Research Article

International Journal of Advanced Computer Research, Vol 9(40)
ISSN (Print): 2249-7277 ISSN (Online): 2277-7970
http://dx.doi.org/10.19101/IJACR.SRS16002

A technical study of remote backup center performance using public virtual private network

Hooki Lee[1], Sungtaek Lee[1], JongHyuk Seong[2], HoGun Rou[1] and GwangYong Gim[3*]
Research Scholar, Soongsil University, Korea[1]
Research Scholar, Kyonggi University, Korea[2]
Professor, Department of Business Administration, Soongsil University, Korea[3]

Received: 31-June-2018; Revised: 30-August-2018; Accepted: 15-November-2018

Abstract
Due to the rapid development of information technology worldwide, the spread of information technology has made the use of information systems prominent throughout the industry. In particular, the public sector claiming electronic government, provided convenient services, prompt administrative processing, and utilizing various public data. As a result, the beneficiary, the citizens, can use more services conveniently and easily, thus the amount of information data has increased sharply. As the reliance on such information systems continues to increase, the demand for stable and continuous service of information resources is also increasing, and due to unexpected natural disasters or cyber-attacks, the operation of various information systems is stopped being unable to function properly. It should also be prepared at all times for damage situations where critical data may be destroyed or lost and recovery may not be possible. Therefore, a safety management system that prevents the loss of credibility and financial loss caused by various disasters and system damage is becoming a necessity, not an option. The safest and most reliable method of technical and administrative safeguards for establishing safety management systems to ensure continuity of information systems against disasters and disasters is to backup the data to remote locations. Backing up data to a remote location in preparation for disruption and restoration of the main information system requires large budgets, such as technical and administrative considerations, network dedicated lines, and construction of a physical remote center. Therefore, for small and medium enterprises, it is a burdensome reality to build as such. Thus, this study is going to propose the composition of the remote backup center model using a public virtual private network and de-duplication backup technology for low-cost and high efficiency remote backup center model that can utilize small and medium sized agencies.

Keywords
Data backup, Remote backup, Disaster recovery, Backup center, Data center.

1.Introduction
According to the international data corporation (IDC), it is expected that an agency investigating trends in the international ICT market, the use of a variety of smart devices and the internet have increased dramatically, and the volume of data, such as digital contents have surged and the amount of stored data is expected to exceed 40,000 ex bytes, which is expected to surge 50 times compared to 2010, and accordingly, it will require more than 10 times as many servers (Korea IT promotion agency, 2012).

In the present situation, if the information system that is closely related to the lives of the public suddenly stops or the important data and digital contents used in various environments are lost in all the public and industry. There will be great damage and loss. The safety management system of information system in preparation for such unexpected disasters and disasters has become a necessity and not an option in modern society. Disaster was called natural disasters such as a typhoon or earthquake in the past. However, in modern society, the result of large-scale anthropogenic accident exceeds the damage of natural disaster, and now it is a concept including artificial disasters and natural disasters [1]. Among them, server disruption of information system, communication disconnection, power failure, and hacking are classified as technical disasters [2]. The

*Author for correspondence

information systems of more than 350 agencies and companies residing in the world trade center at the time of the 9/11 terrorist attacks in 2001 were mostly destroyed and lost together with the collapse of the buildings. However, some companies, including Merrill Lynch and Morgan Stanley, were equipped with a disaster recovery system in response to the crisis management response system, being able to receive the confidence of customers by promptly normalizing works after a disaster. On the other hand, about 150 agencies and companies, which were not fully equipped with the disaster recovery system based on crisis response, went bankrupt consecutively after the incident and their damages amounted to about $ 120 billion [3]. Cyber-attacks that cause national massive damage are occurring every year. In Korea, where the ICT industry developed, the 1.25 Internet crisis in which the national internet network was paralyzed in January 2003 was caused by the Slammer Worm from the USA and Australia due to communication network traffic increase, and the server was infected and the internet was interrupted due to incapability of restoration, and the total damage due to the data destruction that was not backed up amounted to 102.5 ~ 167.5 billion won. Also, in July 2009, the 7.7 distributed denial-of-service (DDoS) attack, which broke the homepage of major financial and government agencies, caused damage of about 36.3 billion won to 54.4 billion won. On March 20, 2013, 3.20 cyber crisis caused by large-scale damage in which attacks such as the disconnection and alteration of websites of media companies, financial corporations and public agencies occurred in a total of KRW 867.2 billion [4]. Such an Internet infiltration accident not only causes disruption such as disconnection or delay of the network, but also causes damage of the software such as the server, the disk or the like, the software, and the destruction of the data containing important information. This damage occurs because the management network and the backup network are not separately constructed, or there is a contact point in some network separated areas. In this way, companies or agencies that cannot systematically organize remote data backup and backup dissemination operations can be identified by recent cyber security incidents that cannot be recovered in case of damage.

According to an infrastructure composition, research project for data center industry prosperity, about 47% of IDCs receive backups at separate centers, while the remaining 53% operates the same server, region backup, which can cause big damage such as business interruption, financial loss, etc. in case of a disaster or a hacking accident caused by a cyber-attack [5].

In order to secure the safety and business continuity of the information system in preparation for such disasters and disasters, a disaster recovery environment that can recover to normal at all times is indispensable, which is possible to utilize the backed up data even in the event of damages of the main distribution facility through the construction of the remote backup center at a distance from the main computer facility. The recovery of infrastructures damaged by various disasters and repairing and recovering various computer equipment is also time-consuming. However, since it may take several days or several months until the lost data is restored, there is only one way to build a remote place backup system as a solution [6, 7]. In addition, most security programs and anti-virus programs that can remove the new cyber threat ransomware are provided, but method to solve the lost damage by completely decrypting the files encrypted by ransomware is the most effective measure to previously backed up files [8]. However, in the rapidly developing IT environment, the recommendation of the remote backup center only suggests a large guideline, and it is necessary for many agencies and companies to utilize a large amount of the budget such as manpower, communication security, which is accompanied by realistic constraints. Therefore, this study proposes an on-line remote backup center for low-cost and high-efficiency data dissemination concept considering realistic methods.

2.Related research

2.1Types of remote backup center

Remote backup centers are subdivided into technical and business types according to their configuration as an important part of business continuity planning in case of accidents such as disasters. There are four types of technology configuration: mirror site, hot site, warm site, and cold site [9]. There is a large difference in coping and recovery ability for main memory centers depending on each implementation type. Therefore, it is necessary to pay close attention to the size of the remote backup center in consideration of the size of the agency, the budget requirement, the importance of the work, and the continuity requirement of the business processing [10]. First, as a type of operating in the active-active mode between the main transfer center and the remote backup center, the mirror site has demerit of the high maintenance cost of operation management

as well as investment costs upon initial construction. Second, hot site maintains the latest data through real-time data replication, mirroring while keeping the information system at the same level as the main center at the standby site at the remote site. It is a method of service by switching the information system of the remote backup center to active when it occurs. The initial investment cost and information system maintenance cost are high, and it is not possible to guarantee continuity of work when a small amount of data loss or failure occurs, but it is cheaper than a mirror site. Third, the warm site is similar to the hot site, but it has the same level of information system as the main center in the remote backup center, and only the important information resources are partially stored in the remote backup center and maintenance costs are low, but there is a disadvantage that the initial recovery level is incomplete and it takes some time to complete recovery. Fourth, the cold site secures only the remote backup center space and keeps the data only at the remote site. It is a method to recover the information system by procuring necessary information resources based on the data backed up at the time of disaster, without securing the information resource for the service or securing it at the remote place at a minimum. Construction and maintenance costs are the cheapest, but they have the longest recovery period and low reliability [1].

2.2Remote backup center configuration technology

Data de-duplication technology is the most widely used technology for data backup. It is a kind of data compression technology that detects duplicated part between each data and improves the efficiency of storage space utilization by eliminating duplicated part. The core of data de-duplication technology is to compare the data of the backup object and save the new or changed contents without storing the backed up contents when they are found, thereby improving the space efficiency of the storage. Therefore, even when a remote data backup is to be performed, it is indispensable to change the total capacity of the backup target and de-duplicate the data to transmit only a small amount of new data [11]. It is divided into source-based approach and target-based approaches, depending on where the core functions of de-duplication technology are located [12]. In other words, the source-based backup method is performed when the client performs the de-duplication function and is classified as the target-based backup when the de-duplication function is performed in the de-duplication server. Source-based de-duplication is a data de-duplication in which backup data is sharply reduced and only the block of the changed data performs the backup. Performing data de-duplication of the source has the effect of overestimating storage capacity, which can significantly reduce the amount of data transferred from the source device to the backup storage, thereby reducing the complexity of the physical infrastructure and network bandwidth [13]. The target-based de-duplication technology is suitable for storage area network (SAN)-based high-capacity database backup, and it is a special-purpose high-speed network that connects with heterogeneous data storage related servers for large network users. When de-duplication is performed on the target, only the newly added lower data is stored, so that the backup disk capacity can be optimized. Like this, the de-duplication data backup technology should be configured to select the location where the de-duplication function is performed according to the size of the agency, the network status, the data nature, and the recovery performance, and when configuring for the remote backup, you should closely consider the target of the backup and the efficiency of storage space depending on network status, and the way in which data is transmitted.

The virtual private network (VPN) required for constructing this research model is a security technology for obtaining the same usage effect as the internal private network by using the public internet network, and you can configure a line excellent and safe with much lower cost than private lines [14]. Tunneling technology, which is the most important technology for configuring VPN, is a technology that creates confidentiality and integrity of data by transmitting mutual information with creation of virtual tunnels that are not exposed to the outside on the Internet. As a VPN technology designed to provide interoperability, high quality, and encryption-based security at IPSec protocol VPN supports IPv4 and IPv6. It is used as an alternative to efficiently use remote networks as the reliability and communication security of branch offices and branches are verified.
However, there are not many technology cases and researches that build a remote backup using virtual private network. In the remote backup field, most of the utilization of virtual private network is composed of a backup center configuration for a single agency, a small amount of data storage, and a temporary network line replacement in case of a dedicated line failure. However, it is hard to find a case in which productive data are applied to the construction of an integrated remote backup center. Most guesses are

that the backup success rate may be low because the backup speed is lowered due to the use of data encryption/decryption and tunneling when the virtual private network is used and safety against the physical private line is not secured. In order to comprehensively derive the implications for constructing a low-cost, high-efficiency remote backup center by analyzing prior researches and various situations of remote backup center status, data backup technology, virtual private network technology, etc., small and medium-sized agency is severely insufficient in budget and manpower, so it is reasonable to construct a remote backup center that can minimize the online data dissemination in order to protect data against disasters, and the three detailed implications derived are as follows.

First, considerations of the commercialization type of remote backup center are that is difficult to construct a remote backup center for each single agency. So, it is reasonable to construct a remote backup center for joint use in centralized places such as head office or parent agency. Second, the consideration of the technical type of remote backup center is similar to the concept of mirror site in terms of performing online backup different from the concept of disaster recovery (DR) center. A mixture of the concept of a warm site is suitable in the aspect of configuring and cooperating remote backup center. Third, technical considerations for configuring the remote backup center are as follows; in order to solve the problems of long distance issues such as small and medium sized branch offices and the high cost of leased line usage budget, VPN. It is a realistic solution to construct a dedicated private network for remote, backup, and de-duplication technology that transmits backup data with limited line capacity using the public virtual private network to remove the same data and transmits only unique data is essential, and this study suggest a model applied to it.

3.Design of research system model

3.1Remote backup center design

This study proposes a remote backup center construction model using VPN and backup data de-duplication. A total of 14 small and medium-sized agencies use VPN to configure a backup dedicated network on the internet line that is actually used, and this study designed a technology model by dividing it into a virtual tape library (VTL) to VTL, S/W agent method considering network configuration connected to the central backup center as well as data volume, line speed owned by agencies. Since each agency has different data capacity and internet line capacity, the main considerations to be designed by applying the technology for optimal remote backup are as follows:

1) Design/construction of the remote backup system considering data capacity
2) Plan to minimize the amount of backup data transfer
3) Virtual private network configuration plan
4) Securing the backup data transfer security
5) Consideration of separation network communication, safety for back-up data collection

As shown in [*Figure 1 & 2*], an agency that holds a large amount of backup data or a large amount of network bandwidth is configured as a reproductive function after applying a source-based de-duplication technique. After installing the backup-only agent on the backup target server, first perform a backup to the VTL installed in the target agency, and then configure the VTL to VTL connection method between the remote backup centers to transfer data by the replication method. *Table 1* shows the target agency status. As shown in [*Figure 3*], an agency VPN is established in the target agency's internet line section, and a public IP is assigned to the untrusted interface, and tunneling communication is established by registering the public VPN on a site to site format. In order to construct a closed network environment for remote backup, after setting the routing of the private IP band in the untrusted section of the agency VPN, the data holding servers give the private IP band different from the existing IP band (for communication, self-backup) do. Also, the backup dedicated L2 switch is configured for the flexibility of the number of backup target servers and the same private network configuration.

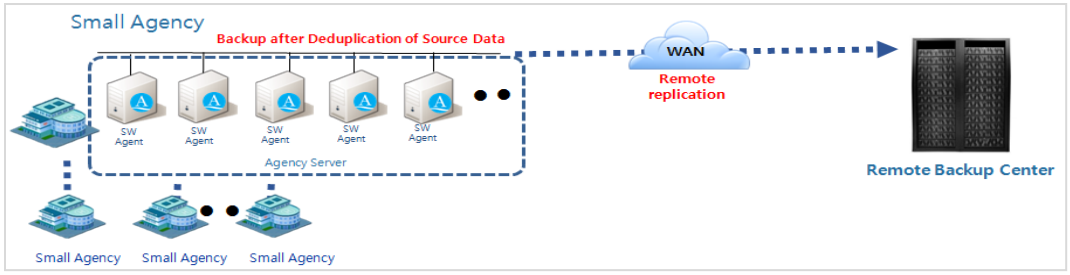

Figure 1 Schematic diagram of the de-duplication technology

Table 1 Target agency's status

No.	Agency	Data capacity (Bytes)	Internet line (MBPS)	Backup type (Daily)	Target hosts(EA)	Remarks
1	A	0.159	100	Full B/U	2	Small
2	B	0.232	100	Full B/U	11	Small
3	C	0.657	100	Full B/U	10	Small
4	D	0.371	150	Full B/U	3	Small
5	E	0.901	150	Full B/U	9	Small
6	F	0.221	100	Full B/U	2	Small
7	G	0.293	50	Full B/U	7	Small
8	H	0.811	100	Full B/U	3	Small
9	I	0.051	100	Full B/U	10	Small
10	J	0.302	100	Full B/U	2	Small
11	K	1.161	150	Synthetic Full B/U	11	Large
12	L	1.576	150	Synthetic Full B/U	8	Large
13	M	1.657	300	Synthetic Full B/U	10	Large
14	N	1.569	200	Synthetic Full B/U	13	Large

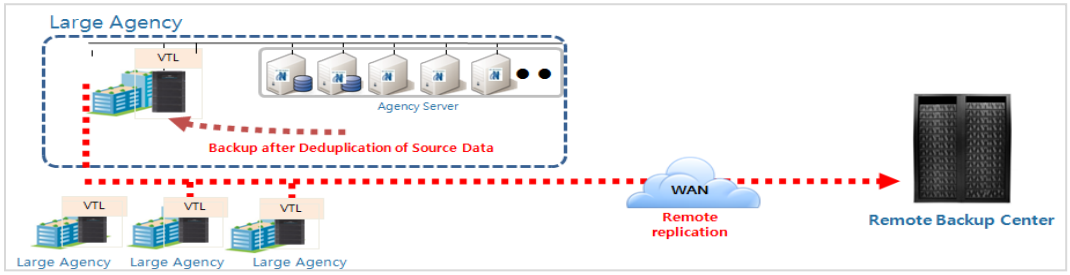

Figure 2 Construction diagram of large-sized agency de-duplication technology

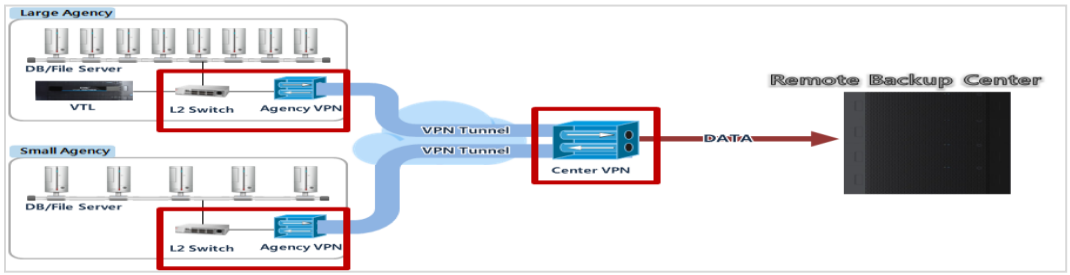

Figure 3 Configuration of public VPN

The operational procedures for backing up and restoring the research model consist of large and small agencies, which are divided into the state of the agency and the backup technology. The operational procedures for backing up and restoring the research model consist of large and small agencies, which are divided into the state of the agency and the backup technology. The backup and recovery procedure flow charts are shown in [*Table 2*].

The advantage of the VTL to VTL backup method is that it is recovered from the VTL installed in the target agency in the recovery task, so that it is processed in a short time. Remote backup of remote backup center and target agency server is performed by communication of dedicated S/W installed in the backup target server and remote backup center because a small amount of data of small agency is held. When an abnormality occurs in a small agency, the speed is slow as it is restored directly from the remote backup center to the target agency backup server against a large agency, but it reduced the condition for bandwidth compared with the standard backup method as it backs up and restores by asynchronously transferring.

This study sets the network traffic transfer threshold as 10Mbps between backup server and remote backup center of each agency to verify the empirical data of this study.

Table 2 Backup and recovery procedure

Size	Procedure	Progress sequence
Large	Backup	① Backup request ② Perform backup ③ Confirm backup information ④ Change check ⑤ Compression, CRC check ⑥ de-duplication backup ⑦ Backup completed
	Recovery	① Recovery request ② Perform recovery ③ Recovery check ④ Recovery completed ⑤ Recovery from the center
Small	Backup	① Backup request ② Perform backup ③ Confirm backup information ④ Change check ⑤ Compression, CRC check ⑥ de-duplication backup ⑦ Backup completed
	Recovery	① Recovery request ② Perform recovery ③ Recovery check ④ Recovery completed ⑤ Compression, CRC check ⑥ de-duplication backup ⑦ Backup completed

4.Research system measurement and verification

4.1Research system measurement and verification configuration

The most important characteristic of the system constituting the remote backup center constructed through this study is that it uses a dedicated virtual private network instead of a closed dedicated line to reduce the cost of construction and operation, that is, the cost of the dedicated line. It can be said that, however, if it does not meet or exceed the general performance requirements of the backup system, it cannot be expected to play a fundamental role as a remote backup system beyond economic logic. Nevertheless, the basic performance of the remote backup system does not define in detail and quantitatively. Generally, at the level of common sense, when the problem of data integrity or failure occurs, the level of quality in the aspect of hardware such as equipment or backbone line is considered as the performance of the backup system in fact.

4.2Study system performance verification items and measurement values

4.2.1 Establishment of research system performance verification items

Applying the generalized Delphi technique, preliminary investigation and total Delphi survey of 3 times are designed. The results are summarized as follows: first, the average and median are high. Second, the convergence is less than 0.5. Third, the consensus is high. Fourth, it was judged as an item in which the opinions of backup expert group are collected based on 4 standards of the item with low performance in coefficient of variation. Finally, expert opinions were gathered as a performance item that the system constituting the remote backup center should have with three items: transfer speed, backup success rate, and de-duplication rate.

Finally, the average value of each performance item considered by the expert group is quantified as the target value of the performance item to be provided by the general remote backup system, and the performance of the remote backup system model presented in this study is measured and the result is shown in the Delphi survey result [*Table 3*] and the target value was compared statistically.

Delphi survey results show that the performance of the selected remote backup system is measured by implementing the backup dedicated network configuration, backup method and backup procedure between the central remote backup center and 14 target agencies designed in this study. The performance measure of this research system is largely three kinds of system, which is designed by using a public VPN. Therefore, it is a de-duplication rate using backup rate among remote sites, storage capacity trend of source data and target data, and source de-duplication technology. In order to collectively judge the three factors and to derive the results, this study measured actual operating data for three months.

There are 8 variables in DB data stored for 63 days, and it is composed of data of 6,399 cases in total. There are 8 variables in total, including backup date, agency name, host name, time, capacity, target storage capacity, de-duplication rate, and backup success rate. *Table 4* shows the comparison items calculated from the variables measured by the research system in response to the performance items of the remote backup system such as the transfer speed, the backup success rate, and the de-duplication rate.

This study measured the amount of backup data, de-duplicated capacity, backup success rate, and backup time data at each of 14 remote sites. *Table 5* shows the backup performance of 4 large VTL-to-VTL backup system and the 10 small S/W-type backup systems, which is measured by actual operation materials.

Table 3 Delphi survey result performance item and target value analysis result

Performance requirements	Average	Standard deviation	Median number	Minimum value	Maximum value	Range
Transfer speed	94.15	3.94	95.00	90.00	100.00	10.00
Backup success rate	98.95	1.70	100.00	95.00	100.00	5.00
De-duplication rate	95.35	3.70	95.00	90.00	100.00	10.00

Table 4 Explanation of measurement parameters of research system performance

Variable Name	Variable Description	Data Type
Backup date	Date of backup	Yyyy-mm-dd
Agency name	agency to back up	T01 ~ T14
Host	Host Backup Server	Ex) T01_book_fs
Lead time	Time to perform backup	00 Min
Volume	Original capacity	GB (gigabytes)
Target storage capacity	Remote Backup Center Storage Capacity	MB (megabytes)
de-duplication rate	Date of backup performed De-duplicated percentage	% (Removal rate)
Backup probability	Success or failure of backup	Success / Failure Ratio (%)

4.2.2 Transfer rate

The transfer rate converts the target storage capacity variable in megabytes into megabits and converts the backup time in minutes into the transfer rate (megabits per second). The transfer rate in seconds calculated like this exists between minimum 0Mbps and maximum 10Mbps, and it is possible to convert it in percentage (%) unit by multiplying 100 to this. Therefore, the transfer rate for one host can be defined as Equation 1.

$$\text{Transfer rate} = \frac{8\,x_{ij}}{60\,t_{ij}} \times 100 \approx \frac{1.3\,x_{ij}}{t_{ij}} \qquad (1)$$

Target storage capacity = x
Processing time = t
Host = i = 1, 2, ..., 101
Backup date = j = 1, 2, ..., 63

According to Equation 1, the average backup data transfer rate of the research system can be calculated as shown in Equation 2.

$$\text{Average transfer rate} = \frac{\sum_{i=1}^{101}\sum_{j=1}^{63}\left(\frac{8\,x_{ij}}{60\,t_{ij}}\times 100\right)}{n_{ij}} \approx \frac{\sum_{i=1}^{101}\sum_{j=1}^{63}\left(\frac{1.3\,x_{ij}}{t_{ij}}\right)}{n_{ij}} \qquad (2)$$

According to Equation 2, the rate obtained by converting the required time and target storage capacity to Mbps and multiplying by 100 was higher than the average of 95% in all the agencies, and the average rate was 96.25% as a whole.

4.2.3 Backup success rate

The backup success rate of the second remote backup system performance item means the success or failure of the backup, and is a percentage of the number of times that the backup is successful among 63 days of execution for 101 hosts, whose equation representing this is as backup success rate $= \frac{s_{ij}}{n_{ij}} \times 100 = \left(1 - \frac{f_{ij}}{n_{ij}}\right) \times 100$ but, $\frac{s+f}{n} = 1$, (3)

Number of Backup Success = s
Number of Backup Failure = f
Host = i = 1, 2, ..., 101
Backup date = j = 1, 2, ..., 63

According to Equation 3, the average backup success rate of the research system can be calculated as Equation 4.

$$\text{Average backup success rate} = \frac{\sum_{i=1}^{101}\sum_{j=1}^{63}\left(\frac{s_{ij}}{n_{ij}}\times 100\right)}{N} \quad (4)$$

Based on the above formula, the number of successes in the number of measurement data times the number of servers multiplied by the number of days of measurement was calculated, and in most agencies, the times of backup failures were 1~2, showing overall success rate of 99.86%.

4.2.4 De-duplication rate

The de-duplication rate, the third remote backup system performance item, means the percentage of the total throughput that needs to be backed up to eliminate duplication and actually store it, and when representing it as a formula, it is like Equation 5.

$$\text{de-duplication rate} = \left(1 - \frac{r_{ij}}{b_{ij}}\right) \times 100 \quad (5)$$

Total throughput = b
Actual storage after de-duplication = r
Host = i = 1, 2, ..., 101
Backup date = j = 1, 2, ..., 63

According to Equation 5, the average de-duplication rate of the research system can be calculated as Equation 6.

$$\text{Average de-duplication rate} = \frac{\sum_{i=1}^{101}\sum_{j=1}^{63}\left(\left(1-\frac{r_{ij}}{b_{ij}}\right)\times 100\right)}{N} \quad (6)$$

According to the above equation, the de-duplication rate is distributed on an average of 93%~100%, according to the above equation, and it has shown that generally 98.94% of duplication is removed.

4.2.5Research system performance item measurement result

As described above, although the number of servers and the capacity vary depending on the characteristics of each agency, we can see that, when obtaining the performance items such as the speed, the backup success rate, and the de-duplication rate, its average value is deducted approximately within a certain range.

Also, the standard deviation is relatively low compared with the measured value of the number of servers and the capacity. So it can be judged that the volatility of the agency characteristics becomes standardized to some degree, thus, it can be viewed that it contains unique as a criteria representing the function of the remote backup system. Finally, the measurement results for the three performance items are shown in *Table 5 & 6*.

Table 5 Summary of measurement results of research system performance items

Agency division	Agency	Number of servers	Measuring days	Measurement material number	Lead time (Min)	Volume (GB)	Target saves volume (MB)	de-duplication rate (%)	Velocity (%)	Backup success
		Pcs	(Day)	(Number of servers ×measuring day)	Average (s.d.)	Average (s.d.)	Average (s.d.)	Average (s.d.)	Average (s.d.)	Times of success (Success rate)
Small agency	A	2	63	126	22.78 (22.69)	44.17 (21.01)	1639.62 (1629.69)	97.52 (2.43)	95.56 (4.48)	126 (100.00)
	B	11	63	693	0.61 (0.72)	41.72 (36.59)	44.18 (51.43)	99.89 (0.09)	96.25 (4.90)	691 (99.71)
	C	10	63	630	5.12 (14.06)	399.23 (396.60)	370.35 (1021.21)	99.76 (0.53)	96.29 (5.28)	630 (100.00)
	D	3.	63	189	8.08 (11.25)	31.39 (24.87)	585.82 (813.60)	98.98 (1.47)	96.38 (5.54)	189 (100.00)
	E	9	63	567	0.75 (0.87)	246.97 (296.70)	53.99 (63.69)	99.98 (0.04)	96.13 (5.10)	565 (99.65)
	F	2	63	126	0.14 (0.32)	126.23 (36.88)	10.06 (22.70)	100.00 (0.01)	96.35 (5.08)	126 (100.00)
	G	7	63	441	4.58 (24.79)	156.08 (112.25)	338.10 (1845.68)	99.63 (1.46)	96.56 (4.57)	441 (100.00)
	H	3.	63	189	90.83 (130.84)	223.08 (184.45)	6437.36 (9254.76)	93.04 (10.02)	96.29 (5.45)	189 (100.00)

Agency division	Agency	Number of servers	Measuring days	Measurement material number	Lead time (Min)	Volume (GB)	Target saves volume (MB)	de-duplication rate (%)	Velocity (%)	Backup success
		Pcs	(Day)	(Number of servers ×measuring day)	Average (s.d.)	Average (s.d.)	Average (s.d.)	Average (s.d.)	Average (s.d.)	Times of success (Success rate)
	I	10	63	630	15.44 (49.84)	131.34 (89.14)	1116.51 (3584.48)	99.62 (0.98)	96.64 (4.97)	630 (100.00)
	J	2	63	126	0.60 (0.45)	40.60 (29.88)	43.02 (32.47)	99.68 (0.33)	96.39 (5.24)	125 (99.21)
Large agency	K	11	63	693	13.71 (26.60)	2454.93 (6969.85)	987.81 (1935.95)	98.83 (1.03)	96.16 (5.24)	693 (100.00)
	L	8	63	504	7.46 (22.46)	129.05 (142.26)	534.36 (1581.83)	99.90 (0.00)	96.05 (5.12)	503 (99.80)
	M	10	63	630	9.63 (10.25)	93.32 (70.36)	695.19 (742.39)	99.90 (0.00)	96.20 (5.06)	629 (99.84)
	N	13	63	819	24.03 (37.91)	92.58 (84.25)	1732.43 (2714.96)	95.89 (8.60)	96.11 (5.00)	817 (99.76)
Total		101	63	6363	11.98 (37.21)	399.01 (2409.75)	861.16 (2656.57)	98.94 (3.97)	96.25 (5.06)	6354 (99.86)

Table 6 Research system performance item measurement result (n=6,363)

Measuring items	Average	Median number	S.D	Range	Minimum value	Maximum value	Coefficient variation
Transfer rate (%)	96.25	97.96	5.06	32.16	76.15	108.31	0.053
Backup success rate (%)	99.86	100.00	0.64	4.76	95.24	100.00	0.006
de-duplication rate (%)	98.94	99.90	3.97	74.00	26.00	100.00	0.040

5.Conclusion and suggestions

In order to verify the efficiency, sufficiency, and stability of the research model as above, 14 agencies, 101 hosts, and about 6,300 backup legacy data were used. In order to verify the statistical significance of the performance, this study arranged variables corresponding to transfer rate, backup success rate, de-duplication rate from about 6,000 times of actual operating system data, and verified the statistical difference by daily sample average comparison. Delphi survey results showed that there was no statistically significant difference between the reference values of the performance factors of the remote backup systems and the differences between the measured values of the remote backup center implemented in the study. However, as this study did not perform performance verification, such as configuration method, recovery target time, cost, security, consistency, and data type which are important elements of remote backup, it feels that it is required to consider the factors, other than technology, in order to verify various review factors. However, it is expected that, through this research model, persons in charge of small and medium size companies and agencies will be able to apply and use in various ways, according to information system environment of them on site to implement data protection policy through the building of a remote backup center.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Kim YM. Information system disaster countermeasures and disaster recovery center construction. Korea Infrastructure Safety Corporation. 2014.

[2] Toigo JW. Disaster recovery planning: strategies for protecting critical information. Prentice Hall PTR; 1999.

[3] Hiatt CJ. A primer for disaster recovery planning in an IT environment. IGI Global; 2000.

[4] Shin YW, Jun S, Lim C, Kim M. National cyber security damage analysis and alternatives. The Korea Association of National Intelligence Studies. 2013; 6(5):135-55.

[5] http://www.ndsl.kr/ndsl/search/detail/report/reportSearchResultDetail.do?cn=TRKO201600014783. Accessed 10 April 2018.

[6] Gordon C. How to cost-justify a business continuation plan to management. Disaster Recovery. 2000.

[7] Hwang BY, Jung BS. A study of remote backup-center design for the continuous banking service and simulation. The Journal of Korean Institute of Information Technology. 2008; 6(3):169-76.
[8] Night B. A study on the effect of ransomware virus infection on individual cloud storage service use. Thesis (Master) - Graduate School of Chungbuk National University: Department of Management Information Science. 2017.
[9] Seo YW, Jin YI, Lim SM, Song MW, Sun SH, Son SH. A study on the remote backup center of the national fundamental information systems for the governmental administration continuity. Informatization Policy. 2002; 9(3):79-97.
[10] Jeong LY, Hwan LM, Min YS. Feasibility of building remote financial backup center in Busan. The Journal of Business and Economics. 2010; 26(2): 123-51.
[11] Tan Y, Jiang H, Feng D, Tian L, Yan Z, Zhou G. SAM: a semantic-aware multi-tiered source de-duplication framework for cloud backup. In international conference on parallel processing 2010 (pp. 614-23). IEEE.
[12] Jung HM, Kim J, Ko YW. Design and implementation of data deduplication backup system supporting multi-mode operations. Journal of KIISE: Computing Practices and Letters. 2011; 17(4):214-24.
[13] Casella G, Berger RL. Statistical inference. Pacific Grove, CA: Duxbury; 2002.
[14] http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&control_no=f484df296438581effe0bdc3ef48d419. Accessed 10 April 2018.

Hoo-ki Lee is a Ph.D. who majored in IT service management of Soongsil University in Korea. His work as cyber security expert at the Ministry of Culture, Sports and Tourism. He is currently interested in research into Security, Control, E-mail Security and Malicious Code Analysis. He published various papers on Cyber Security in the journal.
Email: hk0038@korea.kr

Sungtaek Lee is a Ph.D. who majored in IT service management of Soongsil University in Korea. He is the director of the Startup Support Team of Soongsil University in Korea. He is currently interested in research into the Online Learning Entrepreneurship Education and Commercialization of Technologies related to the 4th industrial revolution.
Email:totona22@ssu.ac.kr

JongHyuk Seong is an attending Ph.D. who majored in the Department of Information Security Systems of Kyonggi University in Korea. He is currently interested in research into Information Security, Cyber Security, Security Control, Service Level Agreement(SLA), Convergence Security, etc.
Email: jhseong@kcisa.kr

HoGun Rou is an attending Ph.D. who majored in IT service management of Soongsil University in Korea. His research areas Cyber Security, Data Security, Big Data, Artificial Intelligence, and IT Policy Management.
Email: infosecu@igloosec.com

GwangYong Gim is a full Professor in the Department of Business Administration of Soongsil University, Korea. He has been published many papers in journals such as Information Science, Fuzzy Sets and System, Journals of The Society of Management Information Systems and a number of books. His current research interests in 4th Industrial Innovation, Big Data Analysis, etc.
Email: gygim@ssu.ac.kr