calculated based on availabilities of its elements. The simplest class of systems is series ones. In reliability analysis a series system is a system that fails if any one of its elements fails, i.e. such a system does not have any redundancy. Its availability A is calculated by the formula

$$A = \prod_{i=1}^{n} A_i, \quad (2)$$

where A_i is availability of the ith element, n is the number of elements in the system.

This formula has a simple but important corollary:

$$A < \min_{i=1,\ldots,n} A_i, \quad (3)$$

that is availability of a system is less than the minimum of availabilities of its elements.

One of the main ways to increase availability is redundancy. Its simplest example is combining few elements in a parallel configuration. All of parallel elements must fail for the parallel system to fail. Availability of such a system is calculated as follows

$$A = 1 - \prod_{i=1}^{n} (1 - A_i). \quad (4)$$

Availability of more complex systems whose structure is a combination of serial and parallel configurations is calculated by the joint application of (2) and (4).

III. WORK PATH FOR CLOUD SERVICES

A. Block diagram

The conceptual aggregated work path for cloud services shown in the form of a block diagram is presented at Fig. 1. It has series configuration. Each of its components (blocks) will be described below.

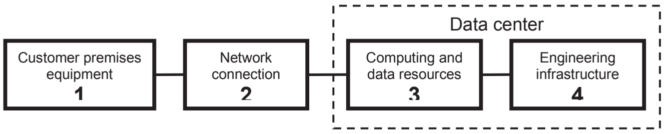

Fig. 1. The work path for cloud services

B. Components

1) Customer premises equipment (CPE): This component includes customer's computer or thin client, LAN, power supply, etc. It is important to note that responsibility for the operation of this equipment usually falls on a customer, not a service provider. Its availability will be considered only in Section VI.

2) Network connection: This component is a connection in a communication network that provides customer interaction with data center. It may be in public Internet, a private network for private cloud, VPN.

Basic Internet connections have availability about 0.99, whereas connections in MPLS VPN are much more reliable and their typical availability is 0.9999 [13]. To achieve higher availability, redundancy can be used, for example, a combination of the two connections from different providers. The appropriate reliability block diagram is shown at Fig. 2. For such situations total availability of a compound network connection can be calculated by the formula (4). Possible options are summarized in the Table II.

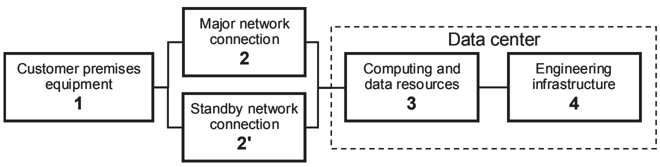

Fig. 2. The block diagram with network connection redundancy

TABLE II. AVAILABILITY OF NETWORK CONNECTIONS

Network connections	Availability
Basic Internet connection	0.99
Two separate Internet connections	0.9999
MPLS VPN connection	0.9999
MPLS VPN and Internet connections	0.999999

3) Computing and data resources: They are servers, storage, etc. within data center. This is the key component that processes and stores customer's data. Typical values of its availability will be presented below in Section IV.

4) Engineering infrastructure: Data center engineering infrastructure includes: ventilation and air conditioning; cooling; power supply, grounding and lighting; structured cabling systems; safety and security systems. They are support systems that are necessary to ensure operation of computing and data resources.

There are several classifications of data centers, which define levels of data centers and describe requirements for data center infrastructure. The most known of them is presented in Uptime Institute's Tier Standards [14]. The Uptime Institute is a data center research and professional-services organization based in Seattle, USA. The higher the Tier level, the greater the expected availability. This classification is applied to each of the three phases of a data center: its design documents, the constructed facility and its ongoing operational sustainability.

The higher the availability needs of a data center, the higher the capital and operational expenses of building and managing it. Therefore, data centers of the highest level (Tier IV) are not very numerous. For example, in Russia there are 27 data centers certified by the Uptime Institute, among them 26 have Tier III level and only one obtained Tier IV certification of design documents [15].

The basic requirements for Tier levels are briefly shown in the Table III. The table also contains typical values of availability and annual downtime. It is worth noting that they are not the standard requirements, but only the estimates made by the Uptime Institute based on its experience.

TABLE III. THE BASIC REQUIREMENTS FOR DATA CENTERS AND THEIR AVAILABILITY

Tier level	Redundancy and protection	Availability	Downtime per year
I	No redundancy	0.99671	1.20 days
II	Partial redundancy in power and cooling	0.99749	21.99 hours
III	N+1 fault tolerant. 72 hour power outage protection	0.99982	1.58 hours
IV	2N+1 fully redundant. 96 hour power outage protection	0.99995	26.28 minutes

C. Two views on end-to-end availability

Speaking about end-to-end availability of cloud services, we should distinguish end user's and service provider's points of view. According to end user's vision, all components of the working path including CPE (blocks 1–4 at Fig. 1) should be taken into account. Therefore, this availability in accordance with (2) is expressed as

$$A_{EU} = A_1 \cdot A_2 \cdot A_3 \cdot A_4.$$

From service provider's point of view, CPE should not be taken into account, because it is outside his scope of responsibility. Therefore,

$$A_{SP} = A_2 \cdot A_3 \cdot A_4. \quad (5)$$

In practice there may be situations when the different components belong to different owners. However, a customer needs a single service provider who will be responsible for all aspects of the service delivery through the whole work path. For this purpose, it is recommended to apply the concept of one stop responsibility [12]. The one stop responsibility, agreed by a provider to a customer within a SLA, allows a user to retain a primary service provider, (with whom he agreed on the SLA) as the only responsible for the overall service received. In its turn, the primary provider, since occurring problems depend on services received by other service providers, can apply the same one stop responsibility to its sub-providers. TM Forum's guidebook [16] also describes the same concept, only it uses the term "Integrator Role" to a primary service provider.

Of course, only service provider's point of view may be present in SLAs. Not without reason, in the above mentioned definition of service availability in [12] it is said about service access points. Usually, in telecommunications such demarcation point between the responsibility of the service provider and the responsibility of the customer is called User Network Interface (UNI). At Fig. 1 it is located between the blocks 1 and 2.

Therefore, further in the paper, availability will be considered mainly from service provider's point of view. End user's point of view will be used only in Section VI.

IV. Target values

A. General classifications

There are different classifications of availability levels in communication and information technologies.

One of them, the so called availability environment classifications (AEC), was proposed by Harvard Research Group (HRG) [17]. HRG provides tightly focused research and consulting services on cloud computing, big data, information and data management, high performance technical computing, highly available systems, etc.

The five Availability Environments (AE) define availability in terms of the impact on the both the business and the end user or consumer [17]. They are briefly shown in the Table IV.

TABLE IV. Availability Environments

AE	Requirement
0	Business functions can be interrupted and availability of the data is not essential. To the user work stops and uncontrolled shutdown occurs. Data may be lost or corrupted
1	Business functions can be interrupted as long as the availability of the data is insured. To the user work stops and an uncontrolled shutdown occurs. However, data availability is ensured. A backup copy of data is available on a redundant disk and a log-based or journal file system is being used for identification and recovery of incomplete transactions
2	Business functions allow minimally interrupted computing services, either during essential time periods, or during most hours of the day and most days of the week throughout the year. This means the user will be interrupted but can quickly relog on. However, they may have to rerun some transactions from journal files and they may experience some performance degradation
3	Business functions require uninterrupted computing services, either during essential time periods, or during most hours of the day and most days of the week throughout the year. This means that the user stays on-line. However, the current transaction may need restarting and users may experience some performance degradation
4	Business functions demand continuous computing and any failure is transparent to the user. This means no interruption of work; no transactions lost; no degradation in performance; and continuous 24×7 operation

Some publications introduce the additional highest level AE5 called "Disaster Tolerant" or "Disaster Resistant" with the requirement that the function must be available under all circumstances [18], [19], [20].

Besides that, there are also availability classes whose numbers correspond to the number of nines in the availability value (see Table I), and the correspondence between these classes and AEC levels [19], [20]. They are summarized in the Table V.

TABLE V. Availability classes and AEC levels

AEC level	Availability class	Availability
0	1	0.9
–	2	0.99
1	3	0.999
2	4	0.9999
3	5	0.99999
4	6	0.999999
5	7	0.9999999

These requirements can be applied to both computing resources within data center (the block 3 at Fig. 1) and cloud services on the whole (see the next subsection).

It should be noted that the upper classes give rather desirable than the actual levels of availability. As J.-C. Laprie pointed out [21], availability of current Web sites is one or two nines, well managed computing systems have four nines, though classical essential infrastructures require availability of at least five nines.

B. Availability levels for cloud services

Most of the cloud providers offer approximately 0.9999 of availability in their SLAs [4]. More rationally to differentiate the target values depending on the situation. As stated in [22], traditional public cloud availability delivers 0.999, enterprise market requires 0.9999 and even more, i.e. 0.99999 for IaaS services availability. In SLAs these three levels sometimes are called Bronze, Silver and Gold respectively (see as example [18]).

V. Calculations of end-to-end availability

A. The case of one data center

The relation (3) allows directly obtain certain limitations. For example, overall availability 0.9999 can be achieved only using the data center of the Tier IV, however, even this data center does not allow to obtain availability greater than 0.99995.

Particular values of end-to-end availability can be calculated using the formula (5). The results of calculations for 10 various variants are shown in the Table VI. The calculations used the data of Tables II, III and V. The values of components' availability in each variant were chosen so that they were not too much different.

These results show that in many cases the bottleneck of the work path is engineering infrastructure of the data center. As already mentioned, the data centers of the highest Tier IV are very expensive and not very numerous, but only such data centers allow obtaining the "Silver" level of availability (0.9999) and even they do not allow achieving the "Gold" level (0.99999).

TABLE VI. END-TO-END AVAILABILITY WITH ONE DATA CENTER

Availability class or Tier level			End-to end availability
Network connection	Computing and data resources	Engineering infrastructure	
2	2	I	0.976880
2	2	II	0.977640
2	3	II	0.986530
4	3	II	0.996390
4	4	III	0.999620
4	5	III	0.999710
6	5	III	0.999809
6	6	III	0.999818
6	5	IV	0.999939
6	6	IV	0.999948

In addition, the use of only one data center is not desirable keeping in mind to ensure survivability, i.e. the ability to continue to function during and after a natural or man-made disturbance. All this leads to the idea of redundancy for data centers with geographical separation of their locations (georedundancy) [23]. The calculations of end-to-end availability for such a case will be given below.

B. The case of two data centers

Consider the case when there are two data centers: primary and backup. Failover occurs in the event of a failure in the primary center. Accordingly, two work paths are considered (Fig. 3). Network connections in these paths may be compound, i.e. each of them can actually contain more than one network connection.

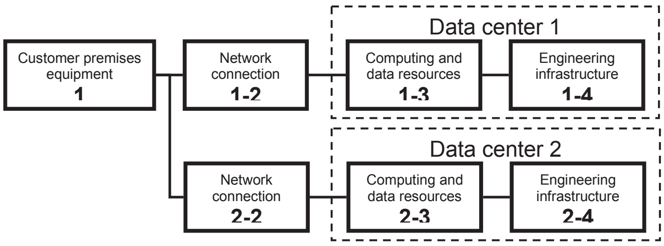

Fig. 3. The block diagram with data center redundancy

The calculation of availability for the case under consideration cannot be made by direct application of the formulas for serial and parallel configurations (2) and (4). Taking into account that the second (backup) work path is used when the primary data center is in down state, the probability of which is its unavailability, the overall availability can be written as

$$A = A_{WP1} + U_{DC1} \cdot A_{WP2} = A_{WP1} + (1 - A_{DC1}) \cdot A_{WP2} = \\ = A_{1\text{-}2} \cdot A_{1\text{-}3} \cdot A_{1\text{-}4} + (1 - A_{1\text{-}3} \cdot A_{1\text{-}4}) \cdot A_{2\text{-}2} \cdot A_{2\text{-}3} \cdot A_{2\text{-}4},$$

where A_{WPi} is availability of the ith work path ($i = 1, 2$), A_{DC1} and U_{DC1} are availability and unavailability of the primary data center respectively ($A_{DC1} + U_{DC1} = 1$), $A_{i\text{-}j}$ is availability of the jth component in the ith work path (j= 2, 3, 4).

For simplicity, it is assumed that similar components in both work paths have the same availability:

$$A_{1\text{-}2} = A_{2\text{-}2},\ A_{1\text{-}3} = A_{2\text{-}3},\ A_{1\text{-}4} = A_{2\text{-}4}.$$

The results of calculations for the same 10 variants as in the Table VI are shown in the Table VII.

TABLE VII. END-TO-END AVAILABILITY WITH TWO DATA CENTERS

Availability class or Tier level			End-to end availability
Network connection	Computing and data resources	Engineering infrastructure	
2	2	I	0.989833
2	2	II	0.989841
2	3	II	0.989993
4	3	II	0.999887
4	4	III	0.999900
4	5	III	0.999900
6	5	III	0.999999
6	6	III	0.999999
6	5	IV	0.999999
6	6	IV	0.999999

These results show that not only the “Silver” (0.9999), but even the “Super-gold” or “Platinum” (0.999999) levels of availability can be obtained by using two Tier III data centers.

VI. EXAMPLE: DAAS

Discussed above, the target values for cloud service availability may seem somewhat abstract. Consider the example of cloud service in which they arise naturally. It is Desktop as a Service (DaaS). This service is not as well known as SaaS, PaaS, and IaaS, therefore, brief information about it will be presented.

Requirements and functional architecture for DaaS are described in ITU-T Recommendations Y.3503 [24] and Y.3504 [25] respectively. According to [24] and [25], DaaS is a cloud service category in which the capabilities provided to the cloud service customer are the ability to build, configure, manage, store, execute and deliver users' desktop functions remotely. Instead of maintaining and running a desktop operating system and applications on customer devices, servers of a cloud service provider located in the cloud are used to execute the instances of users' virtual desktops.

One of the DaaS general requirements is high availability [24], [25]. The transition from a traditional personal computer to DaaS, on the one hand, reduces the availability by adding such components as network connection and data center, but on the other hand, it allows to replace a personal computer with a more simple and reliable thin client. Actually, a thin client is a common customer device for DaaS and such a case will be considered further.

The natural desire of the user is that the availability when using DaaS was not lower than of a conventional personal computer. In this situation availability should be considered from end user’s point of view.

Let A_{PC} and A_{TC} be availabilities of a personal computer and a thin client. Then the above condition means that

$$A_{TC} \cdot A_{DaaS} \geq A_{PC},$$

where A_{DaaS} is DaaS availability (from service provider’s point of view). This inequality can be rewritten as

$$A_{\text{DaaS}} \geq A_{\text{PC}}/A_{\text{TC}}. \quad (6)$$

Therefore, the ratio A_{PC}/A_{TC} can be considered as the target value for DaaS availability.

To calculate availabilities of a personal computer and a thin client one need to know their MTBF and MTTR. The MTBF for a thin client is about 175,000 hours according to [26] and 120,000 hours according to [27], both of these references give 25,000 hours as the MTBF for a conventional personal computer. The MTTR depends not only on the maintainability of the device, but also on the maintenance support performance (effectiveness of an organization in respect of maintenance support) [7]. In order to cover various situations, for both types of devices consider few possible values of the MTTR: 1, 2, 4 and 24 hours. The values of availability for all of the above options calculated by (1) are shown in the Table VIII.

TABLE VIII. AVAILABILITY OF THIN CLIENT AND PERSONAL COMPUTER

MTTR, hours	Thin client with MTBF		Personal computer
	120,000 hours	175,000 hours	
1	0.999992	0.999994	0.999960
2	0.999983	0.999989	0.999920
4	0.999967	0.999977	0.999840
24	0.999800	0.999863	0.999041

Substituting into the right side of inequality (6) values from the Table VIII, we get the results shown in the Table IX.

TABLE IX. POSSIBLE TARGET VALUES FOR DAAS AVAILABILITY

MTTR, hours	Thin client's MTBF	
	120,000 hours	175,000 hours
1	0.999968	0.999966
2	0.999937	0.999931
4	0.999873	0.999863
24	0.999241	0.999178

The results presented in this table show that the target values discussed in Section IV (three, four or five nines) are quite reasonable for DaaS.

VII. CONCLUSION

The main findings of this paper are the following: availability is an important characteristic of cloud services; it should be estimated from end to end, taking into account all components of a work path; achieving high availability requires redundancy, in particular, for network connections and data centers; the cost of this redundancy must be taken into account when conducting a feasibility study for cloud services.

Further work could be devoted to more detailed analysis. In particular, it can be taken into account the time required for switching from failed components to standby ones, availability of network connections between data centers, etc.

REFERENCES

[1] ITU-T Recommendation Y.3501 (06/2016). *Cloud Computing – Framework and High-Level Requirements.*

[2] *Cloud Computing Service Level Agreements. Exploitation of Research Results,* Editor D. Kyriazis, Brussels, European Commission, 2013.

[3] P.T. Endo, M. Rodrigues, G.E. Gonçalves, J. Kelner, D.H. Sadok and C. Curescu, "High availability in clouds: systematic review and research challenges", *Journ. of Cloud Computing: Advances, Systems and Applications,* vol. 5:16, 2016, pp. 1-15.

[4] F.M. Alturkistani, S.S. Alaboodi, "An analytical model for availability evaluation of cloud service provisioning system", *Intern. Journ. of Advanced Computer Sciences and Applications,* vol. 8, 2017, No 6, pp. 240-247.

[5] V.A. Netes, "Virtualization, cloud services and dependability", *Vestnik Sviazy*, 2016, No 8, pp. 7-9 (in Russian).

[6] V.A. Netes, "Service level agreements and dependability", *Dependability*, vol. 17, 2017, No 4, pp. 27-30.

[7] IEC 60050-192:2015. *International Electrotechnical Vocabulary – Part 192: Dependability.*

[8] ISO/IEC 20000-1:2011. *Information Technology – Service Management – Part 1: Service Management System Requirements.*

[9] ISO/IEC 27000:2016. *Information Technology – Security Techniques – Information Security Management Systems – Overview and Vocabulary.*

[10] ISO/IEC 17788:2014 | ITU-T Recommendation Y.3500 (08/2014). *Information technology – Cloud computing – Overview and vocabulary.*

[11] ITU-T Recommendation X.140 (09/92). *General Quality of Service Parameters for Communication via Public Data Networks.*

[12] ITU-T Recommendation E.860 (06/2002). *Framework of a Service Level Agreement.*

[13] J. Dix, "Enterprise WAN connectivity: MPLS VPN vs. public Internet", *Network World,* September 12, 2011, Web: https://www.networkworld.com/article/2220617/tech-debates/enterprise-wan-connectivity--mpls-vpn-vs--public-internet.html.

[14] The Uptime Institute official website, Tiers Program Publications, Web: https://uptimeinstitute.com/resources/.

[15] The Uptime Institute official website, Tier Certification Map, Web: https://uptimeinstitute.com/TierCertification/certMaps.php.

[16] TM Forum GB917, *SLA Management Handbook,* Rel. 3.1., 2012.

[17] Harvard Research Group official website, Availability Environment Classifications (AEC), Web: http://www.hrgresearch.com/pdf/AEC%20Defintions.pdf.

[18] High Availability in Cloud Computing. Zimory GmbH, 2013, Web: http://www.zimory.com/fileadmin/zimory/Downloads/Whitepaper/Product-HighAvailabilityinCloudComputing-270813-1611-134.pdf.

[19] J. Ahrends, Minimal Downtime Migration. Quest Software, 2006, Web: https://www.nloug.nl/downloads/ogh_minimal_downtime_migration.pdf.

[20] J. Kaskade, What Does HA in the Cloud Mean?, Web: http://jameskaskade.com/?p=448.

[21] J.-C. Laprie, Dependability and Resilience of Computing Systems, LAAS-CRNS, 2010, Web: http://webhost.laas.fr/TSF/Dependability/pdf/1-Jean-ClaudeLaprie.pdf.

[22] A.P. Czarnowski, Service availability (in the clouds). Warsaw, 2014, Web: http://ec.europa.eu/justice/contract/files/expert_groups/discussion_paper_service_availability_en.pdf.

[23] H.M. Brotherton, J.E. Dietz, "Data center site redundancy", *in Proc. 10th Intern. Conf. of the Intern. Inst. for Infrastructure Resilience and Reconstruction,* May 2014, pp. 2-6.

[24] ITU-T Recommendation Y.3503 (05/2014). *Requirements for Desktop as a Service.*

[25] ITU-T Recommendation Y.3504 (06/2016). *Functional Architecture for Desktop as a Service.*

[26] White Paper "Thin Clients". 2X Software Ltd., 2005, Web: http://cdn3.dabcc.com/resources/2X-Thin-Client-Guide.pdf.

[27] A Hemispheric Initiative. Thin Client Computing: Installation Manual. La Paz – Bolivia, 2007, Web: http://www.reciclemos.net/docs/pdf%20ingles/thinclientpaper-en6.pdf.