

DATA STORAGE: FROM DIGITAL TO BIOLOGICAL

⁽¹⁾Shweta Rana ⁽²⁾Ruchi Kamra

⁽¹⁾⁽²⁾Amity University Haryana

Abstract: *Digital data has changed the use and access of information. Everyday lot of data is produced and this requires high-density storage devices which can retain values for a long time. The demand for data storage devices is increasing day by day as more and more data is generated every day. Deoxyribonucleic acid (DNA) can be potentially used for these purposes as it is not much different from the conventional method used in a computer. DNA can be used as a robust and high-density storage device even under unfavorable conditions. In this paper, we are going to comparing the Digital Storage and Biological Storage.*

Keywords: *DNA, Cloud, encoding, compression, storage mechanism, digital data, secured storage*

1. Introduction

The total information in digital format in the year 2012 was about 2.7 zettabytes. Presently devices such as optical discs, portable hard drives, and flash drives are used to store data. But silicon and the other non-biodegradable materials used in data storage pollute the environment. Also, they are available in limited quantities. Thus, they would be exhausted one day. The linear density of digital storage device is 10 kb per square mm. Hence, newer technology is needed for data storage and archival process. As the data increases, the current data storage technology would not be enough to store data in future as data is growing every day. Even potentially important information can get lost due lack of storage space. One of the most common cause of data loss is accidental deletion of files without backup. Every day many people lose important data because of deleting files accidentally because they do not have proper backup systems. Poor handling of the optical disk can cause data loss in them. Data loss can occur due to damage of hard drive. Mechanical damage to the hard drive is common as it contains a lot of fragile parts moving at very high speed. Hard drives can get damaged due to accidental drop of computers. Hard drives can get damaged if any liquid enters it. Liquids can cause damage to electronic parts of drive making it difficult to recover data. The hard disk can get damaged due to fire. Solid State Drive has a limited number of write cycles. Thus after write cycle limit, it is not possible to write data on them. A printed book has better life expectancy than best of the data storage method. There are many ways to backup the data. One can use cloud services to store data. But to access data which is stored in a remote cloud, an internet connection is needed all the time.

Thus without an internet connection, it is not possible to access the data which is stored in the cloud. Another way is to store data on an external drive. But external drives are prone data loss too. Scientists and Researchers for over the past decade, have been trying to develop a robust way of storing data on a medium which is dense, robust and everlasting. They are sticking to the storage medium which is used by nature that is DNA. There are many reasons to use DNA as the storage medium such as small size and high density.

2. Cloud Storage

Cloud storage is believed to have been invented by computer scientist Dr. Joseph Carl Robnett Licklider in the 1960s. About two decades later, CompuServe began to offer its customers small amounts of disk space in order to store some of their files. In the mid 1990s, AT&T launched the first all web-based storage service for personal and business communication. Since then, a number of different services have become gained traction. Some of the most popular cloud storage providers as of early 2017 are Apple (iCloud), Amazon (Amazon Web Services - AWS) and Dropbox, as well as Google Docs and Gmail.

Cloud storage is a term that refers to online space that you can use to store your data. As well as keeping a backup of your files on physical storage devices such as external hard drives or USB flash drives, cloud storage provides a secure way to remotely store your important data. Online storage solutions are usually provided using a large network of virtual servers that come with tools for managing files and organizing your virtual storage space.

2.1 How Cloud Storage Works

The simplest type of cloud storage occurs when users upload files and folders on their computers or mobile devices to an internet server. The uploaded files serve as a backup in case the original files are damaged or lost. Using a cloud server permits the user to download files to other devices when needed. The files are typically protected by encryption and are accessed by the user with login credentials and password. The files are always available to the user, as long as the user has an internet connection to view or retrieve them.

2.2 Examples of Personal Cloud Storage Options

Although there are many cloud storage providers, some of the more familiar names include:

- **Amazon Drive** offers users 5GB of free cloud storage to users who want to store files, photos, and videos. If the user has an Amazon Prime account, the free plan includes unlimited photo storage and 5 GB for other file types. Users who want more space on Amazon Drive can purchase 100 GB or 1 TB for a fee.
- **Google Drive** is cloud storage that is built to work seamlessly with Google products and offers an offline mode for some categories of documents, which does not require internet access. Users receive 15GB free to store their documents, photos, music, and videos, all of which are accessible from any device. For a fee, Google Drive can accommodate up to 30 TB of data.
- **Apple iCloud** is cloud storage for Macs, iPhones, iPads, iPod touch and Windows PCs. Users receive 5GB of free storage for cloud storage of photos, videos, documents, app data and more. Several larger storage plans are available for a fee. You can structure iCloud to automatically back up your mobile devices or handle the file transfers personally.

2.3 Cloud Storage Security

There is so much attention on cloud storage today in the digital era because so much of our sensitive personal data is stored in the cloud whether we voluntarily store it there or whether a company we do business with decides to store it there. As a result, cloud security is a major concern. Users wonder whether their information is safe, and increasing data breaches have demonstrated that sometimes it isn't. Users are also concerned about whether the data they have stored on the cloud will be accessible when they need it. While cloud storage may seem vulnerable due to the prevalence of hacking, the alternatives, such as onsite storage, have security vulnerabilities, too. Company-provided cloud storage can actually improve security by giving employees an alternative to using their personal accounts to back up and transfer files that they need to access outside the office. A good cloud storage provider will save data in multiple places so that it survives any human errors, equipment failures or natural disasters. A reputable provider will also store and transmit data securely so that no one can access it without permission. Some users might also require that data be stored in such a way that it can only be read but not changed; this feature, too, is available through cloud storage.

2.4 Advantages of Cloud Storage

1. **Usability:** All cloud storage services reviewed in this topic have desktop folders for Mac's and PC's. This allows users to drag and drop files between the cloud storage and their local storage.
2. **Bandwidth:** You can avoid emailing files to individuals and instead send a web link to recipients through your email.
3. **Accessibility:** Stored files can be accessed from anywhere via Internet connection.

4. Disaster Recovery: It is highly recommended that businesses have an emergency backup plan ready in the case of an emergency. Cloud storage can be used as a back-up plan by businesses by providing a second copy of important files. These files are stored at a remote location and can be accessed through an internet connection.

5. Cost Savings: Businesses and organizations can often reduce annual operating costs by using cloud storage; cloud storage costs about 3 cents per gigabyte to store data internally. Users can see additional cost savings because it does not require internal power to store information remotely.

2.5 Disadvantages of Cloud Storage

1. Usability: Be careful when using drag/drop to move a document into the cloud storage folder. This will permanently move your document from its original folder to the cloud storage location. Do a copy and paste instead of drag/drop if you want to retain the document's original location in addition to moving a copy onto the cloud storage folder.

2. Bandwidth: Several cloud storage services have a specific bandwidth allowance. If an organization surpasses the given allowance, the additional charges could be significant. However, some providers allow unlimited bandwidth. This is a factor that companies should consider when looking at a cloud storage provider.

3. Accessibility: If you have no internet connection, you have no access to your data.

4. Data Security: There are concerns with the safety and privacy of important data stored remotely. The possibility of private data commingling with other organizations makes some businesses uneasy. If you want to know more about those issues that govern data security and privacy, here is an interesting article on the recent privacy debates.

5. Software: If you want to be able to manipulate your files locally through multiple devices, you'll need to download the service on all devices.

3. DNA Storage

DNA storage is the process of encoding and decoding binary data onto and from synthesized strands of DNA (deoxyribonucleic acid). In nature, DNA molecules contain genetic blueprints for living cells and organisms. To store a binary digital file as DNA, the individual bits (binary digits) are converted from 1 and 0 to the letters A, C, G, and T. These letters represent the four main compounds in DNA: adenine, cytosine, guanine, and thymine. The physical storage medium is a synthesized DNA molecule containing these four compounds in a sequence corresponding to the order of the bits in the digital file. To recover the data, the sequence A, C, G, and T representing the DNA molecule is decoded back into the original sequence of bits 1 and 0. Researchers at the European Molecular Biology Laboratory (EMBL) have encoded audio, image, and text files into a synthesized DNA molecule about the size of a dust grain, and then successfully read the information from the DNA to recover the files, claiming 99.99 percent accuracy. An obvious advantage of DNA storage, should it ever become practical for everyday use, would be its ability to store massive quantities of data in media having small physical volume. Dr. Sriram Kosuri, a scientist at Harvard, believes that all the digital information currently existing in the world could reside in four grams of synthesized DNA. A less obvious, but perhaps more significant, advantage of DNA storage is its longevity. Because DNA molecules can survive for thousands of years, a digital archive encoded in this form could be recovered by people for many generations to come. This longevity might resolve the troubling prospect of our digital age being lost to history because of the relative impermanence of optical, magnetic, and electronic media. The principal disadvantages of DNA storage for practical use today are its slow encoding speed and high cost. The speed issue limits the technology's promise for archiving purposes in the near term, although eventually the speed may improve to the point where DNA storage can function effectively for general backup applications and perhaps even primary storage. As for the cost, Dr. Nick Goldman of the EMBL suggests that by the mid-2020s, expenses could come down to the point where the technology becomes commercially viable on a large scale.

To store data in DNA, the concept is the same, but the process is different. DNA molecules are long sequences of smaller molecules, called nucleotides – adenine, cytosine, thymine and guanine, usually designated as A, C, T and G. Rather than creating sequences of 0s and 1s, as in electronic media, DNA storage uses sequences of the nucleotides.

There are several ways to do this, but the general idea is to assign digital data patterns to DNA nucleotides. For instance, 00 could be equivalent to A, 01 to C, 10 to T and 11 to G. To store a picture, for example, we start with its encoding as a digital file, like a JPEG. That file is, in essence, a long string of 0s and 1s. Let's say the first eight bits of the file are 01111000; we break them into pairs – 01 11 10 00 – which correspond to C-G-T-A. That's the order in which we join the nucleotides to form a DNA strand.

Digital computer files can be quite large – even terabytes in size for large databases. But individual DNA strands have to be much shorter – holding only about 20 bytes each. That's because the longer a DNA strand is, the harder it is to build chemically. So we need to break the data into smaller chunks, and add to each an indicator of where in the sequence it falls. When it's time to read the DNA-stored information, that indicator will ensure all the chunks of data stay in their proper order.

3.1 Storing the data

After determining what order the letters should go in, the DNA sequences are manufactured letter by letter with chemical reactions. These reactions are driven by equipment that takes in bottles of A's, C's, G's and T's and mixes them in a liquid solution with other chemicals to control the reactions that specify the order of the physical DNA strands. This process brings us another benefit of DNA storage: backup copies. Rather than making one strand at a time, the chemical reactions make many identical strands at once, before going on to make many copies of the next strand in the series. Once the DNA strands are created, we need to protect them against damage from humidity and light. So we dry them out and put them in a container that keeps them cold and blocks water and light. But stored data are useful only if we can retrieve them later.

3.2 Reading the data back

To read the data back out of storage, we use a sequencing machine exactly like those used for analysis of genomic DNA in cells. This identifies the molecules, generating a letter sequence per molecule, which we then decode into a binary sequence of 0s and 1s in order. This process can destroy the DNA as it is read – but that's where those backup copies come into play: There are many copies of each sequence. And if the backup copies get depleted, it is easy to make duplicate copies to refill the storage – just as nature copies DNA all the time. At the moment, most DNA retrieval systems require reading all of the information stored in a particular container, even if we want only a small amount of it. This is like reading an entire hard drive's worth of information just to find one email message. We have developed techniques – based on well-studied biochemistry methods – that let us identify and read only the specific pieces of information a user needs to retrieve from DNA storage.

3.3 Advantages of DNA data storage

- DNAs can remain stable for very long (10,000s of years) without any special care which makes it a safer place to store digital information. Magnetic tapes used to archive digital data degrade after some years.
- DNA can maintain its integrity without any power supply. Also, its small size and weight make it easy to store and transport.
- DNA is less susceptible to technical failures.
- Digital storage media will soon become obsolete. The different storage media as we know—floppy disks, CDs, DVDs, portable hard drives, thumb drives, and cloud storage all have limited span of life. But as long as living things and biologists exist, someone shall be there to read DNA data.

3.4 Disadvantages of DNA data storage

- High cost of DNA synthesis per data stored (around US\$12,400 per megabyte of data stored).
- Data is read back at low speed.
- DNA is not rewritable, i.e., it can't update the information it holds without redoing the entire information storing process.
- DNA does not allow random access either, meaning, to access a particular part of the data stored, the entire stored information should be decoded.

4. Comparison

Below Table shows a comparison of the digital storage device and DNA storage.

Parameter	Digital Storage Device	DNA Storage
Basic Element	Transistor, Magnetic or Optical Domain	Nucleotide
Basic Element Size	1,000 / 10,000 nm ²	1 nm ³
Addressing Scheme and Organization	Discrete, Hierarchical	Sequential, Parallel Copying.
Redundancy	Multi-Layer	Data Replication
Capacity	Unlimited	0.1 GB / genome
Linear Density	10 kilobits / mm ²	1 Mbit / mm
Areal Density	1 gigabit / mm ²	-
Access Time	0.01 millisecond	-
Internal Data Transfer	10 megabits / sec	100 kilobits / sec
Error Rate	10 ⁻⁴	10 ⁻⁹

Table 1: Comparison

5. Conclusion

Thus, using DNA for data storage, it is possible to store huge amount of data in very less size. As DNA can retain data for millions of years, it is possible to store data for a long time. By using this technique, data is compressed and the security to the data is provided. Parallel reading of files is also possible enabling users to read multiple files at the same time. This technique maintains two copies of data. Hence in case of data damage, its copy can be used to read data. In the case of any errors while encoding the data, the error is restricted to that particular file and no other file is affected due to that error. This technique can be used for all kind of files by making minor changes to adapt to the type of file. This technique can be used to store big data in very small space with little computational overhead. This method is scalable and can be used to store large files too. Also multiple copies can be made easily. This method can be used to store information in archival systems or big data. Instead of using conventional storage devices which have less capacity to store data, DNA- based storage method be used in distant future to store data secured manner and for long time storage and solve the problem of limited space.

References

- [1] J. Gantz, D. Reinsel. Extracting value from chaos. International Data Corporation (IDC), Framingham, MA (2011), www.emc.com/collateral/analyst-reports/idc-extracting-value-fromchaos-ar.pdf.
- [2] C. Bancroft, T. Bowler, B. Bloom, C. T. Clelland. Long-Term Storage of Information in DNA. *Science*, 293, 1763 (2001).
- [3] George M. Church, Yuan Gao, Sriram Kosuri. Next-Generation Digital Information Storage in DNA. *Science*, 337, 1628 (2012).
- [4] Siddhant Shrivastava and Rohan Badlani. Data Storage in DNA. *International Journal of Electrical Energy*, Vol. 2, No. 2, June 2014.
- [5] Mohan S., Vinodh S. and Jeevan F. R. Preventing Data Loss by Storing Information in Bacterial DNA. *International Journal of Computer Applications* (0975 8887) Volume 69 No.19, May 2013.
- [6] Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* 328, 710722 (2010).
- [7] Willerslev, E. et al. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317, 111114 (2007).
- [8] Lilit Garibyan and Nidhi Avashia. Polymerase Chain Reaction. *Journal of Investigative Dermatology* (2013) 133, e6. doi:10.1038/jid.2013.1
- [9] Nick Goldman, Paul Bertone¹, Siyuan Chen, Christophe Dessimoz, Emily M. LeProust, Botond Sipos & Ewan Birney. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494, 7780 (2013).
- [10] Salunke Avinash N., Shruti Gupta, Varsha Agarwal, and Muhammad Rukunuddin Ghalib. A Novel Digital Information Data Storage Approach in DNA. *International Journal of Applied Engineering Research*, ISSN 0973-4562, Vol. 8, No. 19 (2013).
- [11] Mamta Sharma. Compression Using Huffman Coding. *IJCSNS International Journal of Computer Science and Network Security*, VOL.10 No.5, May 2010.
- [12] Nigam Sangwan. Text Encryption with Huffman Compression. *International Journal of Computer Applications* (0975 8887), Volume 54 No.6, September 2012.
- [13] Ailenberg, M. & Rotstein, O. D. An improved Huffman coding method for archiving text, images, and music characters in DNA. *Biotechniques* 47, 747754 (2009).
- [14] Mohan S, Vinodh S and Jeevan F R. Preventing Data Loss by Storing Information in Bacterial DNA. *International Journal of Computer Applications* 69(19):53-57, May 2013.
- [15] J. Reardon, D. A. Basin, and S. Capkun. Sok: Secure data deletion. In *IEEE Symposium on Security and Privacy*, pp. 301-315, 2013.
- [16] Yang Tang, Patrick P. C. Lee, John C. S. Lui, and Radia Perlman. FADE: Secure overlay cloud storage with Access Control and Assured Deletion. In *IEEE Computer Society*, vol. 9, no. 6, pp. 903-916, 2012.
- [17] G. Zhao, C. Rong and Jin Li, et al. Trusted Data Sharing over Untrusted Cloud Storage Providers. In *Proc. of the 2nd IEEE International Conference on Cloud Computing Technology and Science*. pp. 97-103, 2010.