Computer-Aided Disaster Recovery Planning Tools (CADRP)

Omar H. Alhazmi *ohhazmi@taibahu.edu.sa*
Department of Computer Science
Taibah University
Medina, Saudi Arabia

Abstract

Information Technology Disaster Recovery Plans (DRPs) are becoming an essential component for any organization with IT infrastructure. However, DRPs varies in performance and cost; therefore, based on requirements and resources, an organization can design their DRP. Typically, DRPs depends on data and/or system replication, data needs to be backed up frequently, and a plan to restore the system to running state within the allowed time. Hence, DRP designer must know the needed business requirements in terms of recovery time objective (RTO) and recovery point objective (RPO). Then, the appropriate technical requirements will be set. At the same time, the cost factor can play a role in choosing the appropriate DRP. The industry has a widely accepted seven-tier system of how DRP can be designed. In this work, we design and implement a software tool that can simulate the IT DPR systems and therefore help designers to design, optimize, and test their design before it is physically implemented. This tool will run a simulated system with DRP specific design and the designer can exercise with the system to show it's RTO, RPO, and cost that can significantly improve DRP design.

Keywords: Disaster Recovery, Business Continuity, RTO, RPO, Simulation.

1. INTRODUCTION

The disaster recovery plan (DRP): "is a documented process or set of procedures to recover and protect a business IT infrastructure in the event of a disaster", [1]. Therefore, it contains manual procedures usually performed by IT professionals and automated procedures performed by the IT system. Of course, minimizing manual procedures and maximizing automated procedures will reduce recovery time an important factor referred to as Recovery Time Objective (RTO). Moreover, data and system backups are typical part of any disaster recovery plan; more frequent backups will improve another factor which is Recovery Point Objective (RPO) which can also mean lost data.

The lower RPO and RTO, the better the disaster recovery plan; however, the cost also goes up. Therefore, some organizations go for DRP for critical systems and another DRP for non-business essential systems.

In this work, we introduce a Computer Aided Disaster Recovery Planning tools (CADRP) that will:

1) Design and test: Help disaster recovery specialists test and design different plans and to be able to compare them using safe simulation environment.

2) Choose, compare and optimize: Help CIOs and disaster recovery specialists choose among different alternatives which vary in RPO, RTO and Cost; moreover, and also DRP engineer with all these choices and different technologies available in the market, especially cloud services and choose the right solution wither DRaaS or Platform as a service (PaaS) or system as a service (SaaS).

3) Research and develop: to help researchers from industry and academia to conduct research affordably on CADRP platform in order to conduct quality research by using it as a virtual lab.

CADRP is a software tool that will take the environment parameters and allow the DRP specialist or administrator with some intermediate expertise in disaster recovery planning to build a virtual system and a virtual disaster recovery system with some specific given specifications and scenario. Then, CADRP will analyze system's components and calculate some statistics about the system. Moreover, CADRP will run a simulation of a system along with its disaster recovery system and at a certain point the system will assume that a disaster has struck and the original system will stop while the disaster recovery system will run and a "dynamic" analysis will be performed to calculate some metrics about the system including the critical factors of RPO and RTO. Finally, CADRP will produce a detailed report about the DRP plan.

The CADRP system will also consider systems with various tier levels 1-7 which can also make it a valuable research tool for researchers interested to work in this area, it also support cloud DR solutions.

In section 2, we will preview related work, disaster recovery tiers scheme, and disaster recovery cost analysis; next, in section 3, we preview CADRP in details; later in section 4, we will test the system; finally section 5 will discuss the conclusion and point to some future research.

2. DISASTER RECOVERY PLAN

In this section we will preview two aspects of DRP, the performance in DRP tier and the DRP cost. The choice of disaster recovery tier will have direct impact on cost. Some equations will help to test if DRP cost should be justified financially or not.

2.1 Related Work

Before the 1990s several disaster recovery solutions existed, however, they varied in their sophistication, cost and performance; therefore, by 1990 a need to categorize these solutions became necessary. Hence, a tier system of DRPs where established by IBM [5] and later over the past decades others also suggested different way of classification like Novell 4-tier system [6], Hitachi's system [7] also Webornatr [8] and Xiotech [9]. These schemes have similarities and differences; however, IBM's is the oldest and has got some acceptance in the industry. Therefore, when we designed the CADRP tools we considered IBM's to be the main reference. In addition to that, new emerging technologies being introduced to disaster recovery challenging classical DRPs tier schemes [4], such as Disaster Recovery as a Service (DRaaS) provided by major cloud service providers will also be considered by CADRP.

In searching for a research about disaster recovery simulation tools, we have found SYMIAN by Bartolini et. Al. which is a discrete event simulator, it basically simulates an incident happening to a particular application of a system to consider corrective measures [10], we have not came across any other tool that shall serve the purpose we are aiming at.

2.2 Disaster Recovery Plans Tiers

Table 1 below shows Share/IBM scheme, it is simple and yet flexible; this can explain its popularity over other schemes. Table 1, briefly explains about each disaster recovery tiers; we can ignore tier 0 which means that there is no DRP at all. Tier 1, is simple with minimum cost and can be ideal solution cost-wise for data of small and non-critical nature, it what can also be done at the personal level when backing up mobile phones or photo albums. Tier 2, has a little more readiness for recovery, a stand by system that needs to be built back with all necessary software installations, configuration and data restoration; therefore, RTO would be ranging typically from hours to days, while RPO will heavily depending on the frequency of backup (which is manually done at this tier).

Starting from tier 3, will be having a disaster recovery with more predictable RTO and RPO, automated backups at this level allows more frequent backups this improved RPO, also RTO improves with less manual work done. So, as we go to tier 4, 5 and 6. RTO and RPO improve as more frequent backups and more automation of the DRP are done at these tiers, at the same the cost gets higher. Moreover, at tier 7, is concerned with having the system mirrored with fully operational disaster recovery site. In case of a disaster or disruption, the disaster recovery site replaces the original site automatically; this significantly reduces the impact of human factor which usually causes significant delays in the recovery process from the prospective of RTO and RPO.

Disaster recovery plans traditionally fall in one of the seven tiers on IBM's7-tiers system (see Table 1), [4]. In these 7-tiers system RPO and RTO get lower (i.e. improves) as we go up in tier number.

Tier	Technology	Description
0	No off-site data	No saved information, no recovery plan at all
1	Data backup with no hot site	Data are packed up and taken to a remote location for storage, also called PTAM; *the "Pick-up Truck Access Method."*
2	Data backup with a hot site	Same as tier 1; however, the remote site has ready infrastructure capable of restoring operation to the latest backup within hours/days
3	Electronic vaulting	Same as tier 2; however, backups are done via electronic vaulting, and high speed communication (no PTAM)
4	Point-in-time copies	Same as tier3; however, data are backed up more frequently; thus, better estimation of data loss and recovery time.
5	Transaction integrity	This application level tier ensures that original site and backup site are consistent; thus, minimizing loss to zero or near zero level.
6	Zero or little data loss	This tier requires site mirroring, two sites working in sync
7	Highly automated, business integrated solution	Same as tier 6; plus the recovery process is automated; therefore, the system will recover itself with no or minimum intervention.

TABLE 1: Share / IBM Disaster Recovery Tiers.

2.3 Disaster Recovery Plans Cost

For these DRP plans there are different kinds of cost which are:

- The initial cost (Ci) which is the cost to establish the DRP
- The Ongoing cost (Co) the operational overhead including human resources, hardware and software
- The cost of a disaster (Cd) the cost incurred by the incident
- The annual cost (CT) which is the annual cost of DRP

Here we overview some of the equations used in calculating the cost and study the feasibility of the DRP. The total annual system cost (C_T) is the sum of the: Initial cost (C_i), ongoing cost (C_o) and Cost of disaster (C_d), therefore, [4]:

$$C_T = C_i + C_o + C_d \quad (1)$$

On the other hand, the total cost caused by disasters (C_d), is affected by the cost of an incident (C_λ) and the probability of a disaster happening (P_d), [1]:

$$C_d = C_\lambda * P_d \quad (2)$$

DRP	C_i	C_o	C_d	**C_T**	C_λ	P_d	**C_d**	**$C_d \geq C_T$**
1	5k	200	1k	**6.2k**	100k	.01	**1k**	**No**
2	19k	800	200	**21.2k**	1m	.02	**20k**	**Yes**
3	100k	5k	0	**106k**	2m	.03	**60k**	**Yes**

TABLE 2: Examples of Different DRPs for Different Systems.

In order for a disaster recovery planner to decide if a certain DRP is cost effective the equation below should be true, [2]:

$$C_d \geq C_T \qquad (3)$$

To clarify this, let's take those examples shown in Table 2 above. In case 1, the system will cost 6,200$ while the disaster will cost 1,000$, therefore the chosen DRP is too expensive for the system. In case 2, the system will cost about the same as disaster. On the other hand, case 3 show that the chosen DRP will save the organization 102k-60k = 42k.

3. COMPUTER AIDED DISASTER RECOVERY PLANNING

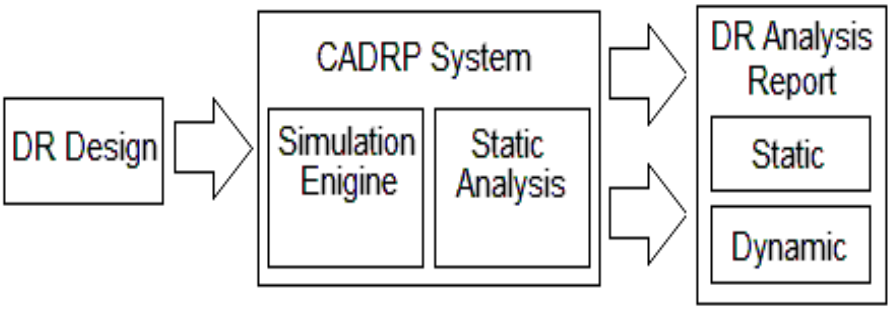

FIGURE 1: CADRP System's Outline.

Computer aided Disaster Recovery Planning (CADRP) system should accommodate disaster recovery system design ranging from the lowest tier (0) to the highest (7) on IBM tier system. CADRP should provide visual drag and drop interface. Then the system should be analyzed statically and dynamically (simulation). Figure 1 above shows an outline of the CADRP system.

3.1 CADRP Design Overview

In this part the DRP designer should design the original system and the DRP system and set the environment factors (See Figure 2): the recovery system may be absent (tier 0), or it can be a memory card, a hard disk, a tape like in lower tiers 1 to 3 or a server like in higher tiers 4 to 7; furthermore, a cloud server or storage can also be selected. Moreover, some data must be entered in order for CADRP system to analyze the DRP and generate correct reports (see Figure 3), these data is mainly about the environment to determine the weight of some factors, and this will help for the feasibility analysis and calculating RTO.

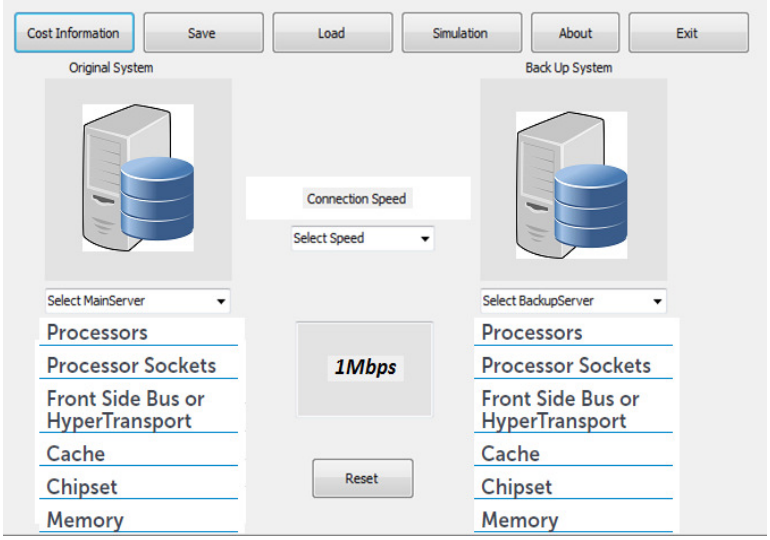

FIGURE 2: CADRP Design Screen.

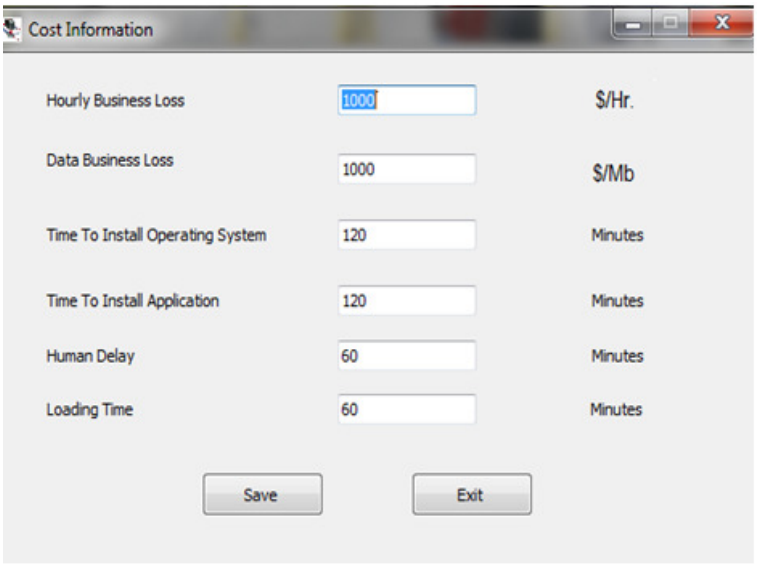

FIGURE 3: Some Data Collected by CADRP.

3.2 Simulation Engine

This module will be responsible for simulating two systems, the original system and the disaster recovery system. After the system is designed, and also the appropriate parameters entered; this module will run a hypothetical application that will run in cycles, at each cycle the original system will keep processing the current transaction depending on its CPU speed, if the transaction is fully processed then the system will process the next transaction and the old transaction will be sent to the disaster recovery back up system (it can be tape, disk, server or virtual server. So, depending on the connection speed from the original site to the backup site sometimes there is some delay; in addition to that there is a speed also for tape, disk or server to process or store the coming transactions. There is one important assumption that must be made, is to have a *sync DR* or *async DR*, each one has advantages and disadvantages, as in sync systems both systems must be in the same transaction, so the slower of the two systems will slow the other, while the async let the DR system work on its own pace, without causing the original system to wait (see Figure

4). On the other hand the sync DR preserves the integrity of the transaction as it will not move to the next one until it is processed and stored on both systems.

```
1:    int Simulate()
2:    {
3:    long i=0; //transaction processed at the original system
4:    long j=0; // transaction being transferred
5:    long k=0; // transaction processed at the DR system
6:    long cycle =0;
7:    do{
8:      cycle++;
9:      Process (Orig_system, transaction[i]);
10:     // process portion of the transaction in this cycle
11:     If processed (Orig_system,transaction[i]) {i++,j=i};
12:     // if the transaction is completed, transfer it
13:     Transfer (Speed,transaction[j]);
14:     //transfer portion of the transaction during the current cycle
15:     If transferred (transaction [j]){ j++,k=j};
16:     // if transaction is transferred process it at DR
17:     Process (DR_system, transaction [k]);
18:     If processed (DR_system,transaction [k]){ j++,k=j};
19:   // if transaction is transferred process it at DR
20:     If (disaster_triggered) disaster =1;
21:     } while (disaster == 0)
22:   RPO= k-i; // number of lost transactions
23:   Return (RPO)
24:   }
```

FIGURE 4: Basic Algorithms for Disaster Recovery Simulation Engine.

3.3 Static Analysis Module

In this part CADRP will calculate the parameters which they do not go through the simulation, including cost, ongoing cost, storage size, RTO. The static provided system can also help developers by estimating the optimal cost of a certain DRP solution.

3.4 Data Analysis Report

The generated report would take the format shown in Figure 5, so the disaster recovery planner would have results coming simulated system

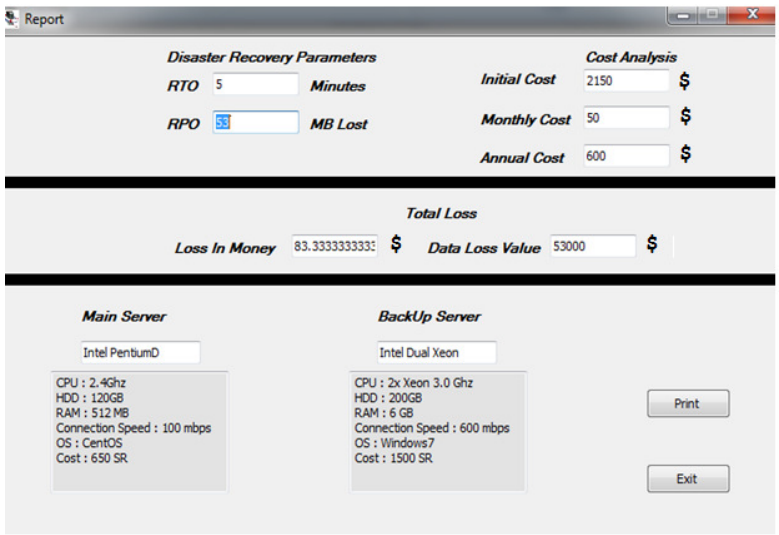

FIGURE 5: The Final Generated Report.

4. TESTING THE SYSTEM

We will evaluate RPO and RTO eight cases. By testing each case on the tool (see Figure 5) and then running the simulation, we will see how the RPO changes when giving different solution and parameters. The cases tested here are:

- Tape backup : cases 1 and 2, (tier 1, Table 1)
- Hard disk back: cases 3 and 4 (tier 3, Table 1)
- Mirrored servers; cases 5 and 6 (tier 6, Table 1)

- Cloud servers; cases 7 and 8 (tier 6, Table 1)

Here we assumed that each transaction has a fixed size of 1 Megabyte, we find that when the original system completed 10,000 transaction the backup system has just finished storing the 9970th transaction, this when the hypothetical disaster happens and therefore we have lost the last 30 transactions (about 30 Megabyte of data); so depending on criticality of the data this can be translated in loss of money.

Case	Disaster Recovery System	Network bandwidth (mbps)	Number of transactions before disaster	Lost transactions	Lost data(mb)	RTO(minutes)
1	Tape	.25	10000	30	30	240
2	Tape	.5	10000	21	21	240
3	Disk	.5	10000	24	24	120
4	Disk	1	10000	8	8	120
5	Server	10	10000	2	2	0-5
6	Server	100	10000	1	1	0-5
7	Virtual Server	0.5	10000	20	20	0-5
8	Virtual Server	100	10000	1	1	0-5

TABLE 3: Simulation Results.

In cases 5 to 8 we can notice the improvement in performance by significant reduction in RPO; however, connection speed is a main bottleneck when we tried low connection speed we have lost 20 transactions. In this simulation we can see which factor has more impact on RPO, we can see that CPU is important; however, the network bandwidth might be the main factor.

In Table 3 above, we can also see the RTO. When calculating an RTO we assumed that there is a 60 minutes operating system restore and a 60 minutes configuration and 120 minutes of system and data restoration. Therefore, the total for worst case RTO is 240 minutes; here, we ignored any impact of absence of system admins during the disaster, most of the time there is logistic delays of traffic and other factor that can add hours or days to manual system restoration. On the other hand, for automated mirrored systems recovery solution the backup server the RTO can be within few minutes.

5. CONCLUSION AND FUTURE WORK

In this work, a computer aided disaster recovery planning tool was presented to be used in practice and in research. Basically, this tool will help to design a recovery plan and also to compare different disaster recovery plans in order to find an 'optimal DRP', which will shall be effective cost-wise and performance wise and can help show the trade-off between RPO, RTO and Cost.

The tool will run a simulation to produce a report showing RPO, RTO, Systems cost and disaster cost, it will also analyze all entered data such as cost per lost megabyte and also cost of losing business time and use it to produce a helpful report.

One limitation that needs to be worked on is that RTO is calculated using some preset parameters and the analysis is straight forward of applying some formulas.

Suggested Future work would be to add more resolution and functionality to CADRP; for example, to take policy and procedures into account and to include more external environment factors. Also the tool can be extended to support more than one disaster recovery sites. In future, we plan to incorporate this tool in system courses to teach students about disaster recovery plans. Moreover, the system can also incorporate COBIT and ITIL business continuity maturity levels into CADRP's next update [11].

6. REFERENCES

[1] Abram, Bill (14 June 2012). "5 Tips to Build an Effective Disaster Recovery Plan". Small Business Computing. http://www.smallbusinesscomputing.com/News/ITManagement/5-tips-to-build-an-effective-disaster-recovery-plan.html.

[2] Lars Albrecht, Bernd Baier, Designing a bullet-proof Disaster Recovery Architecture for business-critical Applications, White paper, http://www.libelle.com/fileadmin/Public/Whitepaper/WhitePaper_Bullet-Proof_DRArchitecture.pdf.

[3] O. H. Alhazmi and Y.K. Malaiya, "Evaluating Disaster Recovery Plans Using the Cloud", Proc. Reliability and Maintainability Symposium (RAMS 2013), Orlando, January 2013, pp. 37-42.

[4] O.H. Alhazmi and Y.K. Malaiya, "Are the Classical Disaster Recovery Tiers Still Applicable Today?", Proc. 25thIEEE Int. Symposium on Software Reliability Engineering Workshop, Nov. 2014.

[5] Robert Kern, Victor Peltz, "Disaster Recovery Levels", IBM Systems Magazine, November 2003.

[6] Novell, "Consolidated Disaster Recovery", http://www.novell.com/docrep/2009/03/Consolidated_Disaster_Recovery_White_Paper_en.pdf, March 2009.

[7] Roselinda R. Schulman, Disaster Recovery Issues and Solutions, A White Paper, Hitachi Data Systems, September 2004.

[8] Montri Wiboonratr and Kitti Kosavisutte, "Optimal strategic decision for disaster recovery," Int. Journal of Management Science and Engineering Management, V ol. 4 (2009) No. 4, pp. 260-269.

[9] XiotechCorporation ,Tiered Data Protection and Recovery, , May 2006.

[10] Claudio Bartolini, Cesare Stefanelli, Mauro Tortonesi, SYMIAN: A Simulation Tool for the Optimization of the IT Incident Management Process, Lecture Notes in Computer Science Volume 5273, 2008, pp 83-94.

[11] Melita Kozina,COBIT - ITIL mapping for Business Process Continuity Management, Proceedings of the 20th Central European Conference on Information and Intelligent Systems,pp113-119,2009.