IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-3 E-ISSN: 2347-2693

New Trends in Digital Data Storage for the Internet of Things

A. Saxena[1], S. Sharma[2], S. Dangi [3], A. Sharma[4], C. Patel[5]

[1,2,3,4,5] Dept. of Computer Science, Dev Sanskriti, University, Haridwar, India

[*]*Corresponding Author: shekharrockin1988@gmail.com, Tel.:7895510180*

Available online at: www.ijcseonline.org

Received: 24/Feb//2018, Revised: 03/Mar2018, Accepted: 23/Mar/2018, Published: 30/Mar/2018

***Abstract*—** As we know in every era, we need knowledge for performing the efficient job. That knowledge comes from past experiences & information, if we don't have proper storage for information then that information will go in vain after few days. Today we have a large amount of data for storing, for which we used a different type of devices. Based on our need we have developed a different type of technologies to store the data such as Cds, DVDs, Floppy Disk, Hard Disks, Flash drive.

As we know that technology is moving towards IoT and experts believe that IoT will consist 30 billion devices by 2020. So the physical world has more direct involvement in the computer world. Now our devices become intelligent systems that can share and analyze data. This analyzed data or information will change our business and daily life. Big possibilities come for analyzing the data across the system. For this, we need a better storage so that we can store a large amount of data easily & retrieve the data without any error. So in this paper we have discussed the technique for converting the information in 2.14*10^6 bytes in DNA oligos.

Keywords—DNA,IOT

I. INTRODUCTION

In this article, we are going to talk about something different and beyond the new inventions in the science field. It is the latest fashion of storing the information that is DNA digital storage. It is the unique combination of two field's bio with information technology. DNA is further going to store the digital information inside its structure. It is unbelievable that the DNA can store a million of gigabytes in 1 gram. This is another way to store the information and it is a more compact system of storing the information than the other systems like magnetic tapes and other storages we have been used. DNA is also the system which has a long life i.e. hundreds of thousands of years it can be if it stores at the cool and dry place. DNA can be live as long as from up to 60,000 years. This is given by the study of the woolly mammoth (an old elephant).

The journey is started to storing the information from the punch cards, magnetic tapes, gramophones, floppy discs and so on and after that, we moved on some new trends of storage such as CD's, DVD's, Blu-ray discs and flash drives. Now we have taken a diversion from the journey of punch cards, magnetic tapes, floppy discs and other hard drives to DNA digital information storage. This DNA is made by using commercially available oligonucleotides synthesis machines for storing the information, sequencing and also used for retrieve the information. But the DNA synthesis is more expensive and laborious and it's more difficult to read the information from the DNA.

In figure 1 we have shown the structure of DNA. Further, we discussed the structure of DNA, which have two strands. On which information stores.

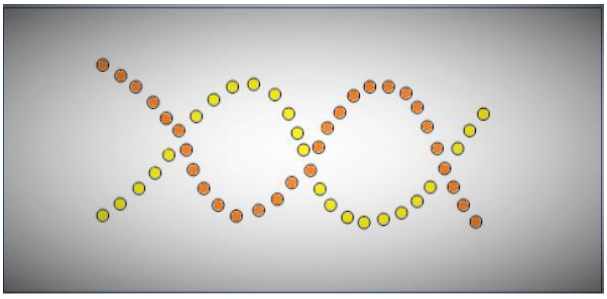

Figure 1: Structure of DNA

DNA is a programming language of our genetic codes. Which actually encodes the information. Each and every day we produced approximately 25 billion gigabytes and DNA stores all this large amount of information in a very small space. So we found the whole information is produced by the world over a year can be stored in only 4 gram of DNA.

Fig 2: DNA storing capacity in MBs

In figure 2, we have shown DNA storing capacity in MBs and found that billions of gigabytes we can store in it. Experimentally to store the data into DNA,we convert our binary data file into base 4 to match them with four building blocks of DNA. Then we included ID tags that allowed them to access any byte within the large pool of data. DNA. DNA basically stands for Deoxyribonucleic acid.

II. Related Work

The bioscientists and leader of the scientist's group Nick Goldmen said that there are some problems to adopting the DNA storage and barriers can occur in DNA digital storage like an enormous amount of information, it takes a lot of time to synthesize the data and it is also difficult to maintain the sequence of that DNA because every day we produce approximately 25 billion gigabytes of data and which is really a very large amount of data to maintain and synthesize [1]. Bio-engineers and a group of scientists encoded the 4 large files like movie, audio, poems etc. The scientist and professor at Columbia university Yaniv Erlich with Dina Zielinski who is the associated scientist at NewYork Genome Centre encode the film of 1885 "Arrival of a train at La Ciotat", a $50 Amazon gift card, a computer virus, a pioneer plague [2]. After encoding the files they were able to access the files almost perfectly. They want to use enzymes to reconfigure the modules building the blocks into patterns that record specific information [3]. Robert Service (2 March 2017) talks about in his article that DNA is the ultracompact system. They talk about how much a DNA can contain the data. DNA contains the 215 million gigabytes in a single gram [4]. In science magazine, DNA could store all of the world's data in one room they published the researchers about the DNA occupies the minimum space [5]. Naik Gautam talks about the base pairs and sense of storing the information into DNA. He defines that how much important to store the information because digits are stored in nucleotides of DNA which is the basic thing set the information in DNA. In both sense such as encoding the data and also decoding the data [6]. In this invention some talks about the leading and logging end of DNA which adds the new existing nucleotides to make the single strand of DNA helix [7]. DNA solves the computation problems. We all urgently need the storage which is efficient and takes a little space. Because the conventional storages consume more space and they were slow. DNA does not produce the pollution rather the conventional storages pollute the environment.so it is another way to understand the advantages of the DNA in perspective of storage.

III. Findings

We know that a large amount of data we can store in DNA and the human DNA has approximately 3 billion base pairs as we have shown in figure 3.

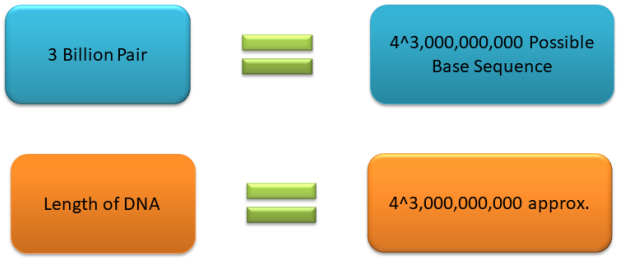

Figure 3: Length of Human DNA

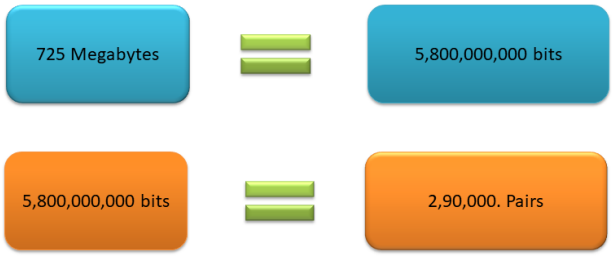

Figure 4: Human DNA Base Pair

In the above figure 4, we have describe that how a human DNA contains the bits on its base pairs. Each base pair contains 2 bits. According to that the number of bits shown in figure 4 is formed in 2, 90,000 pairs. This amount of data contains a single genome of DNA. Which can be compressed 1% and at worst scenario, it compressed maximum 4 megabytes.

The management of data into DNA is based on the building blocks of DNA which are called nucleotides. In figure 5 we have represented them by their different-different colours.

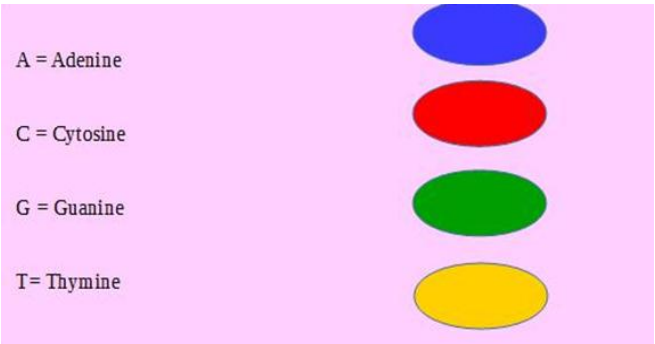

Fig 5:Image to show the colors of DNA nucleotides

In the figure we have seen that Adenine is indicated by the blue color, Cytosine is indicated by the red color, Guanine is indicated by the green color and the Thymine is indicated by the yellow color.

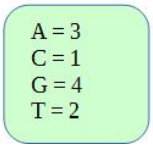

Fig 6: size of molecules

In figure 7 we have shown the size of molecules because in this architecture we did encoding on the basis of the size of molecules.

DNA contains the helix structure and contain 4 building blocks. It looks like one helix is wrapping another helix. Segment of DNA contain {AT,AT,AT,AT,CG,}.

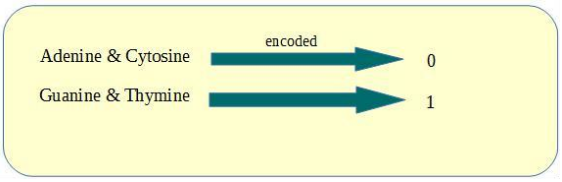

Figure 7: Nucleotides Encoded as Bits

In figure 7 we have shown the building blocks of Adenine and Cytosine are encoded as 0 while Guanine and Thymine are encoded as 1. Similarly we can encode them opposite as well.

In present, the data is executed in the form of binary digits 0's and 1's. But in DNA we store data on the basis of four building blocks A,C,G and T In this whole process we have one very important task is to set the binary digit for the DNA building blocks because without this we cannot connect them together and this leads to generate the fragmentation of the data. In DNA the four building blocks which are the basis of the data storage.

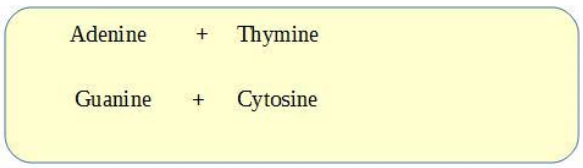

Figure 8: Base Pairs

In figure 8, we defined the pairing among the nucleotides of DNA and found that because of their chemical bonding Adenine makes a pair with Thymine and Guanine makes a pair with Cytosine. In figure 9 we have shown that Adenine and Thymine have 2 hydrogen bonds Guanine and Cytosine have 3 hydrogen bonds between them.

Fig 9: hydrogen bonds associated with base pairs

IV. METHODOLOGY

Base pairs of DNA are work on the complementary base pair matching. In this method, we have eight combinations on which we set the combination of 0's and 1's in the form of A, C, G, and T.

There is also one most important thing is the combination made among the pairs, according to these pairs we assign the set like A with T, C with G. We have a table of all these compliments which is set on the strands of the DNA helices. In figure 10 we have shown all the 8 base pair complimentary. This gives the clear intuition about the base pairs of nucleotides of DNA.

S.R no.	1	2	3	4	5	6	7	8
00	A	A	C	G	C	G	T	T
01	C	G	A	A	T	T	C	G
10	G	C	T	T	A	A	G	C
11	T	T	G	C	G	C	A	A

Figure 10: Base Pairs Matching with Symbols

The above figure is shown, the 8 encoding rules through which we can store the data into DNA. Which sets on the strands of DNA. There are two strands in a DNA. Which contain the massive amount of data in the perspective of future use.

In figure 11 we have shown the single-stranded DNA. Which is positioned in between the two ends of DNA.Where one end is 5' (5 prime) and another is 3' (3 prime). Because the synthesis of DNA is formed as 5' and 3' directions. The DNA polymerase is working only with the 5' and 3' ends.

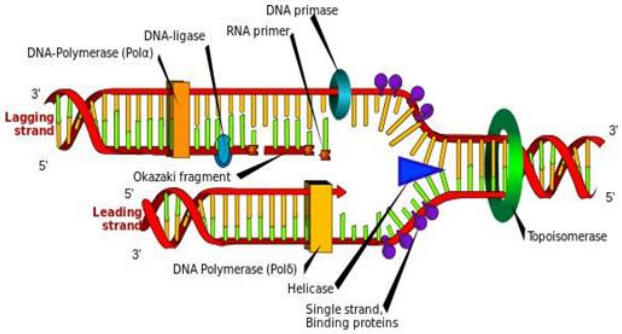

Figure 11:DNA replication

DNA polymerase requires the 3' OH end to add the incoming nucleotides these are added to the 3' OH ends to growing the one by one DNA strands by using the DNA polymerase so the

bonding becomes between the 3' OH end of the first nucleotide and 5' p end of the incoming nucleotide and the resultant polynucleotide chain is always synthesized as in 5'-3' direction. DNA replication is used to adding free nucleotides to the existing strand.

Now we are discussing the procedure for storing the data into DNA. It is thediagrammatical representation of this unique procedure.

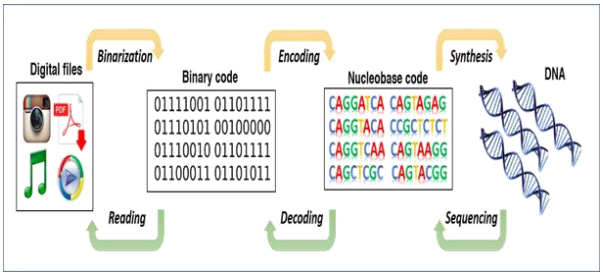

Figure 12: Procedure for Encoding the Data

In the above figure 12 we have shown the procedure of encoding the information into DNA. In this procedure first we take any file such as: text file, audio files, videos etc. we do the binary conversion of all these files in the form of binary digits 0 and 1, then resultant binary code is encoded into the nucleotides of DNA, after that the challenging process comes to synthesise the DNA, according to the information, then we sequenced the DNA which has been synthesized, the nucleotide code again then converted into the binary form and read the information. The output of this procedure is desired output.

It is difficult procedure to encode the information into DNA nucleotides. Because encoding is achieved by placing the nucleotides at each repeated positions. In this procedure the binary digit is converted into base 4 because the nucleotides of DNA contain the base 4. So here we store the binary code of files into the 4 building blocks of DNA. Further, we show the conversion of binary files into nucleotides.

Let's understand the procedure to convert the binary code into base 4 for example if we have the number 5, and the binary representation of the 5 is 101 , so how to 101 is converted as base 4? We can simply do that by adding 0 at the left position of the binary code because it does not affect and now it becomes 101 to 0101.

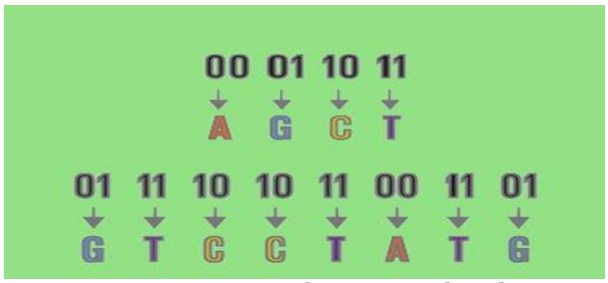

Figure 13: Data Encoding into Nucleotides

After that, we encoded our data into nucleotides. Like here A is set to be 00, C is set to be 10, G is set to be as 01 and T is as 11. For example if we have a binary file = 100111 conversion = CGT.

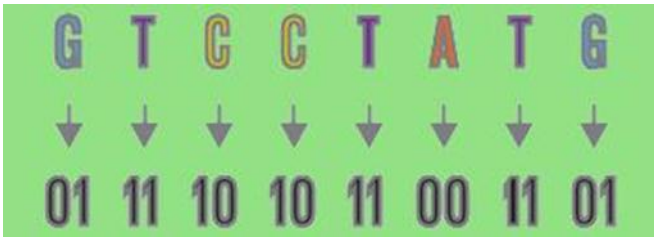

Figure 14: Decoding Information

In figure 19, we have shown the decoding of the stored information into nucleotides. Method of conversion is just opposite of encoding the information while storing the file. It is the method of retrieving the information from the base pairs or the double-stranded DNA.

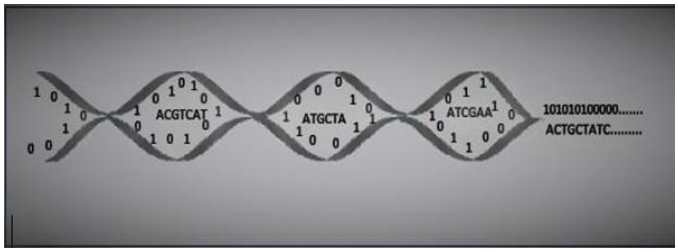

Figure 15: representation of information under the DNA

The above figure 15 shows that how the DNA seems after the information stored in it. The binary digits 0 and 1 looks like shown in figure when they set on the nucleotides of DNA.

V. CONCLUSION

In this review article, we analyse the methods of storing data into DNA. Where data is encrypted in the form of nucleotides. During our analysis, we also got that it is unbelievable that the DNA can store a million of gigabytes in 1 gram and found that this is another way to store the information and it is a more compact system of storing the information than the other systems like magnetic tapes and other storages we have been used. So this paper is also defined the new innovations in the field of information technology and biology. It is also a diversion to moving the old systems of storing the information.

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

