REPUBLIC OF MACEDONIA
UNIVERSITY "ST. KLIMENT OHRIDSKI"- BITOLA
FACULTY OF INFORMATION
AND COMMUNICATION TECHNOLOGIES BITOLA

Saso Nikolovski

RELIABILITY AND PERFORMANCE ANALYSIS OF CLOUD-BASED DISASTER RECOVERY SYSTEMS

- EXTENDED ABSTRACT OF THE DOCTORAL THESIS -

Bitola, 2023

MEMBERS OF THE COMMISSION FOR EVALUATION AND DEFENSE OF DOCTORAL THESIS

PECE MITREVSKI, Ph.D.
FULL PROFESSOR
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ST. KLIMENT OHRIDSKI UNIVERSITY - BITOLA

BLAGOJ RISTEVSKI, Ph.D.
FULL PROFESSOR
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ST. KLIMENT OHRIDSKI UNIVERSITY - BITOLA

NIKOLA RENDEVSKI, Ph.D.
ASSOCIATE PROFESSOR
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ST. KLIMENT OHRIDSKI UNIVERSITY - BITOLA

SASO JOSIMOVSKI, Ph.D.
FULL PROFESSOR
FACULTY OF ECONOMICS
SS. CYRIL AND METHODIUS UNIVERSITY - SKOPJE

BOZIDAR MILENKOVSKI, PhD.
FULL PROFESSOR
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ST. KLIMENT OHRIDSKI UNIVERSITY - BITOLA

In the modern environment of every company whose operation is based on information technologies, there is a need to adopt a plan to maintain business continuity, which includes a disaster recovery plan aimed at providing conditions and procedures for quick recovery in cases after an outage. Motivated by these two aspects for the maintenance of information systems and their information platforms, seeing the area of recovery after an outage or disaster as particularly interesting and challenging for research, within the framework of the dissertation a comparative analysis of recovery systems was made with a special emphasis on the systems which are partially or completely placed in the cloud and their reliability for the reliable and safe performance of the role they have in the overall scenario. Considering the actuality of services for protection and recovery of data and information systems placed in the cloud, for the needs of the research, in a real production environment, two systems for protection and recovery have been implemented.

To identify the values of the key parameters according to which their performability and reliability would be evaluated in given real conditions of their work, in the research simulation software was used to improve the performance of real systems and at the same time a System Dynamics analysis was performed for each of the considered systems.

In that direction, the paper identifies a framework with the parameters according to which a solution would be chosen for the protection and maintenance of data and information systems in organizational structures. The benefits of the framework proposed in this way refer primarily to determining the approach and stages in the selection and selection of an appropriate solution for the protection and recovery of data and information systems, through the possibility of adapting the chosen concept to the environment, according to the outlined time frames set in the plan for business continuity and the recovery plan from an outage. With the conducted analyzes and the conclusions drawn from them, a direct contribution is made to the correct thinking for decision-making when choosing concepts for the performance of recovery systems after an outage or catastrophic interruption, depending on the needs of the entities that implement the same.

1 PUBLISHED PAPERS RELATED TO THE RESEARCH

Scientific papers presented at international academic conferences:

1. S. Nikolovski, P. Mitrevski, "*On the Requirements for Successful Business Continuity in the Context of Disaster Recovery*", Proc. of the 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), Ohrid, North Macedonia, 2022
2. S. Nikolovski, P. Mitrevski, "*Data protection and recovery performance analysis of cloud-based recovery service*", Proc. of the 58th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), Nis, Serbia, 2023

Papers in international scientific journals:

1. S. Nikolovski, P. Mitrevski, "*Modelling and simulation of data protection systems for business continuity with disaster recovery*", submitted for publication in International Journal of Business Continuity and Risk Management (under review)

2 SUBJECT AND AIM OF THE RESEARCH

The constant availability of company information systems for many imposes an impression and opinion that business continuity can be interpreted or understood as protecting the future business and functional survival of the organization from some form of disruption.

With modern work processes that are based on digital systems, zero downtime during operational disruption considered from the perspective of sustained business, is an ideal outcome for all organizations. Such an expectation is not always possible or realistically achievable for numerous reasons (outages due to weather events, cyber attacks, etc.) despite the availability of numerous solutions for protection and recovery both in the local data centers of the organizations, as well as solutions based on using a system in cloud [3]. Therefore, from the aspect of management of such organizations, more and more attention is paid to reducing the impact of outages on the overall work process through the maximum estimated outage time that the organization itself can afford, without having permanent consequences for its further operation.

When analyzing the outages and setting the goals for consistent recovery from them, while making the same acceptable for the organization, it can be noted that the whole process is based primarily on the time in which the organization is out of operation, and at the same time it includes two periods of time dependent elements.

One element is determined by the system or technological aspect presented by the recovery time objective (RTO), and the other, which is more organizationally oriented, represents the time required for a full operational recovery of work processes (Work Recovery Time-WRT).

These two time components of the recovery process determine the Maximum Tolerable Downtime (MTD) provided by the Business Continuity Plan (BCP) and the Disaster Recovery Plan (DRP) and is the summary time of the recovery time objective and the required operational time return to work processes:

$$MTD = RTO + WRT \quad (1)$$

This means that RTO as a parameter is a time interval during which operations are performed in the technical-technological part of the organization, a time during which systems, data and network infrastructure are restored. The remaining time until the maximum tolerable outage time is the operational recovery time (WRT) and in it is carried out recovery of all work processes that are based on information and information systems (Figure 1).

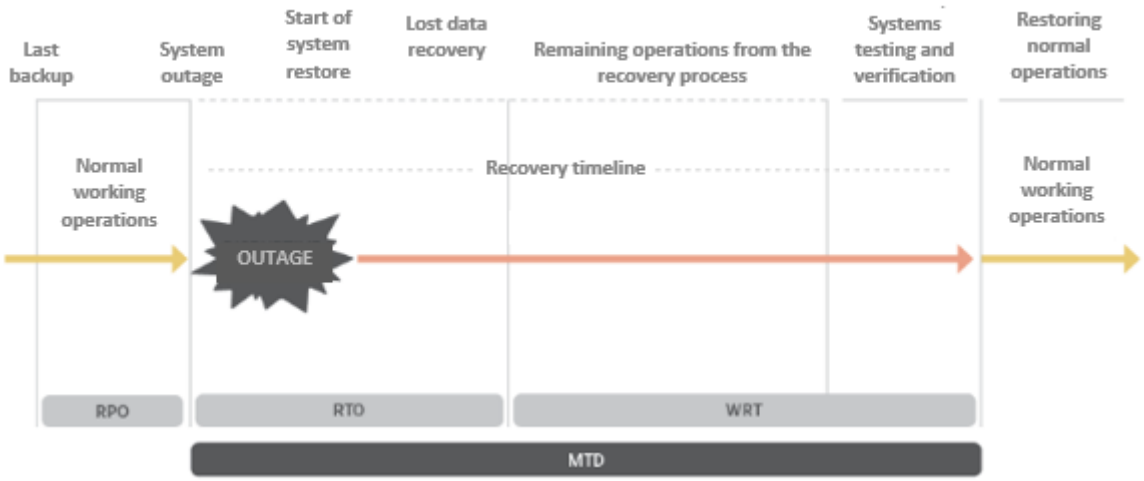

Figure 1 *Maximum Tolerable Downtime (MTD)*

The recovery time frame, which is limited by the framework in which the MTD is set, also includes several components that are an integral part of this process. These components aim to restore data from their backup (backup) which is closest in time to the outage, to carry out subsequent operations as the final part of the recovery and, of course, to check and test the functionality of the systems before they are officially operational to establish normal work in organization.

When defining the needs, in the system design phase, the component that is directly related to the data, and thus to the consistency of the information, is the recovery point objective (RPO). This parameter is time-dependent and gives the "age" of the data in the backup set at the point in time from which it will be restored to the systems for operational use. Considering that the RPO recovery process is backward recovery, loss of data and final information is inevitable (except in synchronous replication situations where data loss is zero). Therefore, when designing systems within the framework of sustainable business continuity planning, the maximum tolerance threshold is set for data loss that the organization can afford (Maximum Tolerable Data Loss-MTDL). Systems that enable zero data loss create an inversely proportional relationship between the amount of data lost during the recovery process and the cost of the systems. This means that the closer the RPO is to the time of the outage, the higher the cost of such systems and vice versa.

For that reason, using the above-mentioned time components as a starting point, within the framework of the research, an analysis of the performance and reliability of two systems for recovery after an outage was carried out, with the aim of obtaining a parametric framework for the selection and installation of systems that will satisfy the requirements of organizations during recovery processes from a technical-technological, organizational and financial aspect, which very often has a decisive influence when making the final choice.

When conducting the research, the overall process included a series of activities consisting of:

1. Creating a detailed project, setting up and configuring a data center including data protection based on a DellEMC DP4400 hardware device as an on-premises solution and Microsoft Azure Recovery Services (MARS) as a fully cloud-based solution
2. Installing and configuring data protection systems for research purposes. The number of samples of time and data parameters retrieved from the systems was matched with the values predicted by the business impact analysis (BIA) and the minimum number of samples supported by the protection systems (14 days from the hybrid system and 7 from the cloud-based system),
3. After a time period of one year (December 2021-December 2022) in which the systems had continuous operation, an analysis of their operation was performed by downloading the data on time and data parameters from the performance of backup and recovery operations) of the data, where the parameters that should be calculated from the values of the parameters taken from the protection systems are determined,
4. For each of the considered systems, two models have been created - one basic model, in which, by using the values taken from the systems for the time and data parameters, the values for the determined derived parameters in the model have been obtained, and one extended model, in which, by using the values of the derived parameters from the basic model are calculated values for the performance of the systems with a given test amount of data in the considered operating environment. Also, during the creation of the models, a model for the reliability of the systems was created, with time frames aligned with used time settings made in the systems,
5. A comparative analysis of the obtained results from the simulations of the models for the two systems was made and conclusions were drawn based on which recommendations for correct dimensioning and placement of the different protection concepts in different situations of use were given in the final text of the dissertation.

According to everything stated above, the focus of the research is set in the area of maintaining business continuity in the operations of organization entities, and the analysis of the reliability and performance of systems for disaster recovery is imposed as a subject of research, while the goal is set to building a parametric framework that will provide precise guidelines when choosing a system solution for the protection and maintenance of data and information systems in data centers.

3 MAIN BENEFITS

The research included in the text of the dissertation is based on a series of texts resulting from the research work in the past ten years, a period in which cloud services became relevant in the daily operations of company entities [4][7][8][20][26][27][28]. In the interest of these researches, several target parameters are constantly circulating, of which availability, data recovery point objective (RPO) and data recovery time objective (RTO) are commonly observed, as parameters directly dependent on the performance of data protection systems and their reliability.

In practice, regarding the applicability of the methods and algorithms developed in the indicated researches that were used as a starting point for the research carried out in the dissertation, the authors have pointed out that, for the most part, such researches were made in simulation conditions with the most common application of values of the key parameters obtained in an isolated environment, i.e. without influence from the environment in which they are located, and as a confirmation of their applicability, their validation is needed, obtained from their practical implementation in a production environment.

The systems used in the dissertation research are hosted in a production data center (Figure 2).

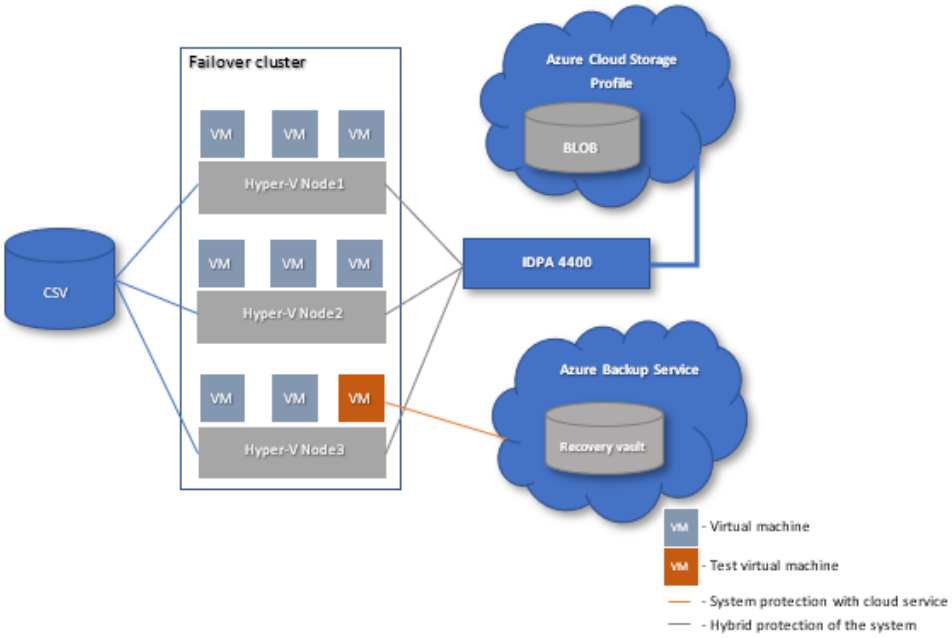

Figure 2 *Production data center – block diagram*

Taking into account that the research is carried out in a real working environment with high importance for stable functionality of all systems in it, for the purposes of the research a separate server system was set up as a test virtual machine (VM) that was used to monitor the processes for data protection (backup) and recovery (recover/restore) within the framework of

the two protection systems, which include protective storages placed in the cloud. In order to determine the performability of data protection systems, within the framework of the research, two basic and two extended models were developed in which, by simulating the operation of real systems, values for the target parameters that are of interest for drawing conclusions were obtained. The practical implementation of the systems used in the research includes the application of a hybrid solution (a hardware device with a data storage layer placed in the cloud) placed in the data center, with an Avamar software agent installed in the server system that is subject to data protection and a solution based entirely on on the Microsoft Azure Recovery Service (MARS) cloud technology, with the MARS agent installed in the server system. When backing up data using the hybrid solution, the backup is initially stored in the device's local storage for a pre-set period of time, after which the copy is transferred to the cloud storage tier for long-term keeping. In the case of backups using the MARS agent, the backups are uploaded and stored directly in the cloud storage, with a preset retention period.

3.1 Models, simulations, and results

When setting up the models for the protection systems used in the research, the main emphasis in them is placed on the time to perform backup operation and time to restore/recover data. What is significant about the developed models is that they use values taken from the real systems as a source for the input values of the variables. The initial values for data quantities are taken from the agents, for a time sample of 14 days in the case of the hybrid system and 7 days in the case of the cloud-based system, in order to cover the expiration of the copies, but also to show the relationship of the repositories in cloud with the conditions set in the data center for the transfer of data to them. As a starting point according to which the parameters taken when setting up the models are determined is the business impact analysis of outages (BIA), implemented within the protection policies in the data center. Values that are set as targets in BIA are given in Table 1, where made a separation of the values according to certain parameters for each of the systems separately.

In the table, for the hybrid system, the Avamar agent is listed as an agent installed in the server system that is being backed up, and MARS agent for the cloud system.

Agent	Backup frequency	Backup retention time	Recovery points in time	Cloud tiering policy	RPO	RTO
Avamar	daily	14	7+7+60	> 14 days	≤ 7 days	≤ 5 hours
MARS	daily	7	2*7+3	/		

Table 1 *Parameters values in BIA*

The backup scenario is fully compliant with the values given in Table 1. To monitor and evaluate the process of recovering lost data from a selected recovery point in time (RPO), a simulation of damage (deletion) of document folders is made, on which the recovery process has been carried out. Figure 3 shows, in a general form, the input-output parameters in the models of the two considered systems.

Figure 3 *Input-output parameters in the models*

From Figure 3, it can be noted that the amount of data transferred on a daily level to the storage systems and the time for which they were transferred to the storage systems, are set as input parameters for both models. Derived parameters for the time needed to protect a given amount of data and the time for its recovery within the system that is subject to protection are defined as output parameters from the models.

During creation of models for both systems, the basic models are initially set, which include a series of variables connected to each other with relations through which they change states of the resulting components and storage systems in their concepts. In order to determine the performability of the systems during the protection and recovery of large amounts of data, an extended model has been set for each of the systems in which additional components have been added through which specific values for the time parameters related to the processes for making backup copies and recovery have been obtained.

3.1.1 Hybrid system model

The basic model of the hybrid system is shown in Figure 4, where there are four derived components that cover the process of making backup copies of the data, as well as the process of recovery in case of data damage or lose. The time frame in which the simulation process takes place in the given model is aligned with the BIA, where the process of creating backup copies takes place in 14 time periods, and the process of recovering the data and moving it to the cloud storage level is set in one period subsequent to the backup time frame.

Values of the variables in the basic model are shown in Table 2 with separate views of the components of the backup process, the restore process and values of the derived variables in the model.

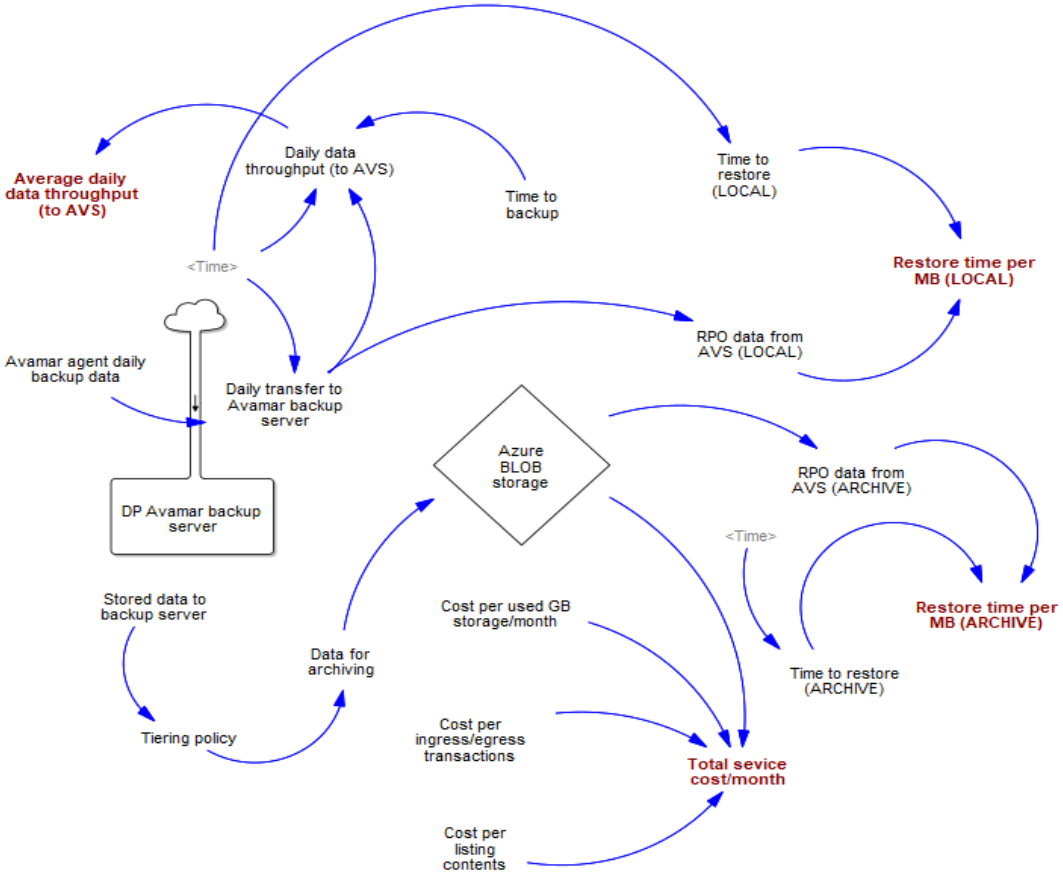

Figure 4 *Basic model of hybrid system with data tier in cloud*

In Figure 4, *Average daily data throughput (to AVS)* appears as derived component in the model, a derived variable obtained as a mean value of the amount of data transferred to the protection storage in the system (Avamar backup server - AVS) by the Avamar agent in the server system, *Restore time per MB (LOCAL)* variable which shows time required to restore a 1MB amount of data retrieved from the local storage of the DP4400 system, *Restore time per MB (ARCHIVE)* variable which shows the time required to restore a 1MB amount of data that was moved to the storage level of the device placed in the cloud and of course the derived variable *Total service cost/month* which shows the costs of using a storage service placed in the cloud.

Figure 5 provides a graphical states representation of input components in the model to present parameter values changes inside of the specified time frame.

Variable	Value														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Data backup process															
Avamar agent daily backup data (MB)	26956	25712	27194	26710	25147	24520	26529	19711	27342	19574	27024	25262	24644	26082	
Daily transfer to Avamar backup server (MB)	26956	25712	27194	26710	25147	24520	26529	19711	27342	19574	27024	25262	24644	26082	
Daily data throughput (to AVS) (MB)	51.3448	40.3022	49.8059	51.8661	55.7633	57.5587	45.7428	74.6629	43.0596	76.1693	47.9149	58.3472	55.132	51.4438	
Time to backup (minutes)	8.75	10.633	9.1	8.583	7.516	7.1	9.666	4.4	10.583	4.283	9.4	7.216	7.45	8.45	
Data restore process															
RPO data from AVS (ARCHIVE) (MB)															1824
RPO data from AVS (LOCAL) (MB)														1824.01	
Time to restore (ARCHIVE) (seconds)															470.1
Time to restore (LOCAL) (seconds)														38.24	
Derived variables															
Average daily data throughput (to AVS)															54.2224
Restore time per MB (LOCAL) (seconds)															0.0209649
Restore time per MB (ARCHIVE) (seconds)															0.25773
Total sevice cost/month (dollars)															1.57912

Table 2 *Values of variables set in the basic model*

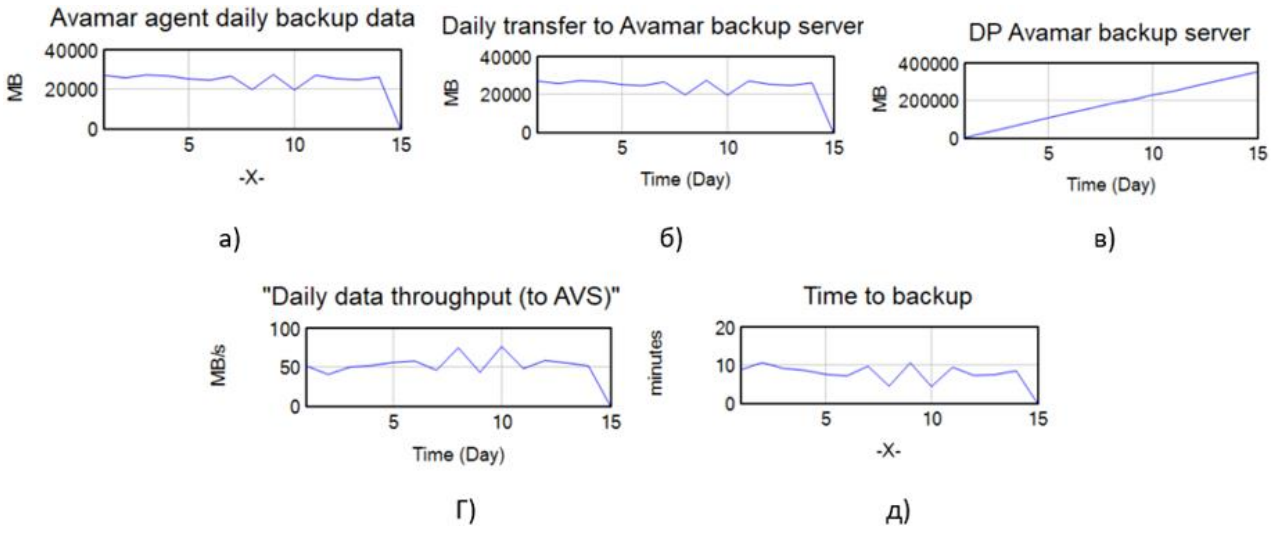

Figure 5 *Graphic representation of parameters states in backup process*

According to the subject and objectives of the research to determine the performability of the solution, the basic model of the system has been upgraded with five new components that have no influence on the values in the basic model (Figure 6).

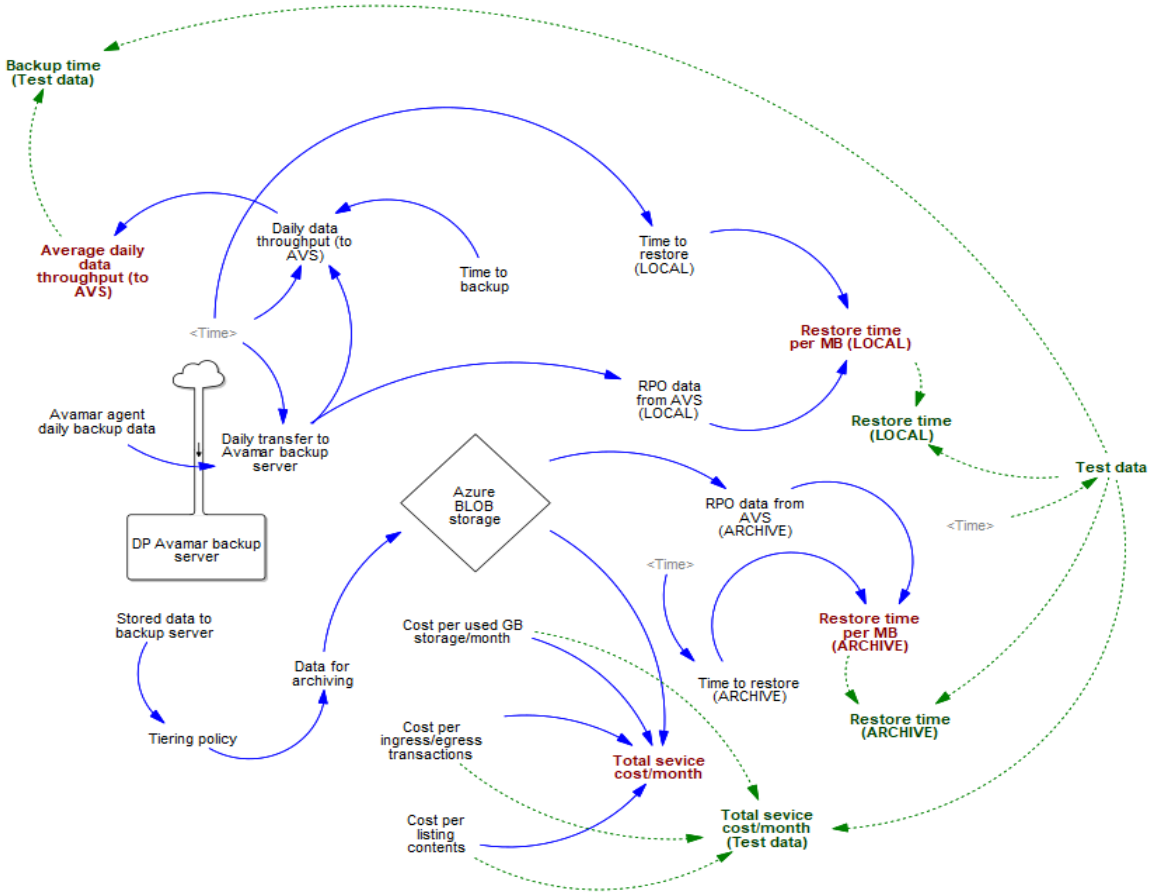

Figure 6 *Extended model of hybrid system with data tier in cloud*

From the presented view of the extended model, the relationships between the derived components of the basic model and the additional components for a given test amount of data in the extended model can be observed. As additional components in the model, the *Test data* component that enters the test amount of data (531 GB), *Backup time (Test data)* as a component that calculates the time required to make a backup copy of a given amount of data, *Restore time components (LOCAL)* and *Restore time (ARCHIVE)* in which the time needed to recover a given amount of data is calculated in the case when the process is performed from the local storage of the system with RPO ≤ 14 days or from the level of storage placed in the cloud where 15 ≤ RPO ≤ 60 days and the component *Total service cost/month (Test data)* which gives the total monthly costs for using the storage service of the cloud-based system.

The values of the additional components set in the expanded model of the system are shown in Table 3, from where it can be noted that some of the resulting components in the

expanded model repeatedly exceed the maximum allowed values for them provided for in the BIA, which makes them useless for the needs of the organization .

Variable	Value
Derived variables (basic model)	
Average daily data throughput (to AVS) (MB/s)	54.2224
Restore time per MB (ARCHIVE) (seconds)	0.25773
Restore time per MB (LOCAL) (seconds)	0.0209649
Test data simulation values (extended model)	
Test data (MB)	531012
Backup time (Test data) (hours)	2.72034
Restore time (ARCHIVE) (hours)	38.016
Restore time (LOCAL) (hours)	3.09239
Total sevice cost/month (Test data) (dollars)	11.6602

Table 3 *Simulation results of an extended model for a hybrid system*

Specifically, the *Restore time from (ARCHIVE)* component has a value many times higher (38 hours) than the maximum value provided in the BIA (≤5 hours) and therefore, the level set in the cloud, in this case could not be used for quick recovery of systems in the organization with acceptable downtime and returning its operation and functionality to the level before the occurrence of the outage. Against this value, the value of the *Restore time (LOCAL)* component is within the limits predicted by the BIA (3 hours and 5 minutes), which indicates that the system placement in the data center structure can satisfy the requirements seted in the analysis.

From the point of view of creating a backup copy, the values of the *Backup time (Test data)* component fully satisfy the requirements for quick creation of the copies in the terms provided for their implementation.

It should be taken into account that such systems have a complex architecture not only in hardware, but also in software, where through a series of algorithms over time of use, the time for which the copies will be made can be drastically shorter than what was taken as the average time in the basic model.

3.1.2 Model of system in cloud

The basic model of the cloud-based system is shown in Figure 7, where, just like in the hybrid system, four derived components appear, two of which relate to the process of creating a backup copy, one to the process of data recovery, and one to the monthly costs for using such a service completely set up in the cloud.

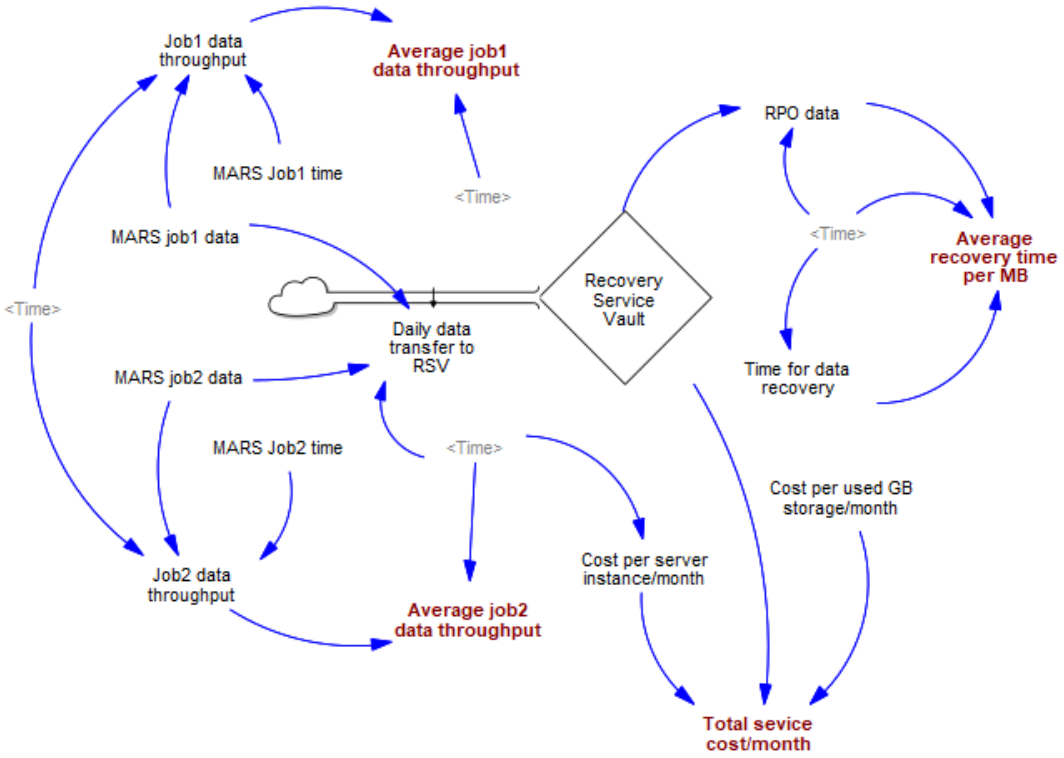

Figure 7 *Basic model of cloud-based system*

The presentation of the components from which the model is built is made in the same way as the hybrid solution was described. *Average job1 data throughput* and *Average job2 data throughput* appear as derived components in the model, as variables that show the average value of the amount of data transferred to the system's backup storage (Recovery service vault - RSV), the variable *Average recovery time per MB* that shows the time required to recover 1MB amount of data placed in the protective storage of the system and *Total service cost/month* as a derived variable that shows costs of using the storage service placed in the cloud.

The time period in which the procedures for making copies and restoring data from them are performed includes 8 time points, 7 of which are intended for performing the policy for making backup copies, and the last time point is intended for showing the change in quantity data that are placed in the recovery vault after the completion of the 7-day cycle, according to the values given in Table 1. The states of the values of the variables in the basic model and their changes in the set time frame are shown in Table 4, where a separate section also shows the final values of the derived components after a completed simulation with the given input parameters.

Figure 8 provides a graphical representation of the states of the input components in the model associated with the *MARS job1* process, and Figure 9 provides a graphical representation of the states of the components associated with the *MARS job2* process.

Variable	Value							
	1	2	3	4	5	6	7	8
Data backup process								
MARS job1 data (MB)	8362	8409	8463	8516	8569	8622	8678	
MARS Job1 time (sec)	3270	2993	3025	3110	2976	3105	3307	
Job1 data throughput (MB/s)	2.55719	2.80956	2.79769	2.73826	2.87937	2.77681	2.62413	
MARS job2 data (MB)	353	375	478	553	551	394	780	
MARS Job2 time (sec)	351	365	489	628	232	358	387	
Job2 data throughput (MB/s)	1.0057	1.0274	0.977505	0.880573	2.375	1.10056	2.0155	
Daily data transfer to RSV (MB)	8715	8784	8941	9069	9120	9016	9458	
Data recovery process								
RPO data (MB)								7690
Time for data recovery (sec)								1380
Derived variables								
Average job1 data throughput (MB/s)								2.57731
Average job2 data throughput (MB/s)								1.83045
Average recovery time per MB (sec)								5.57246
Total sevice cost/month (dollars)								7.82701

Table 4 *Value states of the variables in the basic model*

The obtained values from the simulation relating to the derived components shown in Table 4 are the basis for conducting an evaluation of the performability of the system when manipulating extreme values of the data components, which as such have a direct impact in changing the performance of the system.

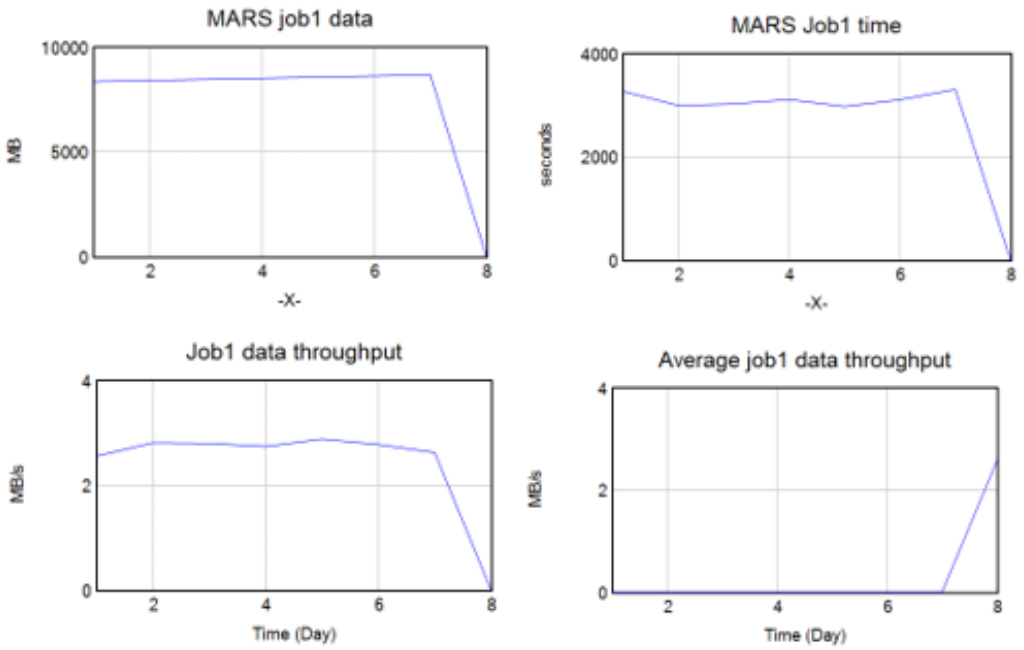

Figure 8 *Graphical representation of component values related to MARS job1*

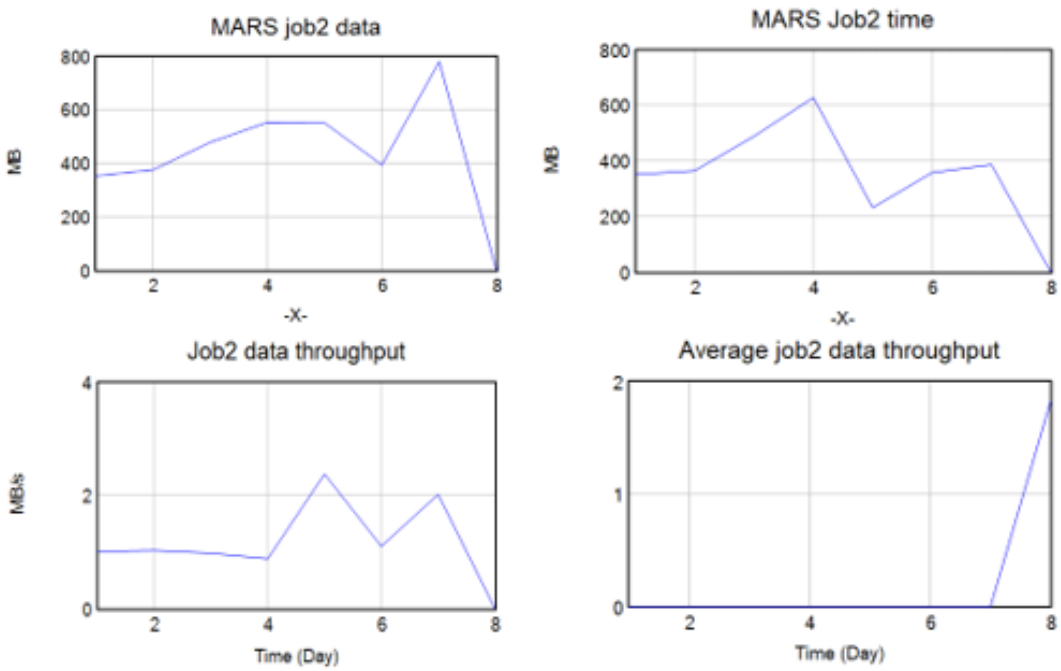

Figure 9 *Graphical representation of component values related to MARS job2*

For this purpose, five new components have been added to the model that is presented as basic and from which the derived components are used, of which:

- Two components *(Backup time Job1(Test data)* and *Backup time Job2(Test data))* for time calculation in backup creation with both jobs,
- a component that will refer to the data recovery process *(Recovery time (Test data))*,
- *Test data* as a common component for all, which will carry the value of the amount of data that a server system has in the organization (Test data = 531 GB) and
- one component *(Total service cost/month (Test data))* for calculating the costs of using the service with the new amount of data.

The view of the expanded model with the new components is given in Figure 10, and the results of the simulation are shown in Table 5.

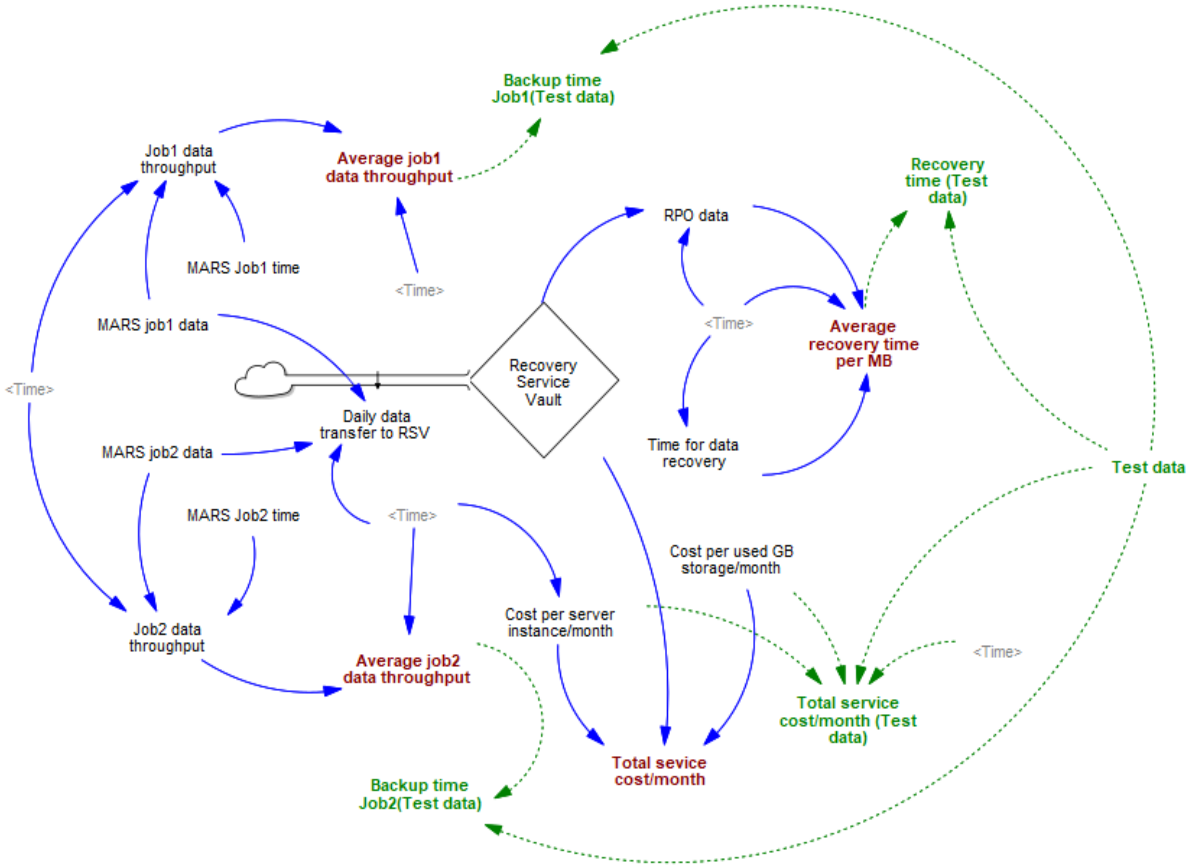

Figure 10 *Extended model of cloud-based system*

Variable	Value
Derived variables (basic model)	
Average job1 data throughput (MB/s)	2.57731
Average job2 data throughput (MB/s)	1.83045
Average recovery time per MB (sec)	5.57246
Test data simulation values (extended model)	
Test data (MB)	531012
Backup time Job1(Test data) (hours)	57.2315
Backup time Job2(Test data) (hours)	80.5831
Recovery time (Test data) (hours)	26.47
Total sevice cost/month (Test data) (dollars)	43.7893

Table 5 *Simulation results of an extended model for a cloud-based system*

Due to the high values of the time components, representation of these values in Table 5 is given in hours, and the data values in MB.

3.1.3 Comparative analysis

In order to obtain a complete picture of the performability of the two systems, a comparative analysis of the obtained results from the performed simulations of the presented models was made in the dissertation. By applying the values obtained from the data protection systems, in a comparative overview divided by operations in three tables, components that are marked as performed in the simulations are placed. In the first table (Table 6) a comparative overview of the amounts of data transfer in both systems during the process of creating a backup copy of the data is made.

SYSTEM	Hybrid (DP 4400)	CLOUD-BASED (Azure recovery service)	
Component	Average daily data throughput (to AVS)	Average job1 data throughput	Average job2 data throughput
Value (MB/s)	54.2224	2.57731	1.83045

Table 6 *Values of data transfer when creating a backup copy*

According to the given review, it is evident that the hybrid system has many times more data transfer per unit time between the Avamar agent and Avamar server (AVS) when performing the process for creating a backup copy. This is primarily due to the communication connection between the two endpoints of the process, where in the hybrid system the connection is through the local network, and in the cloud system the starting point is located in the data center (MARS agent), goes through the Internet operator (ISP) and it ends in the cloud-based service. Such changes to data transfers inevitably bring performance degradation, and the result is a total bandwidth with small values, which directly affect the duration of the process by which the copies are created. The comparative presentation of the simulation results of the data recovery processes with the resulting components that are derived from values obtained from these processes in both systems are shown in Table 7.

SYSTEM	Hybrid (DP 4400)		CLOUD-BASED (Azure recovery service)
Component	Restore time per MB (LOCAL)	Restore time per MB (ARCHIVE)	Average recovery time per MB
Value (sec)	0.02096	0.25773	5.57246

Table 7 *Time component values in recovery process for 1MB data*

From the results shown in the table, it can be seen that with the hybrid system, the process of recovering data from the device's local storage has a much shorter time than the process that requires data placed in the cloud storage.

The use of cloud storage by the described data protection and recovery systems also entails costs that result from the use of the cloud service. Table 8 shows these costs and the parameters according to which they were made.

SYSTEM	Hybrid (DP 4400)	CLOUD-BASED (Azure recovery service)
Cost per server instance/month	/	10
Cost per used GB storage/month	0.02	0.0448
Cost per ingress/egress transactions ($/10K operations)	0.54	/
Cost per listing contents ($/10K operations)	0.5	/
TOTAL COST for TEST DATA ($/month)	11.66	43.79

Table 8 *Monthly costs for using the cloud service for both systems*

The last row in the table shows the cost of using storage space equal to the amount of data (531 GB) that was used in the simulations for both systems in the cases of their extended models.

3.1.4 Reliability analysis

According to the layout of the components in the system that shows the concept of data protection from outages and disasters, it includes three components: the data center (DC), the cloud service (Azure) and the connection to the global network (ISP). With this concept of the system, it can be shown as a series connection between its components where the failure of any component in the system (not the operational binary state of the component, for which R=0 will apply), will mean the failure of the entire system . This way of thinking about the operation of the entire system is the basis for analyzing its reliability. When building a model for calculating and assessing the reliability of the system, in our case it implies setting more variables in it that will represent the conditions on which the reliability of each of its components depends. The initial conditions for setting up the concept of the model include: the period in which the data center will be used (predicted 7 years or 61320 hours), the period in which the system analysis is performed (15 days or 360 hours), the operational time of the connection operator to the global network (based on the usage contract of 24 months or 17520 hours) and the level of Service Level Agreement (SLA) of 99.95%.

The reliability analysis model of data protection concepts with the initial conditions set in this way is shown in Figure 11, and the results of the performed reliability simulation of each of the components and the system as a whole are given in Table 9.

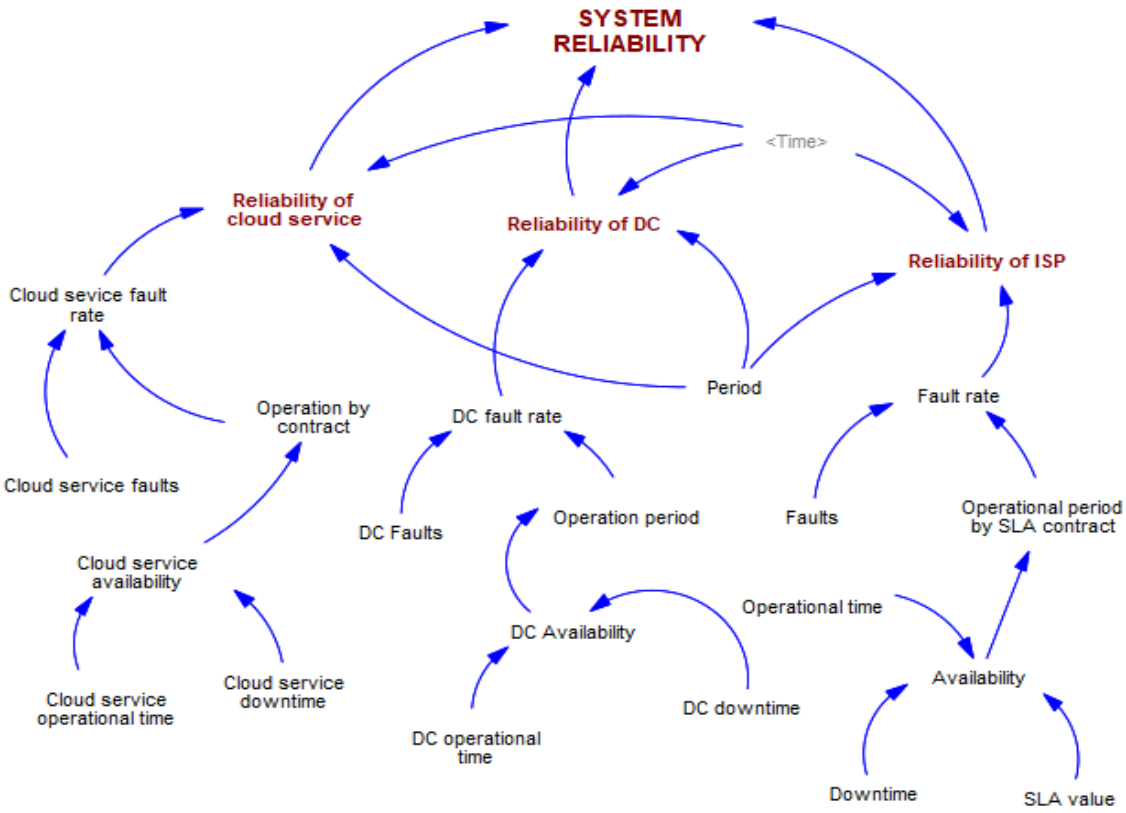

Figure 11 *A three-component system reliability model*

COMPONENT	VALUE
Reliability of cloud service	**0.993952**
Reliability of DC	**0.993952**
Reliability of ISP	**0.978981**
SYSTEM RELIABILITY	**0.967174**

Table 9 *Reliability of components and system for a given period*

From the presented results in Table 9, it can be concluded that despite the high reliability values of the cloud service and the data center (reliability for them in the initial conditions is set to R=1), the lower reliability value of the operator has a negative impact on the reliability of the whole system. It is due to the serial connection of these components in the execution of the processes that use the cloud service. In the Disaster Recovery Journal (DRJ)[1], it is stated that any scenario or concept of such a solution is as reliable as its connection to the global network.

Therefore, in concepts of systems that are based on a cloud service or a service of any kind that uses a connection to the global network, when designing such solutions, a careful assessment

[1] https://drj.com/journal_main/backup-and-disaster-recovery-services-only-as-reliable-as-your-network-connection

of the values in the SLA documents is mandatory, because small deviations in them contribute for noticeable changes in the characteristics of the rest of the components, and thus of the entire set solution.

3.1.5 Concluding observations

In modern business operations and its accelerated digital transformation, the importance of data and the systems in which they are found has been highlighted, placing data at the center of information exchange. This dependence of business processes imposes the development and installation of solutions for their protection, in order to maintain their business continuity. There are many solutions of systems with this purpose, each with its own advantages and disadvantages in the way they realize the protection of business continuity. Some of them are fully aimed at safe and secure storage of data (backup data), some are intended for their protection by storing active data in two or more locations, related to processes for permanent or occasional synchronization between them (replication), and part are intended for maintaining two or more data centers of the organizations, for maintaining uninterrupted business continuity. The main benefit of the research carried out in the dissertation refers to the determination of the parameters that have a key role in building the right approach to choosing a solution or solutions that will fully satisfy the requirements and needs set in the BIA, in order to successfully maintain the business continuity of the organizations whose operation is completely based on information systems. Based on this, the concluding observations in the dissertation are divided into conclusions from the technical aspect, from the aspect of characteristics and performance, and from the financial aspect, so that in the final part, insights and recommendations are given for possible future research to advance the approach to obtaining higher values for performability and reliability of data protection systems. Considering that the amount of data has a direct impact on the time components of the protection solutions, as future research directions should be the mechanisms that will contribute to a greater reduction of the data that will be exchanged between the systems in the entire protection process. From a technical point of view, as well as from the point of view of the performance of the protection solutions, according to the results obtained from the models developed for the two considered systems, use of a solution that will be placed locally by distributing the data in the cloud as a level of the entire storage of the device (single data namespace), provides sufficient protection in the event of an outage and meets the needs for a quick return to an operational and functional state of information and information systems used for daily operations. Regarding the costs of using the cloud service, it should be taken into account when such services are used to store large amounts of data, that the coefficient that increases the value of the Total service cost/month component the most is exactly the part that refers to the billing for occupying storage space, that is, the amount of data that will be placed in the storage.

4 STRUCTURE AND CONTENT OF THE DOCTORAL THESIS

Due to the high rate of failure in the recovery processes of companies' information systems in the event of a more serious interruption, the doctoral dissertation deals with disaster recovery as an integral part of business continuity, and in seven chapters, an analysis is made of the performability and reliability of the systems that have been processed and their practical implementation in real scenarios. As an introduction to the subject, in the first three chapters, an overview of the theoretical foundations on which the research is based, related to data centers, information systems and business continuity concepts in organizations whose operations are based on digital foundations, is made.

In the **first chapter** of the dissertation, in a total of five sub-chapters, a brief overview is made of what will be discussed in the rest of the paper. In doing so, a reference is made to the subject of the research, the objectives of the research are stated, a reference is made to the applied methodology, and finally a brief overview of the results and benefits of the conducted research is made. Within this chapter, an overview of previously conducted research on solutions that are related to cloud services and solutions that address the process of disaster recovery as a concept is given.

The **second chapter** deals with the foundations on which the dissertation is based, namely data centers (Data Centers-DC) and the information systems installed in them. In doing so, data centers are defined and divided according to the method of performance and their characteristics as important parameters in maintaining business continuity, and in addition, a review is made of several key aspects related to their work, with a special emphasis on the processing of information systems. Information systems will be processed by defining them as systems, defining the terms information, data, and system as basic for their functioning, in order to further review and classify them according to their use, as important data when building business continuity plans.

The first part of the **third chapter** deals with business continuity as a concept, briefly describes methods for ensuring business continuity achieved by widely applied conventional methods as an introduction to modern systems for planning sustainable business continuity, based on key parameters in the planning process, but also of the conclusions obtained after conducting the analysis of the impact of the interruption in the overall operation (Business Impact Analyses-BIA). In the second part of the chapter, a review is made to the process of recovery from outages and disasters through its description and definition, and in the sequel, the key parameters by which the recovery process is evaluated are determined and described in detail. In addition to correctly setting the basic aspects of the recovery process, the requirements of the process are considered and specified, for its proper planning and implementation.

According to the topic, in the **fourth chapter** a description of practically implemented solutions for recovery from outages and disasters is made by defining the needs for realization, the constructive elements and their characteristics applied in the solutions. In the following, two scenarios of recovery systems are considered, one of which is a hybrid solution based on a system placed in the data center on site, with a level of security set in the cloud as an endpoint, and the other system is a completely cloud-based scenario.

Within the framework of the **fifth chapter**, the key parameters are determined and defined according to which the analysis and evaluation of the proposed systems is done. In the following, their connection is made in a functional coupling through the application of System Dynamics approach for each of the considered systems separately, in order to finally make their comparative analysis. Performing "what-if" simulations to test certain policies on such a model greatly contributes to understanding how the system changes over time, and thus to determining the parameters to be targeted in recovery plans. and maintaining business continuity in the operations of the organizations. In the course of summarizing the obtained results and their graphic presentation, the performance of the components that participate in the data protection process is considered in parallel. Such summarization is aimed at obtaining a clear idea of the performance of the systems in the data center security processes.

The **sixth chapter** provides the concluding observations based on the disaster recovery systems in place and the analysis of their performance. In doing so, the key advantages and disadvantages of each of them are highlighted in the processed application scenarios. Within this chapter, certain results and conclusions from the research are summarized. In the final part, directions for further work are given, with the aim of obtaining better performance of the used solutions.

In the **seventh chapter**, an overview of the entire literature that was used in the research and writing of the paper is given. In a separate section of this chapter, the Internet sources that were used as a necessary information source for modern achievements in the areas that were covered by the research are listed.

5 BIBLIOGRAPHY

[1] A. Bokhary, "Cloud Service Reliability and Usability Measurement", Computer Science and Engineering Theses and Dissertations, Southern Methodist University, 2018

[2] A. Marcham,"Understanding Infrastructure Edge Computing: Concepts, Technologies and Considerations", First Edition, John Wiley & Sons Ltd, New Jersey, 2021.

[3] A. Mathew, C. Mai, "STUDY OF VARIOUS DATA RECOVERY AND DATA BACK UP TECHNIQUES IN CLOUD COMPUTING & THEIR COMPARISON", 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, 2018, Bangalore, India

[4] A.A. Tamimi, R. Dawood, L. Sadaqa, "Disaster Recovery Techniques in Cloud Computing", IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019, Amman, Jordan.

[5] A.Hiles, "The Definitive Handbook of Business Continuity Management Second Edition", John Wiley & Sons, Chichester, West Sussex, England, 2007

[6] A.Lenk, "Cloud Standby Deployment:A Model-Driven Deployment Method for Disaster Recovery in the Cloud", 8th International Conference on Cloud Computing, 2015, New York City

[7] A.Mishra, V.Sharma, A.Pandey, "Reliability of Cloud Computing Services", IOSR Journal of Engineering, Volume:04, Issue:01, pp.51-60, 2014

[8] A.Mishra, V.Sharma, A.Pandey, "Reliability, Security and Privacy of Data Storage in Cloud Computing", International Journal of Advanced Research in Computer Science, Volume 5, No.3, 2014

[9] B. Chakraborty, Y. Chowdhury, "Introducing Disaster Recovery with Microsoft Azure: Understanding Services and Tools for Implementing a Recovery Solution", Apress, 2020

[10] C. Bertrand, M. Keane, "Cloud-ready Data Protection with Dell EMC", ESG White Paper, The Enterprise Strategy Group, 2018

[11] C. Garnier, M. Aggar, M. Banks, J. Dietrich, B. Shatten, M. Stutz, and E. Tong-Viet, “Data center life cycle assessment guidelines,” The Green Grid, White Paper-45, 2012.

[12] C. M. Pearson and J.A. Clair, “Reframing crisis management”, Pearson,1998

[13] D.M.Vistro, A.U.Rehman,S.Mehmood,M.Idrees, A.Munawar, "A LITERATURE REVIEW ON SECURITY ISSUES IN CLOUD COMPUTING: OPPORTUNITIES AND CHALLENGES", JOURNAL OF CRITICAL REVIEWS, Volume:7,Issue:10, 2020

[14] D.Sutton, "Business Continuity in a Cyber World: Surviving Cyberattacks", Business Expert Press, New York, 2018

[15] Data Protection and Management”, Participant guide, Dell Technologies Inc, USA, 2021

[16] E.Bauer, R. Adams, "Reliability and availability of cloud computing", IEEE Press, John Wiley & Sons, New Jersey, 2012

[17] E.Dubrova, "Fault-Tolerant Design", KTH Royal Institute of Technology, Sweden, Springer Science+Business Media, 2013, New York

[18] E.J. McClusky, S. Mitra, "Fault Tolerance" in Computer Science Handbook 2ed. ed. A.B. Tucker. CRC Press, (2004)

[19] H. Geng, "Data center handbook: plan, design, build, and operations of a smart data center", John Wiley & Sons, 2021

[20] H.B.Rebah, H.B.Sta, "Disaster Recovery as a Service: A Disaster Recovery Plan in the Cloud for SMEs", Global Summit on Computer & Information Technology, 2016, Sousse, Tunisia

[21] H.C.Lucas, Information Systems Concepts for Management, McGraw-Hill, New York, 1978, p.5

[22] H.E. Miller, K.J. Engemann, "Using reliability and simulation models in business continuity planning",International Journal of Technology, Policy and Management, Vol. 5, No. 1, 2014, pp.43-56

[23] I.Gertsbakh, "Reliability Theory", Springer-Verlag Berlin Heidelberg, NewYork, 2005

[24] J.Chan, M.Reith, "Cyber Concerns With Cloud Computing", Proceedings of the 21st European Conference on Cyber Warfare and Security, Reading, Chester, United Kingdom, 2022

[25] J.Li , P.Li, R.J.Stones, G.Wang, Z.Li, X.Liu, "Reliability Equations for Cloud Storage Systems with Proactive Fault Tolerance", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 15, 2018

[26] J.Mendonça, R.Lima, E.Andradey, J.Araujoy, D.S.Kim, "Multiple-criteria Evaluation of Disaster Recovery Strategies Based on Stochastic Models", 16th International Conference on the Design of Reliable Communication Networks (DRCN), IEEE, 2020

[27] J.Mendonca, R.Lima, E.Queiroz, E.Andrade, D.S.Kim, "Evaluation of a Backup-as-a-Service Environment for Disaster Recovery", IEEE Symposium on Computers and Communications (ISCC), 2019, Barcelona, Spain

[28] J.Mendonca, R.Lima, R.Matos, J.Ferreira, E.Andrade, "Availability Analysis of a Disaster Recovery Solution Through Stochastic Models and fault enjection experiments",32nd International Conference on Advanced Information Networking and Applications, IEEE, 2018

[29] Jaroslav Menčík, "Concise Reliability for Engineers", Intechopen, 2016

[30] Joseph M. Firestone, Enterprise Information Portals and Knowledge Management, Elsevier Science, Burlington, 2003, p.4

[31] K.Sharma, K.R.Singh, "Online Data Back-up and Disaster Recovery Techniques in Cloud Computing: A Review", International Journal of Engineering and Innovative Technology (IJEIT), Volume: 2, Issue: 5, 2012

[32] L.Tomás, P.Kokkinos, V.Anagnostopoulos, O.Feder, D.Kyriazis, K.Meth, E.Varvarigos, T.Varvarigou, "Disaster Recovery Layer for Distributed OpenStack Deployments", IEEE Transactions on Cloud Computing, Volume: 8, 2017

[33] Le secours du SI dans le Cloud, Faut-il faire le grand saut du DRaaS ?, Livre blanc, Lexsi, 2014.

[34] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges", Information Sciences, Vol.305, 2015, pp. 357-383

[35] M. Whitman, H. Mattord, "Principles of Incident Response & Disaster Recovery", 3rd Edition, Cengage Learning, 2021

[36] M.A.Khoshkholghi, A.Abdullah, R.Latip, S.Subramaniam, M.Othman, "Disaster Recovery in Cloud Computing: A Survey", Computer and Information Science, Vol. 7, No. 4, 2014

[37] M.Ali, K.Bilal, S.U.Khan, B.Veeravalli, K.Li, A.Y.Zomaya, "DROPS: Division and Replication of Data in Cloud for Optimal Performance and Security", IEEE Transactions on Cloud Computing, Volume: 6, Issue: 2, 2018

[38] M.M. Alshammari, A.A. Alwan, A. Nordin, I.F. Al-Shaikhli, "Disaster Recovery in Single-Cloud and Multi-Cloud Environments: Issues and Challenges", (ICETAS), 2017, Salmabad, Bahrain

[39] M.S. Fernando, “IT Disaster Recovery System to Ensure the Business Continuity of an Organization,” National Information Technology Conference (NITC), 13-15 September,2017, Colombo, Sri Lanka

[40] M.Wallace, L.Webber, "THE DISASTER RECOVERY HANDBOOK SECOND EDITION", American Management Association, USA, 2011

[41] N.Dhanujati, A.S.Girsang, "Data Center-Disaster Recovery Center (DC-DRC) For High Availability IT Service", International Conference on Information Management and Technology (ICIMTech), 2018, Jakarta

[42] O.H.Alhazmi, Y.K.Malaiya, "Assessing Disaster Recovery Alternatives: On-site, Colocation or Cloud", 23rd International Symposium on Software Reliability Engineering Workshops, 2012, Dallas

[43] O.V.G. Swathika, K. Karthikeyan S. Padmanaban, "Smart Buildings Digitalization: Case Studies on Data Centers and Automation", CRC Press, 2022, Abingdon

[44] P. Da, P.M. Khilar,"Virtualization and fault tolerance for cloud computing", Proceedings of 2013 IEEE Conference on Information and Communication Technologies, 2013

[45] P.J.Mitrevski, I.S.Hristoski, “Behavioral-based performability modeling and evaluation of e-commerce systems”, Electronic Commerce Research and Applications (13), p.320-340, Elsevier, 2014

[46] P.A.Longley, M.F.Goodchild, Geographical Information Systems and Science, John Wiley&Sons, West Sussex, 2005, p.30

[47] R.Asciento, A.Lawrence, ”Uptime Institute global data center survey 2020”, Uptime Institute, Seattle, USA, 2020

[48] R.H. Bowman Jr., “Business Continuity Planning for Data Centers and Systems”, John Wiley & Sons, New Jersey, 2008.

[49] S. Kambhampaty, "Infrastructure Architecture Essentials for Data Center and Cloud", Independently published, 2022

[50] S. Shahzadi, G. Ubakanma, M. Iqbal, T. Dagiuklas, "Autonomous, Seamless and Resilience Carrier Cloud Brokerage Solution for Business Contingencies During Disaster Recovery", IEEE 20th International Conference on High Performance Computing and Communications, 2018, Exeter, UK.

[51] S. Snedaker, C. Rima, "Business Continuity and Disaster Recovery Planning for IT Professionals", Second Edition, Elsevier, 2014

[52] S.A.M. Kasim, I. Mohamed, "Level of Readiness in IT Disaster Recovery Plan", Cyber Resilience Conference, 2018, Putrajaya, Malaysia

[53] S.Al-Kiswany, M.Ripeanu,"A software-defined storage for workflow applications", IEEEInternational Conference on Cluster Computing, IEEE International Conference (CLUSTER), Taipei, Taiwan, p. 350–353, 2016

[54] S.Anthoniraj, Dr. S.Saraswathi, M.Anandraj, “Disaster recovery planning using fault tolerant mechanism in virtualization environment”, Fourth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom2012), 2012, Bangalore, India.

[55] S.C. Joshi and K.M. Sivalingam, "Fault tolerance mechanisms for virtual data center architectures", Springer Science+Business Media, 2014, New York

[56] S.D.Lowe et al.,"Building a Modern Data Center Principles and Strategies of Design", ActualTech Media, Bluffton, 2016

[57] S.Wei, H.Yang, J.Song, F.Mikayilov, Z.Xu, "System Dynamics Simulation Model for Assessing Socio-Economic Impacts of Different Levels of Environmental Flow Allocation in the Weihe River Basin",European Journal of Operational Research, Volume 221, pp.248–262, 2012

[58] Scalable Data Protection for Microsoft Windows Server Software-Defined Solutions”, White Paper H17894, Dell Technologies, USA, 2019

[59] T. Wood, E. Cecchet, K. Ramakrishnan, P. Shenoy, J. Van Der Merwe, A. Venkataramani, "Disaster recovery as a cloud service: Economic benefits & deployment challenges", 2nd USENIX Workshop on Hot Topics in Cloud Computing, 2010, Boston.

[60] T.Tsubaki, R.Ishibashi, T.Kuwahara, Y.Okazaki, "Effective disaster recovery for edge computing against large-scale natural disasters", IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), 2020, Las Vegas, NV, USA

[61] U. C. Kozat,G. Liang,"Building Reliable Storage Clouds: Models,Fundamental Tradeoffs, and Solutions",Now Foundations and Trends,2015

[62] V. Kumar, R. Vidhyalakshmi, "Reliability Aspect of Cloud Computing Environment", Springer Nature Singapore Pte, Singapore, 2018

[63] V.Sivaraj,A.Kangaiammal, A.Kashyap, "Enhancing Fault Tolerance using Load Allocation Technique during virtualization in cloud computing",7th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2021

[64] W. T. Coombs, “Ongoing crisis communications: planning, managing and responding”, 5th ed, Sage, 2019

[65] W.P.Turner, J.H.Seader, K.G.Brill, "Industry standard tier classifications define site infrastructure performance", White Paper, The Uptime Institute, 2005

[66] Weill, P., Broadbent, M., “Leveraging the new infrastructure”, Harvard Business School Press, Boston, 1998

[67] Z.Liu, G.Fan, H.Yu, L.Chen, "An Approach to Modeling and Analyzing Reliability for Microservice-Oriented Cloud Applications", Hindawi Wireless Communications and Mobile Computing, Volume 2021, Article ID 5750646, 2021