Disaster Recovery Techniques in Cloud Computing

Abdelfatah A Tamimi
Faculty of Science and IT
Al-Zaytoonah University of Jordan
Amman, Jordan
drtamimi@zuj.edu.jo

Raneem Dawood
Faculty of Science and IT
Al- Zaytoonah University of Jordan
Amman, Jordan
Adawood09@hotmail.com

Lana Sadaqa
Faculty of Science and IT
Al -Zaytoonah University of Jordan
Amman, Jordan
Sadaqa.lana15@hotmail.com

Abstract— Electronic data has been created today in large quantities requiring data recovery services organization's work may experience the various type of disasters whether it was natural or man-made, which may result in huge loss of data. The purpose of recovery technology is the possibility of retrieving information from the backup server when the main data server is lost in the event of disasters. There are some difficulties such as time and cost complexity that make it difficult to implement such techniques. When you use disaster conditions as a service, these disasters can be remedied and data recovery speeds at low cost. In this paper, we compared and discuss the various techniques to create a unique backup and recovery system. In general, all these techniques focus on three different aspects: cost control, data replication, and security issues.

Keywords: **Cloud Computing, Disaster recovery techniques, Disaster recovery as a service.**

I. Introduction

Cloud computing is becoming more common in day-to-day computing because of its ability to share globally distributed resources. Cloud computing is a set of policies and procedures that are typically supported by a physical or technical infrastructure that enables the company to recover quickly from disaster and ensure business continuity if the system crashed or any type of natural or human- made disaster occurred then there is chance of data loss and it may also cause the financial loss. Cloud computing processes are interconnected systems with shared resources There are many users who share the same storage and other computing resources. Therefore, we need a powerful mechanism to prevent other users from accessing your important and useful data. Cloud-based storage and recovery solutions allow you to back up important business files and restore them if they are compromised. Thanks to its high flexibility, Disaster Recovery (DR) allows the organization to maintain or resume critical task functions quickly after a disaster. The goal with DR is to keep the company working as close to normal as possible. Data is stored in a secure cloud environment designed to provide high availability. The service is available on demand, enabling organizations of different sizes to design DR solutions according to their needs. [6, 12]

In this paper, we have found many techniques that have their unique ways of creating backup and recovery. In general, all these technologies focus on three different aspects, such as cost control, data duplication, and security issues. Both of this technique has a complete focus on its purpose of backup and recovery.

II. Causes of data loss

A. Natural Disasters

Natural disasters are the most uncontrollable cause. Such as, fires, floods, earthquakes, even brownouts, all of them are out of our control. Fortunately, according to the survey, only 2 % of users lost data because of natural disasters. [6]

B. Mission critical application failure

Sudden damage to the application may occur when left unused for days, resulting in loss of data that may be important in some organizations. [6, 11]

C. Network failure

When the network crashes, cloud-related systems are disrupted, and cloud-based data and applications are lost because the cloud and clients are connected via the Internet. If the network fails will also suffer IP-based telephony and telecommunications.

D. Network intrusion

When viruses are invaded by applications, a disaster is created. To avoid a disaster, anti-virus applications are used and programs are placed on the disaster monitoring list .[6, 11]

E. Hacking or malicious code

You know that computer viruses can slow down your and steal credit card information. In fact, computer viruses or other malware also can spread like wildfire causing partial or complete damage to your important data. Therefore, it is essential that you should install good antivirus software and keep it updated. [6]

F. System failure:

Operating systems are affected if the organization's infrastructure fails, resulting in the failure of the organization-wide systems. [6, 11]

G. Human Errors:

Believe or not, human error is also one of the most common causes of data loss. Normally, the main reason for the occurrence of a disaster is human, 60% of the data centers are failed. There are two kinds of human errors causing data loss, one is clicking Delete or Format button to erase something we don't mean to, and another one is causing physical damages because of dropping or failing our storage device by accident. [6]

978-1-5386-7942-5/19/$31.00 ©2019 IEEE

III. Disaster Recovery Requirements

For good DR services according to the cost of system downtime or data loss, while others are directly tied to application performance and correctness., these matrices are used: Recovery Time Objective (RTO), Recovery Point Objective (RPO), performance, Consistency and geographic separation [4].

Recovery Point Objective (RPO): Calculates the maximum amount of time it takes to lose data in the event of a disaster. It can be time to lose data between seconds to hours or even days. The goal of a recovery point is to calculate the amount of data lost in the event of a disaster..[11]

The Recovery Time Objective (RTO): Recovery time objective: Is to calculate the amount of downtime and return to it in the event of a disaster. It may be minutes, hours, and days. It may determine the actual failure and settings of the backup servers.[4 ,8] As shown in "Figure 1", RPO designates the variable amount of data that will be lost or will have to be reentered during network downtime. RTO designates the amount of "real time" that can pass before the disruption begins to seriously and unacceptably impede the flow of normal business operations.

Performance: Is to perform the system without failure and give correct results using the application's simultaneous replication to the backup location for full performance without errors so that the application is ready to use. [4]

Consistency: the application which is taken a backup when the disaster occurred should be replicated on the same site after clearance of disaster at the consistent state. DR mechanism is useful to take a backup when a disaster occurs [4]

Geographic Separation: Sites should be separated from the main data locations geographically to avoid the impact of disasters that may occur in the main sites on the data and this will lead to more network response time. The greater the geographical distance between the main site and the backup location, the greater the response time and data recovery. Delay in limited data transfer back and forth due to the speed of light, data replication is only possible when the backup location is 10 seconds from the base Asynchronous technologies can improve Performance over longer distances but can lead to greater data loss during the disaster. [4]

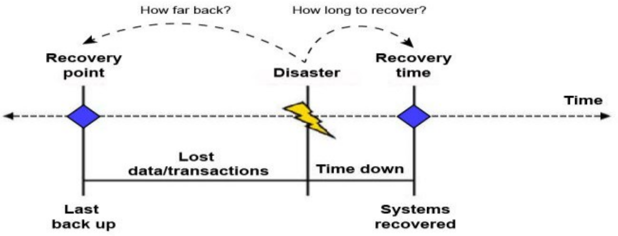


Figure 1: Disaster recovery requirements

IV. Disaster Recovery Challenges

1. Dependency

Customers do not have control over the system and their data. This is one of the disadvantages of cloud services. Backup of data is executed by and at the service provider. This problem makes reliance on CSPs for clients such as enterprises and when data is lost will be a concern for customers. [6, 16]

2. Cost

One of the key factors for selecting the cloud as a DR Service is its low price. So, cloud service providers always find cheaper ways to provide recovery mechanisms by reducing different cost types. The annual cost of DR systems can be divided into three categories: Start-up and implementation costs, which are amortized over a period of years Continuous cost: storage cost and processing cost Ongoing operating costs of data transfer and Cost of potential disasters: the cost of disaster recovery and cost of non-refundable disaster have significant costs. [16, 6]

3. Security

Is to protect data from disasters that may occur to the system. Man-made disaster may be a data breach via the Internet One of DR's goals is to protect or restore this data in case of disaster. [6]

4. Replication Latency

Backup replication techniques are divided into two synchronous and asynchronous types on which the DR mechanisms depend. Synchronous replication, RPO and RTO are very good, but expensive, and can also affect system performance due to large overhead. These problems appear in multi-layer web applications Due to increased time to go back and return between the primary site and the reserve. Asynchronous versions are less expensive than concurrent versions, but the quality of DR Service is low, so replication is an undeniable challenge in cloud disaster solutions. [8].

5. Reliability

Cloud computing design is optimized to be suitable for business continuity and disaster recovery (BCDR). We use multiple locations to store frequent data. [14]

6. Failure Detection

The system downtime can depend on the time of failure detection, so it is necessary to detect the failure quickly and report it to correct it faster. [6.16]

7. Data Storage

With the increased amount of data required for storage companies need more storage to store a large amount of data on the cloud. To ensure the security of the data and to perform the purpose of the distribution of computing centrally. [6]

V. Cloud Computing Legal issues

Cloud computing using a hybrid, community or public cloud model" The new dynamic legal environment affects the relationship between the organization and its information from the public and the preservation of laws., Including protecting the rights of customers who have become cloud service providers are a major challenge to cloud computing.

In understanding how laws are applied how information is managed these laws may relate to data transfer or storage location, as well as the extent to which such data is protected by confidentiality. Therefore, you need to consider legal issues, especially about which data has been collected, stored and processed. There are government, national or international laws that you need to keep in mind to ensure your legal obligation. The following main law Cloud computing challenges:[17]

1. **Liability for illegal data**: In many jurisdictions, cloud providers can be held liable for the illegal data they may be hosting, special protection is focused on storage, and does not take into account processing activities
2. **Compliance issues:** Infrastructure as a Service (IaaS): Data retention obligations, Tax related storage requirements, Labor law related storage requirements. Software as a Service (SaaS): electronic invoicing legislation, ecommerce legislation, electronic signature legislation
3. **Contracting issues:** Cloud computing services offer low barrier to entry and easy scaling possibilities, many publicly available clouds computing, Cloud computing contracts resemble typical software licenses, although potential risk is much higher
4. **Applicable law:** Define the rules, policies, and laws applicable to cloud computing. It can be addressed by the geographical location of stakeholders, and the rights and obligations of each stakeholder shall be determined by regulations in the country concerned.

VI. Disaster Recovery techniques

In our literature survey, we found many techniques that are having their unique ways to create backup and recovery. Broadly speaking, all those techniques focus on three different aspects, such as cost control, data duplication and security issues. Each of the technique has the complete focus on their aim of backup and recovery.

1. Parity Cloud Service

The privacy protection is a crucial issue for providing a personal data recovery service, a plain data backup-based recovery service is not adequate for public service. Users are not expected to upload their critical data to the internet backup server until they can fully trust the service provider in terms of privacy protection. a novel privacy-protected personal data recovery service framework developed, Parity Cloud Service (PCS). There are four considerations for designing personal data recovery service. 1-Reliability .2-Economical efficiency.3-Convenience. 4-Privacy protection. The PCS is extremely simple, can completely relieve users of their concern about privacy protection, easy to use, requires a reasonable server-side cost, and can recover user data with sufficiently high probability. Figure 2: shows the conceptual architecture for PCS. [1, 9, 14]

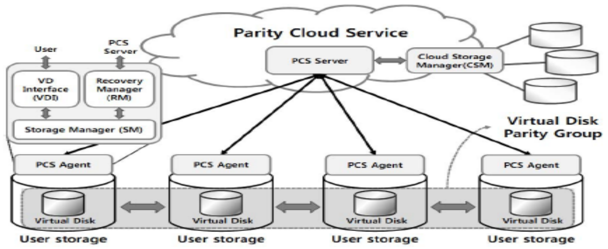
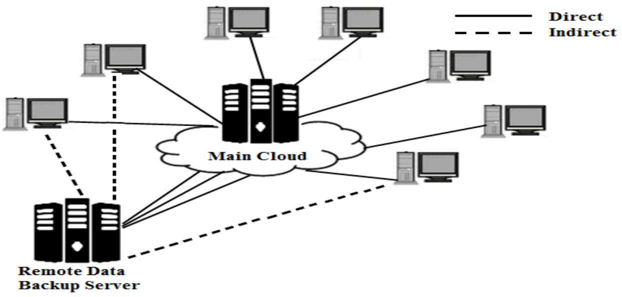


Figure 2: PCS architecture [1]

2. Seed Block Algorithm (SBA):

Is an algorithm used in the proposed system to ensure a secure backup of the data on the cloud and the remote server, this method is based on the concept of an exclusive-OR process (XOR) for digital computing? Consists of three main parts 1. Cloud Server Master 2. Cloud Clients and 3. The Remote Server. The mechanism of this work is by connecting each unique client with a client and when registering a new customer, the customer ID gets a unique random XOR number to retrieve the lost data in the event of disasters in the main cloud through the Xoring data with a particular Seed block for that particular client for data acquisition. The advantages of this technology are that it is able to recover data files with high accuracy and maintain its integrity. But it is inefficient because the storage space is wasted due to the same storage space used in the cloud and remote server [2,7,9] as shown in Fig. 3 clients can access the files from the remote storage repository if the data is not found in a central repository

Figure. 3 Remote Data Backup Server [13, 16]



2.1 EXPERIMENTATION AND RESULT ANALYSIS

It is observed that memory requirement is less in a remote server as compared to the main cloud's server because files will be stored on remote server after compression this will lead to reduced memory requirement for the backup server. [2, 9, 12]

3. Multi-tier Web Application

Multiple web applications consist of the ends of the web interface that are linked to a database and are intended to analyze the cost of DR by calculating replication costs and setting failure we operate the ROBIS standard on the web. We calculate the costs using the registered resources from RUBiS (an e-commerce web application that can be run using multiple Tomcat servers and MySQL database) with 300 clients. As shown in Figure 4, in the normal operation of the primary data center, the cloud manages to restore the ability to work after disasters using two types of resources: backup mode resources to obtain backup before an active disaster; failure mode resources that will be activated only after the occurrence Disaster. [8]

Uptime Cost: compared by using public cloud and Colocation. the yearly cost of the cloud DR service comes to only $1,562, compared to $10,373 with the colocation provider [4].

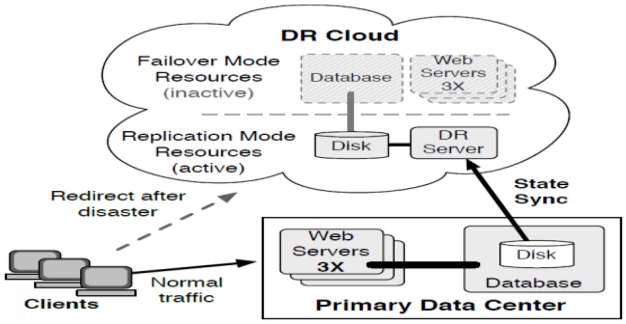


Figure. 4: Overviews of RUBiS system architecture [4]

4. Data Warehouse

The data warehouse contains the data that applications generate and are added to the repository at regular intervals. Reports are created based on the incoming and current data set. Inventory costs are estimated according to the size of the data.
Uptime Cost: by comparing the failover mode cost it is cheapest to use a colocation center as the primary site of the data warehouse ($5,853 per year in the cloud versus $3,202 per year in a colocation center. [4]

5. Carrier Cloud Brokerage

Multi-Cloud Broker Orchestration and Cloud Carrier Architectures can play an increasing role in delivering Backup as a Service. Ensuring that clients can not only recover their cloud infrastructure and services on the fly when disaster strikes; but also, be assured of reliable access to backup data resources as vital services are restored. During disaster recovery situations every second count and restoring cloud resources can become a matter of survival for a business. This requires an autonomous seamless and resilience carrier cloud broker-age solution as shown in Figure 5, where multiple clouds are connected to a single cloud brokerage solution which can provide resources to accommodate different IaaS requests. Cloud service providers can provide scalable resources to accommodate the requirements within minutes. The entire process is required to be initiated dynamically. To achieve this, we have proposed a real-time cloud brokerage that provides seamless, autonomous (self-service and self-managed) [3].

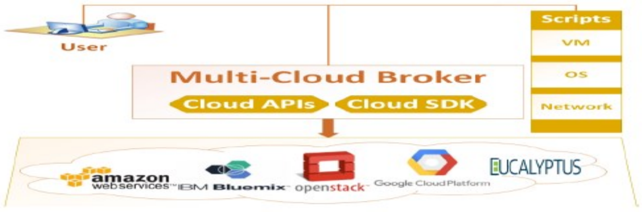


Figure 5: Multi-Cloud Broker Solution [3]

6. High -Security Distribution and Rake Technology (HS-DRT):

The HS-DRT is a backup concept that uses a very wide distributed data transfer mechanism and high-speed encryption technology. Its advantages include: 1) does not require the use of expensive leased lines. 2) Spatial scrambling and random distribution techniques are used to encrypt important data files. 3) With the increase in the number of users, this technology is safe, and the encryption speed is high. This model can be used for transferees such as laptops and smart phones. This technology cannot be considered ideal for backup and recovery in cloud computing because of the increased cost of data recovery and increased redundancy. When the number of duplicate copies of file data increases from the corresponding processor, the performance will be reduced. [5, 14, 7, 8]

7. Efficient Routing Grounded on Taxonomy (ERGOT)

Is a semantic system for discovery of service in distributed infrastructure in cloud computing It is not a backup technology, but it provides an effective retrieval of data that relies on semantic similarity between service descriptions and service requests. ERGOT components: 1) DHT protocol (distributed fragmentation table), which we use to declare the description of the annotated semantic service using the concepts of ontology. 2) SON (Semantic Overlay Network) enables the compilation of an analog that contains a service description that is linguistically similar. SON is increasingly being created as a product to advertise services via DHT. 3) The semantic similarity between service descriptions. [5, 9]

1. Linux box:

It's cost-effective, disaster-protected, and easy to migrate data from another cloud provider. It's useful for small and medium businesses (SMBs). This solution eliminates consumer support on the ISP and its associated backup cost. A Linux box that will synchronize data at the cluster / file level from the cloud service provider to the consumer. It includes an application on a Linux box that backs up the cloud on local drives. [5, 9]

2. Cold backup:

The cold backup recovery process runs when service failure is detected. You must purchase everything required to restore the service to your users and deliver it to the site before the recovery process begins. There may be a delay in hours or days when devices are ejected from storage or redirected from test and development systems. Cold backup sites are the least costly. [11, 15]

3. Warm backup:

In this technology, the site is provided with devices that represent a replica of your data center. , The last backups of the off-site storage facility are delivered to restore the service. It takes about 8 hours to 24 hours (depending on complexity, location, and data size) to recover data after a disaster. The hot backup location may keep the state up-to-date with either synchronous or asynchronous replication systems depending on RPO is necessary. [11, 15]

Hot backup:

A hot backup site typically provides a set of mirrored stand-by servers that are always available to run the application once a disaster occurs. a set of mirrored stand-by servers that are always available to run the application once a disaster occurs. There will also be a requirement for a resilient network connection into the Hot Site. The environment is running concurrently with your main data center. [11, 15]

Table2. Comparison of various techniques of recovery properties

Technique	Advantage	Disadvantage
HS-DRT	Used for movable clients	Costly, increase redundancy
PCS	Reliable, Privacy, Low cost	High complexity
LINUX BOX	Simple, Low cost	High bandwidth, Complete server backup at a time
ERGOT	Perform exact-match retrieval, Privacy	Cost increases as data increases, Increased complexity
Multitier Web Application	minimizing costs	-
Data Warehouse	minimizing costs	-
Hot Backup Service	Available immediately	very expensive
Warm Backup Service	Low cost	time Wasting
Cold Backup Service	Available immediately	Cost, Complexity management.
Seed Block Algorithm	Simple to implement, minimum time to the recovery process.	inefficient
Carrier Cloud Brokerage	Significantly provides resilient and robust cloud performance. reduce the capital	-

Table-3. Comparison of cloud-based DR technologies in terms of different

Techniques	Security techniques	Cost	Redundancy	Time to recovery	Complexity
High-Security Distribution and Rake Technology	High	High	High	-	-
Parity Cloud Service	High	Low	High	Low	High
LINUX BOX	Low	Low	High	Low	Low
Efficient Routing Grounded on Taxonomy	High	-		High	High
Multi-tier Web Application	-	Low	-	-	-
Data Warehouse	-	Low	-	-	-
Hot physical Backup Service	High	High	Low	Low (minutes)	High
Hot virtual Backup Service	High	Lower than physical	-	Low (1 hour)	
Worm Backup Service		low	-	Medium (1-24) hours	-
Cold Backup Service	-	High	-	High (more than 24 hours)	-
Seed Block Algorithm	High		High	Low	-
Carrier Cloud Brokerage	High	Low	-	Low	-

After considering all the above techniques in table 1 and 2, some briefly summarized tables regarding the advantages, disadvantages and different properties of these techniques have been made Show in Table2 and Table 3. Among all the technologies, PCS is one of the most sophisticated technologies to maintain the privacy of every supplier and to reduce the cost of infrastructure, but its disadvantages are that it is unable to control the complexities. HSDRT was found for users of laptops and smart phones but failed in terms of cost reduction in the implementation of recovery because it is unable to control the duplication of data. ERGOT relies on semantic analysis and is unable to focus on time and complexity of execution. Linux Box is a highly cost-effective data recovery technology. All these techniques have been tested to keep them as low as possible. Some technologies increase in cost as data increases. For example, a hot and cold backup strategy that addresses backup and recovery based on detection of operating failure. SBA is to take a minimum time for the recovery process. We have compared the costs of running DR services using public cloud or privately-owned resources and shown cost reductions of up to 85% by taking advantage of cloud resources in Multi-Tier Web Application and Data Warehouse.in Carrier Cloud Brokerage This proposed solution will not only reduce the capital expenditure but also provides a reliable and efficient way to access the data during disaster.

VI. CONCLUSION

This paper discuses and differentiates between various type of some disaster recovery techniques in cloud computing and highlights some of the common causes of data loss. It also addresses some of the difficulties and limitation of the used techniques knowing that disaster recovery is becoming one of the essential aspects in an organization.

VII. References

[1] C.-W. Song, S. Park, D.-W. Kim, and S. Kang, “Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service,” 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications, 2011.

[2] N. Vedashree, P. Kumar, G. Anilkumar,” Data Recovery in Cloud Environment Using Seed Block Algorithm”, (IJCSIT) International Journal of Computer Science and Information Technologies, 2015.

[3] S. Shahzadi, G.Ubakanmay, M. Iqbalz, T, Dagiuklasx “Autonomous, Seamless and Resilience Carrier Cloud Brokerage Solution for Business Contingencies during Disaster Recovery”

[4] T. Wood, E.Cecchet, K.Ramakrishnany, P. Shenoy, J. Merwey, and A.Venkataramani,” Disaster Recovery as a Cloud Service: Economic Benefits & Deployment Challenges”, University of Massachusetts Amherst.

[5] 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), 2012

[6] A. A. Gharat, D. E. Mhamunkar.” Disaster Recovery in Cloud Computing”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2015.

[7] D. K. Bagave, U. R. Naik, P. P. Shedge, L, S. Naik,” Survey on Cloud Data Recovery and Security”, International Journal of Engineering Science and Computing(IJESC), 2017.

[8] M. A. Khoshkholghi, A. Abdullah, R. Latip, S. Subramaniam, and M. Othman, “Disaster Recovery in Cloud Computing: A Survey,” Computer and Information Science, vol. 7, no. 4, p. 39, Mar. 2014.

[9] S, P. Badhel, V.Chole,” An Efficient and Secure Remote Data Back-up Technique for Cloud Computing” International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 4, 2015

[10] M. M. Alshammari, A. A. Alwan, A. Nordin, and I. F. Al-Shaikhli, “Disaster recovery in single-cloud and multi-cloud environments: Issues and challenges,” 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2017.

[11]A.Srinivas, Y.S. Ramayya, B.Venkatesh.” A Study on Cloud Computing Disaster Recovery” International Journal of Innovative Research in Computer and Communication Engineering IJIRCCE, 2013

[12] P.S. Challagidad, A. S. Dalawai1, M.N. Birje,” Efficient and Reliable Data Recovery Technique in Cloud Computing” Internet of Things and Cloud Computing. Special Issue: Advances in Cloud and Internet of Things. Vol. 5, No. 5-1, 2017

[13] Y. Gite, A Pawar, S. Ghumbre.” Efficient Data Backup Technique for Cloud Storage”, International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)Vol 5, Issue 3, 2018

[14] Y. Sambrani, Rajashekarappa.” Efficient Data Backup Mechanism for Cloud Computing”, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) ,Vol. 5, Issue 7, 2016.

[15] cloudhpt: Traditional Disaster Recovery versus Cloud based DR,2014.

[16] R. V. Gandhi, M Seshaiah, A. Srinivas, C. ReddiNeelima,” Data Back-Up and Recovery Techniques for Cloud Server Using Seed Block Algorithm”, Gandhi et al Int. Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 5, Issue 2(Part 3), 2015.

[17] Hourani,H.& Abdallah,M.(2018). Cloud Computing: Legal and Security Issues. 2018 IEEE 8th International Conference on Computer Science and Information Technology (CSIT). doi:10.1109/CSIT.2018.8486161