Make backups or pay!

András László Keszthelyi

Obuda University, Keleti Károly Faculty of Business and Management,
Keszthelyi.Andras@kgk.uni-obuda.hu

Abstract: Cybersecurity company Sophos has published its fifth annual report on The state of Ransomware 2025. Their research investigated the impact of ransomware attacks on different companies. The average amount of money that companies paid for getting their data back was 1 million dollars, so it is clear that cybercrime in general is a huge „business field" and ransomware attacks in particular are very serious problems for companies. Defence technologies evolve quite quickly but 100% security level cannot be reached. In this paper I investigate how data backups can help minimizing the losses in case of a successful ransomware attack – or any other problem resulting in data loss.

Keywords: ransomware, data bacup

1 Introduction

What can cause data loss? A lot of things. Only a few examples will follos:

A natural catastrophy such as an earthquake (see Fukushima), fire, firemen's activity in case of fire, pipe burst. Sudden cut-off of electricity. Secondary effect of a lightning strike.

Wear and tear of hardware elements, especially hard drives (including solid state disks). Even in ideal working conditions the hardware elements will go wrong after a while.

Someone can steal a computer not because of the data stored in it but for the price of the hardware.

A computer can be stolen in order to get the data stored in it. It can be, in some circumstances, the simplest way to acquire the data.

Human errors. Someone can erase the data accidentally. Or as a form of revenge. There may be software errors as well.

Malicious software, commonly called as "viruses". [5] Among them quite a new "star": ransomware. [6] ItFdu is a special kind of malware (malicious software) that uses encryption in the victim's system making the data unavailable for regular use. If the victim pays a ransom s/he will be sent the decryption key. At least theoretically. There are cases when the victims didn't get the decryption key and there is at least one case when the encryption process had a bug and it was mathematically impossible to decrypt, even having the would-be decryption key.

Ransomware is quite a new branch in the field of cybercrime. Sophos antivirus company ran an annual research in 2024 (as well), asking some thousand companies in all over the world about their ransomware related experiences. [7] They found that 59% of companies experienced ransomware attacks in the previous year. More than half of them agreed to pay the ransom. “The average (median) payment has increased 5-fold over the last year, from $400,000 to $2 million” [7] depending on the size of the company. According to Ákos Bódis a Hungarian cybersecurity expert [8] the ransomware business could be the third largest economy in the world, after that of the US and China.

It is because there are some very serious differences between the traditional, analogue, physical world and the virtual world. In this situation the most important difference is that in the virtual world both the crime and the payment can be done anonymously and without useful tracks because of the special attributes of the virtual world. [9] No need to be physically on the spot as Ronald Biggs had to be on the Glasgow-London mail train. And no need to meet in person to get the ransom or find out so special tricks as D. B. Cooper had to do on the Portland-Seattle flight. As a result, cybercrime means significantly less risk than traditional crime, generally speaking.

2 Value of data

IT-systems of everyday life consists of, at least, three main parts. First one is the hardware. It is of little value, even if it has its price that can be even very high. In other words, if hardware goes wrong, gets too old, disappears, etc., you can buy new hardware. It is only a matter of (enough) money.

The second component is the software. It is very similar to the hardware from this point of view. Has its price but if you have the licence for the software you use you can re-install it. Or, if you don't have any licences for it, you can get it from the same source you got it previously. From the practical point of view the result is the same: after having bought new hardware and re-install your software you have your working environment again. Let it be either a single home computer or quite a big server farm of an enterprise. Software is also a matter of money only.

The third component is the data you store in your computer system. It is the most valuable part of your system. A possible classification of data is the following: you can divide your data into two groups.

The first one is the so-called reproducable data while the other group is the irreproducable data. It means that if you have lost your, e.g., accounting data (supposing you are a company) or you have lost, e.g., your just finished final thesis two days before the deadline just before printing it out then you are in trouble. Good news that this kind of trouble can be managed and you can reproduce your data. You can get the original bills an invoices so you can enter the accounting data for the second time into your newly-built system and then you can continue the regular everyday work. Obviously, this process needs human resources, time, etc., in one word: money, but finally you can restore the original state of your working system and you can continue your work in the normal way. If you are a university student you'll have to re-writre your final thesis and you can apply for the final exam – maybe a semester later, but you'll realize how easier to write the final thesis for the second time. To minimize the cost of such an event you want to develop your own business continuity plan, BCP (see, e.g., ISO 22301:2019). In some cases you can try some data recovery services, such as the Hungarian Kürt Zrt.

As for the second group of data, the irreproducable data, the situation is more complex. There is no way to get your irreproducable data if you have lost it. If you have lost your incoming orders (as a company running a webshop) before processing it, you will never have or make the possibility to get them once again. Or if you have lost the meteorological data you measure hourly in different points of the Carpathian Basin, there are no possibilities to re-measure them. It is also impossible to re-take the family photos of the last summer holiday.

So the irreproducable data needs to be protected the most.

What can you do? Or a better question: what do you have to do? At least three important things: a) make backups, b) protect it, c) use your mind (if there is any).

The most important activity you are supposed to do is to make data backups. If you have the appropriate backups then you will have the possibility to restore your data after any disasters. Secondly, if you have a well-planned (and done) backup process then you can start to work on not needing the backups. In this step you'll set up a firewall, anti-virus software, intrusion detection system [1], intrusion prevention system [2] or whatever you think important and can do.

The third element in preventing data loss is the human factor: Notice every unusual circumstance. A good example of this from the near past is the case of Andres Freund, Last year (in 2024) a computer user, working for Microsoft as a developer, installed a new, beta Debian. He found that to set up an ssh connection became longer than it was previously. The difference was about half a second, just a man can perceive. And he started to investigate what the reason was and found a serous backdoor that could have been a real world-wide catastrophy if he hadn't discovered it. [3] [4]

3 Data backups

The most important task is to have a proper backup procedure either you are a big company or a private person. If you have backups then the second most important task is tu develop and use defending procedures at the physical, administrative and algorithmic levels of defence. The first one means that you are supposed to close the door of the server room (don't leave it on the seat of your car) and provide your equipment with the optimal or, at least, satisfactory working conditions. The second means you must have the appropriate regulations about who and when can enter the server room at all. While the third means that you can, and have to, use the computers themselves in the defence, firewalls, etc.

Data backups belong to the third level, mainly. But without the appropriate regulations (second level) it will not be satisfactory. How to plan, organize a good-enough backup system? The main parameters are: the amount of data, how quickly the data changes, how much time we have for recovery (after a data loss event). In following part of this paper I'll investigate two simple procedures that can be used in general, office like situations including, of course, the “home office”.

The “general office situation” means that the total amount of data is a few gigabytes on a single office computer, consisting of, mainly, docx, xlsx, pptx, pdf files, there maybe saved html files and some jpg pictures. E.g. my workspace, as an associate professor in IT, contains about 20 thousand files (of which about 4.600 is created or last modified this year) in about 6 gigabytes, after many years of work.

The very basic backup is to copy all the files onto another storage equipment and put that equipment far from the original computer. There are more effective ways of backing up, let's see two of them: the incremental and the differential backups.

3.1 Incremental backup

In the first step a full backup must be performed, obviously. Then in regular intervals, e.g. daily, you have to copy the files that have been modified or created since the previous backup was created. If the backups are done at the end of the office hours then, e.g., the backup created on Tuesday evening will contain all the files that were created or modified on Tuesday (supposing that on Monday evening the previous step was accomplished correctly).

It is quite fast, needs little time and not much storage space as one incremental backup step will contain a small number of files. On the other hand, if you have to restore your workspace after a data loss, you'll have to copy back the full backup created in the first step and then each incremental steps in the appropriate order overwriting possibly existing files. So the backup process needs little time while the restoring process needs significantly more. If any of the backup files is corrupted or disappeared then your data after that link or point may not be correctly restored.

3.2 Differential backup

The first step is the same, you must make a full copy of your workspace. Then in the subsequent steps you copy all the files that differ from the ones in the full backup of the first step. It means that as time goes on the backup packages will grow bigger and bigger. For example you create the first, full backup on Monday evening. The Tuesday backup package will contain all the files that were created or modified on Tuesday. The Wednesday backup package will contain all the files that were created or modified on Tuesday and Wednesday and so on.

In this case the backups need (a bit) more and more time and storage space every day. But after a data loss the restoring process is simpler and quicker: you'll have to copy back the first (full) backup package and the very last one. The risk of a broken backup chain is significantly less than in case of the incremental backup. Looking at the fact that in case of general and regular office-style circumstances a full data loss does not happen frequently (luckily), in such circumstances I'd prefer the first, incremental, method. In case of both methods you'll have the earlier versions of a file on which you were working for long, so it can give additional possibilities against logical or human errors.

3.3 Backup tools

In case of Windows there is the xcopy command you can use in a Command window – yes, Windows has a command line!

In case of Linux you have more possibilities: one is mirroring tool rsync but it needs a remote computer to copy to and quite a good network connection. And you can do the incremental/differential backups with find and tar and gzip.

On Windows the xcopy command [10] can do the job. Each file in a Windows filesystem has a so-called archive bit. It is set by the operating system every time when the given file has changed (including its creation as well). The xcopy command can use this archive bit to decide whether a given file should be copied or not.

The command “xcopy /?” displays its help. The way it should be started is “xcopy xcopy <Source> <Destination> <options>”. Among its options the very basics ones are: /s copies directories and subdirectories, unless they're empty. The /e option copies all subdirectories, even if they're empty and can be used together with /s. The /a option copies only *source* files that have their archive file attributes set. **/a** doesn't modify the archive file attribute of the source file. The /m Copies *source* files that have their archive file attributes set. Unlike **/a**, **/m** turns off archive file attributes in the files that are specified in the source.

In addition to the xcopy command there is an attrib command that can be used to set (or clear) the archive bits, recursively, if needed.

So we have the possibility to make either incremental or differential backups with the /a or the /m option.

In case of Linux there are other possibilities as we have no archive bits. Without an archive bit we can use the timestamps of the files. When starting the backup in the first step we create a file to preserve the date and time of the given backup. Every file that is newer than this should be copied, recursively, next time. If we do incremental backup the timestamp of this file should be set at the beginning of each process while in case of differential backup we set the timestamp only in the first step (the full backup).

Then we can use the find command [11] to find the necessary files. This find command has the option "-newer timestamp-file". Based on this option we can find those files (with the find command:) that are newer and so they must be backed up. We can redirect the output of the find command to a text file and then we can use the tar command [12] to put these files one by one into a single tgz file (something like a zip file). Then, or even in real time, we can send this package to another place, either to a remote computer via the network or to a local storage device what can be even (simplest case) a usb stick.

4 Cloud

The so called "cloud" is someone else's computer. Using the cloud can be very comfortable for a company, especially when different users must access the same files regularly, so keeping then in one place that everyone can access simplifies the organisation of the work very much. The question is: In case of using a "cloud" as a service provided by someone else, should we do our own backups or should it be enough for us that we have a contract with them containing guarantees for the availability of the cloud and keeping our files safe?

I would say that guarantees are nice but to have one (more) copy of my data in my hands is nicer. As the old rule says: In case of digital data one copy is not a copy, two copies is half a copy, three copies is one copy (and security starts at four copies). It may sound sort of funny but there is an old Hungarian wise saying: "Better safe than sorry".

Conclusions

In the virtual or digital world the digital data is a (the?) real value so we are supposed to keep our data safe. As there are no storage devices that could give us a guarantee for being able to retrieve our data from it, not even for a limited time period, we have to make backup copies if we want a higher level of security. It has become more important since ransomware attacks started.

There are a lot of aspects of developing a backup process for a given person or company, in this paper I gave an outline for this, supposing an average, general office like environment in general, and also in particular both for Windows and Linux.

One should keep in mind that the price and security level function is similar to the function $f(x) = -1/x + 100\%$. It never, never reaches the 100% level but it can go quite close to it. The higher security level you want the much higher investment you have to do.

Acknowledgement

I have to say a big “Thank you, professor!” to Mr. Ervin Szűcs, my late professor at the Department of General Technics who let me use the supercomputer of the department when I was a first year student. Without that, most probably, I haven't become an IT guy. That supercomputer was an IBM compatible XT with 640 kB RAM and a 10 MB hard drive and was the only one our department had at those times.

References

Online sources and links were checked 1 Oct 2025.

[1] Hung-Jen Liao, Chun-Hung Richard Lin, Ying-Chih Lin, Kuang-Yuan Tung (2013). Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, Volume 36, Issue 1, pp. 16-24. ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2012.09.004.

[2] Safana Hyder Abbas, Wedad Abdul Khuder Naser, Amal Abbas Kadhim (2023). Subject review: Intrusion Detection System (IDS) and Intrusion Prevention System (IPS). Global Journal of Engineering and Technology Advances, eISSN 2582-5003, 14(02), pp. 155–158.

[3] Hern, Alex (2024). TechScape: How one man stopped a potentially massive cyber-attack – by accident, The Guardian, 2 April 2024, https://www.theguardian.com/global/2024/apr/02/techscape-linux-cyber-attack

[4] Goodin, Dan (2024). Backdoor found in widely used Linux utility targets encrypted SSH connections, Ars Technica, 29 March 2024, https://arstechnica.com/security/2024/03/backdoor-found-in-widely-used-linux-utility-breaks-encrypted-ssh-connections/

[5] Young, A.; Yung, M. (2004). Malicious Cryptography: Exposing Cryptovirology. Wiley. ISBN 978-0-7645-4975-5.

[6] Min, Donghyun; Ko, Yungwoo; Walker, Ryan; Lee, Junghee; Kim, Youngjae (July 2022). "A Content-Based Ransomware Detection and Backup Solid-State Drive for Ransomware Defense". IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 41 (7): 2038–2051.

[7] Sophos (2025). The state of ransomware 2025. Findings from an independent survey of 3,400 IT and cybersecurity leaders across 17 countries whose organizations were hit by ransomware in the last year. .A Sophos Whitepaper. June 2025. https://assets.sophos.com/X24WTUEQ/at/9brgj5n44hqvgsp5f5bqcps/sophos-state-of-ransomware-2024-wp.pdf

[8] Mucsi, Viktor (2025). Ez lenne a világ harmadik legnagyobb gazdasága, ha nem lopnák el az összes pénzt, Index, 19 Sep 2025, https://index.hu/techtud/2025/09/19/kiberbiztonsag-kiberbunozes-hacker-bodis-akos-szakerto-tortenelem-bitcoin/

[9] Keszthelyi András (2022). Some Special Results Of ICT Revolution In: Živan, Živković (eds.) XVIII. International May Conference on Strategic Management – IMCSM22 Бор, Srbija : University of Belgrade, Technical Faculty in Bor, Engineering Management Department (EMD) (2022) pp. 479-484.

[10] Xcopy (2024). Microsoft Learn, https://learn.microsoft.com/en-us/windows-server/administration/windows-commands/xcopy, 29/05/2024.

[11] find(1p) Linux manual page (no date) Linux man pages online, https://man7.org/linux/man-pages/man1/find.1.html

[12] tar(1) Linux manual page (no date) Linux man pages online, https://man7.org/linux/man-pages/man1/tar.1.html