Grzegorz Koziel[1]

COSTS OF DATA PROTECTION

No organization can exist without the data. Each kind of activity demands data processing and its storage. Loss of the data is very dangerous. It can even result in an organization's collapse. It is very important to protect the data against loss or damage. Each organization should have appropriate means to secure the data. The choice of solutions depends on the level of safety requirements. Each solution has its advantages, disadvantages and costs. It is very important to choose the best means of protection within the costs accepted by an organization. The presented paper shows the most popular data protection solutions and estimates their costs. The proposals of data protection solutions are presented as well.

Keywords: *data protection, information security.*

Гжегож Козєл

ВИТРАТИ НА ЗАХИСТ ДАНИХ

У статті стверджується, що кожен вид діяльності вимагає обробки даних та їх зберігання, а втрати даних дуже небезпечні, вони навіть можуть призвести до краху організації. Тому дуже важливий захист даних від втрати або пошкодження. Кожна організація повинна мати відповідні засоби для захисту даних. Вибір рішення залежить від рівня вимог безпеки. Кожне рішення має свої переваги, недоліки і витрати, тому необхідно вибирати найкращі засоби захисту, доступні організації. Показано найпопулярніші рішення захисту даних і оцінено їхню вартість.

Ключові слова: *захист даних, інформаційна безпека.*
Форм. 1. Табл. 1. Літ. 10.

Гжегож Козел

РАСХОДЫ НА ЗАЩИТУ ДАННЫХ

В статье утверждается, что каждый вид деятельности требует обработки данных и их хранения, а потери данных очень опасны, они даже могут привести к краху организации. Поэтому очень важна защита данных от потери или повреждения. Каждая организация должна иметь соответствующие средства для защиты данных. Выбор решения зависит от уровня требований безопасности. Каждое решение имеет свои преимущества, недостатки и издержки, поэтому необходимо выбирать лучшие средства защиты в пределах средств, доступных организации. Показаны самые популярные решения защиты данных и оценена их стоимость.

Ключевые слова: *защита данных, информационная безопасность.*

1. Introduction. The data is a very important factor in each organization's activity. The bigger an organization is, the more important its data is. Very often it is impossible to operate without data processing. The data is necessary to control the stock, to process accountant operations, to control processes (Laskowski, 2011). It is a standard to use the data in its electronic form. Each organization owns an information system to process and store data. This system usually is based on the IT system. Moreover, according to current laws, it is allowed to resign from paper form in many cases.

The problem of data security must be considered from 2 points of view: technical and human (Juszczyk, 2011). The main issue of human factor in data security is

[1] Dr. Inz, Institute of Computer Science, Lublin University of Technology, Poland.

the authentication, authorization and accounting. Unfortunately, the level of acceptance of these elements is not sufficient in Polish companies (Milosz & Milosz, 2011; Juszczyk, 2012). The main point of technical data security is to ensure its security in 3 main domains: integrity, availability and confidentiality.

Integrity means that the data includes no errors and describes the real situation as precisely as it is possible or necessary. Data usage must ensure the possibility of controlling real situation in organization.

Availability is interchangeable with ensuring the access to the data to every authorized person at each time and each place it is demanded, but with the condition that it is compatible with the organization procedures and security rules (Ping An Wang, 2010).

Confidentiality is the most difficult condition to ensure. Each data set has to be accessible only to authorized people. It means that various data sets will be accessed with various privileges by various people. Unauthorized people can not have access to data in any form – electronic, paper or any other.

In each organization there exists a set of threats. This set varies form the organization and its data value. To protect a good security system it is necessary to analyze what kind of threats exist in the organization, calculate the budget available to buy protection and finally choose the appropriate means and workout the security system.

2. Security threats. It is necessary to identify the security threats before a form of protection is chosen. The most often occurring dangers are (Polaczek, 2006):

- Failures of the equipment being a part of the information system, especially parts of the IT system.
- Software breakdowns.
- Power supply failures.
- Unauthorized data reading.
- Unauthorized data modification.
- Data damage.
- Fabrication of data.
- Transmission overhearing.
- Impersonating a user.
- IT system hacking.
- Unauthorized access to IT system.
- Malicious software.
- Data theft.
- IT system's parts theft.
- Passing access to data by authorized users to unauthorized users.

The first step in preparation of the security plan should be the analysis of the listed above security threats, grade the risk of their occurrence and the costs of possible loses if they happen. On the basis of this analysis it is possible to adjust appropriate protection and rate the reasonable costs of implementing the solution.

The risk of the particular security threat occurrence is usually graded on the 3-point scale: 1 – low probability of risk occurrence, 2 – medium probability of risk occurrence, 3 – high probability. 3-degree scale is used to rate the results of each risk: 1 – insignificant consequences, 2 – medium consequences, 3 – significant consequences. Of course, the scale of rating the risk and its consequences can be more pre-

cise, but it is very difficult to precisely rate these values. Consequences of each security threat are usually calculated on the basis of possible losses for an organization. It includes financial losses, amount of work necessary to restore data if possible, difficulties in further work and is connected with these losses and law consequences of data loss.

Significance of each risk is calculated on the basis of calculated values according to the formula 1:

$$S = P \times C. \quad (1)$$

In the formula 1 the symbols used mean: S – significance of the risk; P – probability of the risk occurrence; C – consequences level of each risk.

If the significance of the risk ranges from 1 to 3 it is a low risk. Values ranging from 4 to 6 mean medium risk. Values above 6 define high risk. The higher the risk is the more needed is the protections against a particular risk.

3. Security threats and protections against them. Hardware malfunction is one of the most often occurring data security breakdowns. Most often it causes lack of the resources availability. If one of devices stops working it results in lack of communication. Sometimes, hardware malfunction can be a cause of data confidentiality or integrity breakdown, especially when one of devices responsible for delivering protection breaks. In practice these threats are very rare because almost all the devices that contain the data have protection included.

It is impossible to avoid hardware breakdown. Each equipment breaks because of defects and age. Even the best solutions, tested and certified can break. The defense against breakdowns relies on building IT systems in such a way that allows continuing work even if one of the parts breaks. Of course, it is the ideal solution but very expensive. In real cost-effective solutions, IT system is built in a way that allows quick system restart after failure. It allows for saving money spent on protections and reducing costs caused by failure.

One of the most susceptible parts of computer systems is hard disc drive. It is the most common data storage. Usually this device is the basic part of each data storage. Hard disc failure is the common reason of data loss. It is necessary to protect the data written on the disc against loss. One of the simplest solutions is backup. It is a copy of the data, written on the other carrier. If the hard disc breaks, the data from the backup can be copied back to the IT system. Of course, it is impossible to be up to date with the backup. The data copy can be done with certain frequency. It is possible to restore data up-to-date in the time of the backup creation. It is necessary to complete the data after restoring to be up to date. Overall, backup is a very effective way of protecting data against loss (Chang, Cung-Yen, 2005, Cherkasova, Zhang, Xiaozhou, 2010). Costs of creating backups contain:

- Costs of a device used to backup data – to create backup it is necessary to have a device to record data. The most popular are DVD recorders or streamers. DVD recorder is cheap. It costs about 25$. Its disadvantage is relatively small disc's capacity which is equal to 4,7 gigabytes. Streamer is a much more expensive device. Its cost ranges from 1100$ to 3000$. Streamer allows recording magnetic tapes that have capacity equal to 200 gigabytes. That device allows compressing recorded data what results in 400 gigabytes tapes capacity.

- Costs of carrier – each backup has to have a carrier to record it. Cost of the carrier depends on the kind of a carrier. DVD single costs about 0,3$. Rewritable DVD costs about 0,4$ but allows multiply recording. Magnetic tape used in streamers costs about 30$ and allows multiple recording as well. It is possible to record backups on USB hard discs. Cost per one disc having 2 terabytes capacity is about 125$. Disc usage does not demand additional devices to record data. It is possible to attach the disc directly to the computer.

- Costs of employees who prepare backup scripts and take care of backup carriers storage. Cost of time used by people is variable depending on the region and worker qualifications. In each type of copy the worker has to spend almost the same amount of time to complete the procedure. His responsibilities include: putting a carrier in a recording device, marking a copy and delivering it to the safe place to keep.

- Costs of carriers storage – it is necessary to protect backups against theft, damage, or interference. It is only possible by placing them in a secure place. Usually, safe boxes are used to keep backup carriers. A safe box protects against theft, fire, flood and other detrimental factors, but in big disasters even the safe box can not protect copies. Some companies rent bank lockers and place there the backups. Usually, backup is prepared once a day. Once a week, one backup is put into a bank locker. Cost of a bank locker rental is not very high. It is about 40$ per month. Safe box cost depends on its quality, robustness, and size and starts from 400$. Of course, it is possible to buy cheaper safe box, but usually it does not suit requirements. While buying a safe box it is necessary to take care about the law. Some kinds of data must be secured at a certain level. For example, in Poland, the law restricts conditions of personal data storage.

Other possibility to protect data in the case of disc failure is to create a redundant array of independent discs (RAID). RAID is a technical solution that allows combining a few hard discs drives to work as if they were one device. Devices included into an array can have various logic organizations called RAID levels. The most important levels from the point of view of data safety are (Xiao, Ren, Yang, 2009):

- RAID 1 – an array composed of 2 hard discs. Each of the discs is used to keep one copy of data. As a result, 2 copies of data are saved in array. It allows for obtaining robustness against one disc failure – the data is not lost and array is still working. Because each data is written twice, one disc capacity is used for copies. It means that available space on the array is equal to only one disc space. The space offered by the second disc is the cost of array creation. Costs of the array sum up to the additional disc's cost and the cost of the array controller device if needed. On Linux machines it is possible to create software RAID. Then the controller role overtakes software. Cost of an additional hard disc drive having 2 terabytes capacity is about 125$. RAID controller costs from 60$ to 2000$. The bigger the bandwidth and number of discs possible to connect are, the nigher the price of a controller is. For a small file server having up to 4 discs and one network connection it is enough to supply it with one of cheaper controllers.

- RAID 5 – an array composed of at least 3 discs. In the matrix containing n discs, data is spread on $n - 1$ discs. Each disc contains $1 / (n - 1)$ part of the data. The last disc is used to write parity bits. It allows calculating the data stored on one of the discs when it broke down. Raid 5 allows obtaining robustness to one disc failure.

Array capacity is equal to $n - 1$ discs capacity. The lost of the space available to store data is always one disc capacity. The percentage of this loss depends on the discs number in the array. The greater discs number is, the smaller is the loss percentage. This RAID type allows increasing efficiency of the array because of parallel writing and reading operations done on a few discs at the same time. Costs of array creation include costs of additional discs and RAID controller. Particular devices cost is the same as presented in above point concerning the RAID 1. The difference is in minimum discs number. Here it is necessary to poses at least two additional discs.

- RAID 6 – an array composed of at least 4 discs. In the matrix containing n discs, data is spread on $n - 2$ discs. Each disc contains $1 / (n - 2)$ part of the data. 2 additional discs are used to write double parity bits. It allows calculating the data stored on 2 of the discs when they broke down. Raid 6 allows obtaining robustness to 2 discs failure. Array capacity is equal to $n - 2$ discs capacity. The lost of the space available to store data is always 2 discs capacity. The percentage of this loss depends on the discs number in the array. The greater discs number the smaller the loss percentage. This RAID type allows increasing efficiency of the array because of parallel writing and reading operations done on a few discs at the same time. Write time is better in RAID 5 matrix because of smaller parity bits calculating complication and smaller disc number to control. Costs of the array creation include costs of additional discs and RAID controller. Particular devices cost is the same as presented in two above points concerning other discs arrays. The difference is in minimum discs number. Here it is necessary to put at least 3 additional discs.

The most advanced solution is a computer cluster. It is a set of computers that works as if it were a single machine. In the cluster it is possible to use redundancy of components. It means that 2 or more computers can be "mirrors". In this case each of computers is a copy of the second one. They work in the same way and keep the same data. It is possible due to high speed network connection that allows synchronizing them in real time (Stallings, 2003). A cost of that solution is the highest of presented above solutions. It demands additional computer and fast network connection between computers. Additionally, it is worthy to place computers in separate rooms or even buildings to protect them in the case of fire. Each computer included in a cluster should be supported with discs array. Cost of an additional computer is at least 1000$. In return, owner gets very high level of reliability.

Table 1. **Various protections comparison**

Solution	One disc failure robustness	Two discs failure robustness	Real time data copy	Computer failure	Work continuity while breakdown of:			Cost of cheapest solution: Devices cost at start/ backup cost per 1GB copied data each time
					One disc	Two discs	Computer	
Backup	yes	yes	no	yes	no	no	no	25$/0,1$ + work at least 10 minutes per backup
RAID 1	yes	no	yes	no	yes	no	no	310$/0
RAID 5	yes	yes	yes	no	yes	no	no	405$/0
RAID 6	yes	yes	yes	no	yes	yes	no	530$/0
Cluster*	yes	no	yes	yes	yes	no	yes	2000$/0

*Cluster contains 2 computers, each of them contains 1 disc.

To obtain high level of data safety it is necessary to combine at least 2 presented solutions. It is known that creating backups is necessary. Additionally, it is worthy to use disc matrix and/or computer clusters. Each solution advantages and costs comparison is presented in Table 1. Of course, together with presented hardware solutions it is necessary to use additional protections against malicious software, intrusions, power failures, theft, overhearing and other identified security threats which might occur.

4. Conclusion. Data is really a strategic asset. Without it further organization functionality is very difficult or even impossible. It is necessary to protect the data against lost or damage. Costs of protection are not low but in comparison with costs of possible data loss they are not high. It is worthy to protect the data at least by creating backup copy. It gives us possibility for restoring the data at low cost. Of course, the presented solutions do not exhaust all the possibilities but allow avoiding main, most frequent security threats.

References:

Chang, C.-Y. (2005). A survey of data protection technologies. Electro Information Technology, 2005 IEEE International Conference. Pp. 1–6.

Cherkasova, L., Zhang, A., Xiaozhou, L. (2010). DP+IP = design of efficient backup scheduling. Network and Service Management (CNSM), 2010 International Conference on Topics: Communication, Networking & Broadcasting; Computing & Processing (Hardware/Software); General Topics for Engineers (Math, Science & Engineering). Pp. 118–125.

Juszczyk, M. (2011). Impact of human factor in data security. Actual Problems of Economics, 120(6): 359–364.

Juszczyk, M. (2012). Digital identity acceptance at Polish large enterprises: the survey results. Actual Problems of Economics, 132(6): 496–502.

Laskowski, M. (2011). Critical date bugs and their impact on computer-based economy. Actual Problems of Economics, 120(6): 335–339.

Milosz, E., Milosz, M. (2011). Digital identity management in Polish SMEs. Actual Problems of Economics, 120(6): 340–345.

Ping, A.W. (2010). Information security knowledge and behavior: An adapted model of technology acceptance. Education Technology and Computer (ICETC), 2010 2nd International Conference on: Communication, Networking & Broadcasting; Computing & Processing (Hardware/Software); Robotics & Control Systems, Vol. 2. Pp. 364–367.

Polaczek, T. (2006). Audyt bezpieczenstwa informacji w praktyce. Helion, Gliwice.

Stallings W. (2003). Systemy operacyjne. Wydawnictwo Robomatic, Wroclaw.

Xiao, W., Ren, J., Yang, Q.K. (2009). A case for continuous data protection at block level in disk array storages. IEEE transactions in parallel and distributed systems, 24: 898–911.

Стаття надійшла до редакції 14.12.2012.