International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-3 Issue-2, July 2013

Use of DNA for Computation, Storage and Cryptography of Information

Piyush Saxena, Amarpal Singh, Sangeeta Lalwani

Abstract— ***DNA computing was proposed [1] as a method of solving a group of inflexible computational tribulations in which the computing time can grow up exponentially with respect to the problem size. A DNA can also be used as a next generation Digital Information Storage Medium that has tremendous storage capacity and low maintenance cost. This process of artificial manufacturing and decoding of DNA's can also be used to encode data by use of an extremely advanced and naturally existing cipher mechanism.***

Index Terms—***DNA Computing, DNA Cryptography, Logic gates, DNA chip, DNA Microprocessor.***

I. INTRODUCTION

DNA computing, in the literal sense, is the use of DNA (Deoxyribose Nucleic Acid) molecules, the molecules which encode genetic information for all living things, in computers. This is accomplished in a suspended solution of DNA, where certain combinations of DNA molecules are interpreted as a particular result to a computational problem encoded in the original molecules present. DNA computing is currently one of the fastest growing fields in both Computer Science and Biology [1], and its future looks extremely promising. A highly interdisciplinary study incorporating the research results of computer scientists and biologists. (Fig 1)

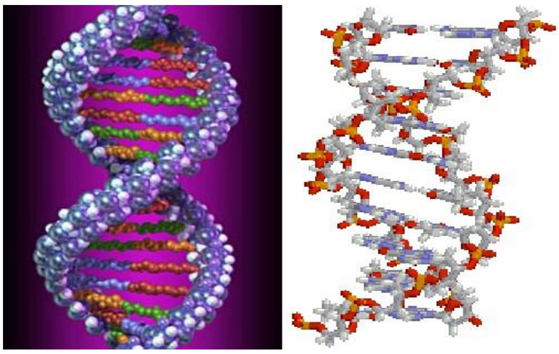

Fig 1: 3-D model of a DNA

Artificial Genes and DNA computing is proposed as a means of data storage and cryptographic data transportation for a very high level of security and classified data transmission that can be misused by the terrorist organizations for illegal activities, and this type of data needs to be protected with the state of art advanced security solutions.

DNA data computing and storage can also be used for large size data warehouses and for computing and solving scientific calculations and problems that is not possible to solve on existing methods of data storage. The speed of read and write on an artificial gene is quite high and the method to make and decode is extremely costly so can be used by limited number of people as it will need proper labs for it. This technique also involves the immobilization and manipulation of combinatorial mixtures of DNA on a support.

II. BACKGROUND

In the future, computers may weigh no more than 1.5 tons. Science made enormous development in efficiency since the days of room-sized computers, yet the fundamental computational framework has remained the same. The supercomputers used in date still use sequence of logics that were used by the mechanically working dinosaurs of 1930's. Several researchers are looking for new concepts and think beyond these boundaries. They are mainly investigating about entirely new media and computational models. These new media and computational models include quantum, optical and DNA-based computers. In this paper we mainly focus on last of these developments.

DNA computing was first implemented at the end of 1994, when Len Adleman of USC solved a small instance of a computationally intractable problem by means of a small vial of DNA. By representing information as sequences of bases in DNA molecules, Adleman discovered that how the existing DNA-manipulation techniques use to implement a simple, massively parallel random search.

A. DNA Fundamentals

DNA, Deoxyribonucleic Acid, is the molecular basis of heredity and localized especially in most cell nucleus. DNA molecules consist of two long chains held together by complementary base pairs. A DNA chain is a long, un-branched polymer composed of only four type subunits. These are the deoxyribonucleotides containing the bases adenine (A), cytosine(C), guanine (G), and thymine (T). The nucleotides are linked together by covalent phosphodiester bonds that join the 5' carbon of one deoxyribose group to the 3' carbon of the next. The four kinds of bases are attached to this repetitive sugar-phosphate chain. The two long chains of a DNA molecule are held together by complementary base pairs. Three hydrogen bonds form between G(Guanine) and C(Cytosine), and two hydrogen bonds exist between A(Adenine) and T(Thymine). The base pairing mechanism is the basis for DNA replication. (Fig 2)

Manuscript published on 30 July 2013.
*Correspondence Author(s)
Piyush Saxena, Master Of Technology (Computer Science and Engineering) Amity School of Engineering Technology, Amity University Uttar Pradesh, Noida, India.
Amarpal Singh, Master Of Technology (Computer Science and Engineering) Amity School of Engineering Technology, Amity University Uttar Pradesh, Noida, India.
Sangeeta Lalwani, Master Of Technology (Computer Science and Engineering) Amity School of Engineering Technology, Amity University Uttar Pradesh, Noida, India.

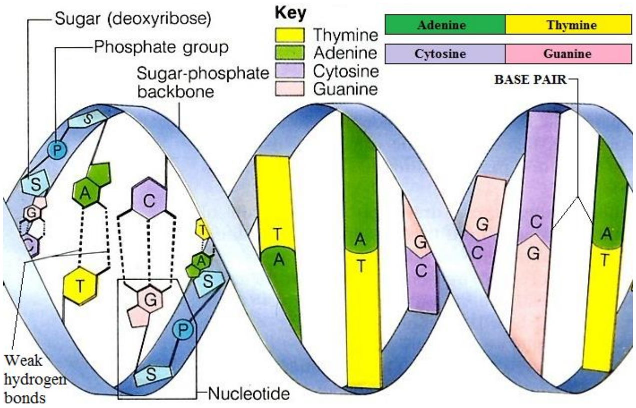

Fig 2: Structure of DNA

As a direct consequence of the base-pairing mechanism, it becomes evident that DNA carries information by means of the linear sequence of its nucleotides. Each nucleotide-A, C, T, or G – can be considered a letter in a four-letter alphabet that is used to write our biological messages in a linear "ticker-tape" form. Organisms differ because their respective DNA molecules carry different nucleotide sequences and therefore different biological message. Since the number of possible sequences in a DNA molecule, which is n nucleotides long, is 4n, the biological variety that could in principle be generated using even a modest length of DNA is enormous. A typical animal cell contains a meter of DNA ($3*10^9$ Nucleotides). Written in a linear alphabet of four letters, an unusually small human gene would occupy a quarter of a page of text, while the genetic information carried in a human cell would fill a book of more than 500,000 pages.

B. Logic Gates of DNA Computing

- In a DNA computer the basic units of data interpretation are logic gates made up of DNA strands.
- In DNA logic gate inputs are replaced by single stranded DNA molecule. The way by which these molecules bind dictates the operation.
- Chemical binding of two DNA strands in end-end operation illustrates the operation.
- DNA ligase seals the gap between any two input strands to give single strands to give a single strand which is output.

Logic gates are physical devices that implement the Boolean functions and convert the binary data into a series of signals that a computer uses to perform its operations.

DNA logic gates are small DNA dispensation centers that have capability to identify definite fragments of the genetic code as input, and then fix together these fragments into a single output. A DNA ligase seals the space between the ends of the two input strands, yields a single new strand [2].

Using regular gel electrophoresis, the length of this new strand can be accurately calculated, providing the DNA computer's output to the two input strands.

III. DNA COMPUTATION

DNA itself does not carry out any computation. It rather acts as a massive memory. But, the way complementary bases react with each other can be used to compute things.

Proposed by Adelman in 1994 "It is the Year 2020. Just imagine a scenario in which your state of the art computer has a problem and you call for tech support. Who turns up at the front door? Your friendly neighborhood biologist! If you think that is not possible - think again, because scientists are hard at work at this very moment trying to use DNA to power the computers of the future.

The race to build a supercomputer faster than the silicon based supercomputers are already on and the prize for the winner is beyond mere riches. It is a computer so powerful it can simulate the most complex and mysterious aspects of the universe, and that too without breaking into a sweat, it should also leave the best computers of today looking like ageing horses.

IV. DIGITAL INFORMATION STORAGE IN DNA

As digital information continues to gather, higher density and longer-term storage solutions are necessary [3]. DNA has many essential advantages as a way for immutable, high latency information storage needs [4]. (See fig 3) For instance, DNA storage is very intense. DNA has ability encode 2 bits/nucleotide (nt) or 455 Exabyte/gram of single-stranded DNA [5]. Contrasting a large amount of digital storage media, DNA storage is not limited to a planar layer and is often readable despite degradation in non-ideal conditions over millennia [6]. Lastly, the main biological role of DNA is to provide the access to natural reading and writing enzymes and ensures that DNA will remain a readable standard for the foreseeable future.

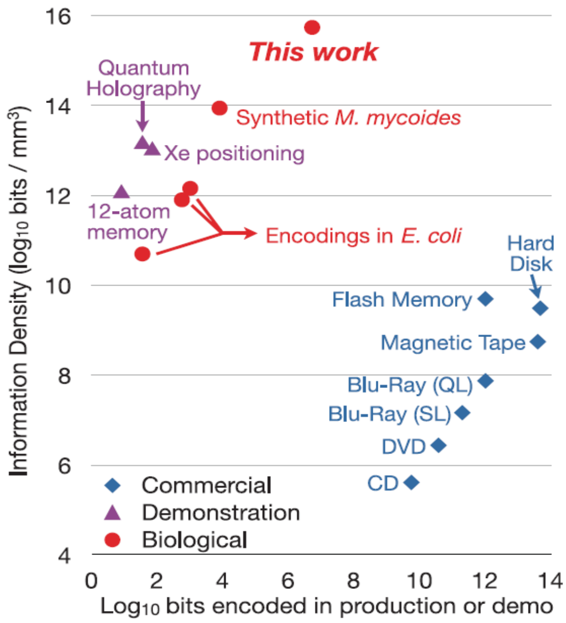

Fig 3: Comparison to other technologies. Graph plotted information density versus current scalability.

The previous work that was of a small scale gives rise to the difficulty in reading and writing long and faultless DNA sequences and has a small spectrum of applications. This issue led to the development of an approach to encode digital information by using encoding schemes like base-2 and ascii encodings with the use of next-generation DNA synthesis and sequencing techniques.

The following new method has some advantages over the previous DNA storage approaches.

1) This new method allows us to encode one bit per base (A or C for zero, G or T for one), instead of two and also enable to encode messages many ways in order to avoid sequences that are difficult to read or write.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

2) It splits the bit stream into addressed data blocks, so that we eliminate the need for long DNA constructs that are complex to assemble at this scale.
3) In order to keep away from cloning and sequence verifying constructs, the manufacture and use of stored and sequenced copies of each individual oligo was done. As errors in synthesis and sequencing are not often coincident, thus each molecular copy corrects errors in the other copies.
4) The Use of a purely in vitro approach helps us to avoid cloning and stability issues of in vivo approaches.
5) The use of next-generation technologies in both DNA synthesis and sequencing to allow for encoding and decoding of large amounts of information for ~100,000-fold cost less than first-generation encodings.

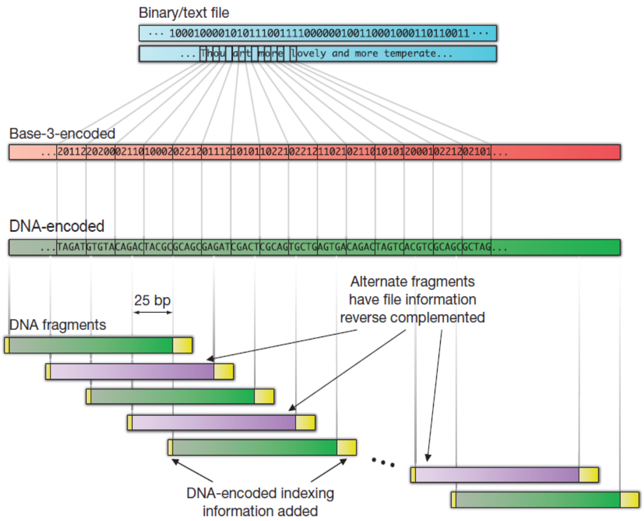

Fig 4: Digital information encoding in DNA.

V. DENSE INFORMATION STORAGE

- 1 gram of DNA can hold about 1*10^14 MB of data.
- The number of CD's required to hold this information, lined up edge to edge, would circle the earth 375 times, and would take 163,000 centuries to listen to.
- One gram of DNA can store 700 terabytes of data. That's 14,000 50-gigabyte Blu-ray discs… in a droplet of DNA that would fit on the tip of your finger.
- To store the same kind of data on hard drives — the densest storage medium in use today — one would need 233 pieces of 3TB drives, weighing a total of 151 kilos.

VI. OPERATIONS THAT CAN BE PERFORMED WITH A DNA

- *Synthesis* of a desired strand.
- *Separation* of strands by length.
- *Merging*: pour two test tubes into one to perform union.
- *Extraction*: extract those strands having a given pattern.
- *Melting/Annealing*: break/bond two ssDNA molecules with complementary sequences.
- *Amplification*: use PCR to make copies of DNA strands.
- *Cutting*: cut DNA with restriction enzymes.
- *Ligation*: Ligate DNA strands with complementary sticky ends using ligase.
- *Detection*: Confirm presence/absence of DNA in a given test tube.

The above operations can be used to "program" a DNA computer [7].

Example of computation

Fig 5 shows graduated PCR of the final product of the computer. Graduated PCR allows one to "print" the results of a computation, using A as the right primer and ~B through ~G as the left primer lanes 1-7, respectively. The graduation shows the progression from A>B>C>D>E>F>G.

Fig 5: Procedure to compute DNA

VII. HOW DNA COMPUTERS WILL WORK

Today the microprocessor and computer chip manufacturers are in a race to develop the next generation IC's that will topple the speed records, but this race will soon hit the boundary of limits of speed and miniaturization. To avoid this chip manufacturer need to find a new material that will produce faster computing speeds. This new material was found out to be a natural DNA that the scientists claim would be our next-gen material for making of IC's instead of silicon used currently. There are millions of natural supercomputers that exist inside living organisms. DNA (deoxyribonucleic acid) molecules, the complex chemical compound the genes are made of, have the prospective to carry out calculations many times faster as compared to the world's fastest and powerful man made computer [8]. DNA will be incorporated in a computer chip to make a biochip that will speed up the computers even faster. DNA molecules have already been harnessed to compute complex mathematical problems.

VIII. DNA MICROPROCESSORS

Silicon microprocessors captured the market of the computing world more than 40 years. During that era, manufacturers have packed full more and more electronic devices onto their microprocessors. According to Moore's Law [9], the number of electronic devices put on a microprocessor has doubled every 18 months. Moore's Law is named after Intel founder Gordon Moore, who predicted in 1965 that in every two years the complexity of microprocessors would double. Many have even predicted that Moore's Law will soon reach its end, because of the physical speed and miniaturization limitations of silicon microprocessors.

DNA computers have the latent to receive computing to fresh levels, selecting up where Moore's Law leaves off.

There are number of advantages of DNA in comparison with silicon as follow:

- There will always be a supply of DNA, as long as there are cellular organisms.
- The huge contribute of DNA makes it an inexpensive store.
- DNA are bio- chips with are the toxic materials that can use to make traditional microprocessors.
- DNA computers are much more convenient than today's computers.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

DNA's computers are smaller than any computer that has come before them, while at the same time holding more data. One pound of DNA has the ability to accumulate supplementary information than all the electronic computers ever constructed; and the computing power of a scratch drop sized.

DNA logic gates, is more powerful than the world's most powerful supercomputer. More than 10 trillion DNA molecules can accommodate into an area no larger than 1 cubic centimeter (0.06 cubic inches). With this tiny amount of DNA, a computer would be able to store 10 TB of data; along with it can perform 10 trillion calculations at a time. By accumulating more DNA, more calculations could be done.

Contrasting traditional computers, DNA computers execute calculations parallel to extra calculations. Conventional computers work linearly, captivating on tasks one at a time. The use of parallel compute allows DNA to analyze and clarify compound math's problems in limited time of hours, where as it would take multiple electric computers some hundreds of years to complete them.

The earliest DNA computers are improbable to characteristic about word processing, e-mailing and solitaire programs. As a replacement for, their influential computing power will be used by national governments for cracking secret codes, or by airlines to map more efficient routes. Studying DNA computers may also guide us to an improved understanding of a more complex computer i.e. the human brain.

IX. DNA CRYPTOGRAPHY

- RSA and DES are the two algorithm used in cryptography. Out of these DES has been broken using DNA computers.
- Tremendous parallel processing and enormous data capabilities are the key to breaking cryptographic algorithm.
- Generate all possible 64 bit keys (DES) using DNA memory strands. For each of 24^64 possible keys, compute the cipher text using the current key.
- The use of a defined strength of phosphoric acid for the manufacture of artificial gene for data storage will also need the exact strength of the chemicals during decoding or the gene will be destroyed or corrupted.

X. APPLICATIONS

A. DNA Integrated Circuit Chip

DNA chip is the massive parallelism to enable to simultaneously detect the expressions for a large number of genes. This is done by using a simple technology of the hybridizations to complementary DNA strands bonded to a glass surface in an array format. Nevertheless, DNA chip technology has a much potential for various applications including gene discovery and disease diagnosis.

In DNA computing researches, some methods have been developed to represent and evaluate Boolean functions on DNA strands. By employing this evaluation methods, we are able to deal with logical operations such as logical-"and" and logical- "or" for gene expressions on DNA strands.

The orthonormal sequences have uniform melting temperature and no folding potential to minimize computational error in DNA computing. DCN-encoded genome information is then analyzed with a power of the massive parallelism of DNA computing.

B. Genetic Programming

Genetic Programming applies GAs to a \population" of programs - typically encoded as tree-structures. Trial programs are evaluated against a _tness function" and the best solutions selected for modification and re-evaluation. This modification-evaluation cycle is repeated until a \correct" program is produced. GP has demonstrated its potential by evolving simple programs for medical signal alters, classifying news stories, performing optical character recognition, and for target identification.

C. Pharmaceutical Applications

Pharmaceutical Applications deals with emerging new technologies for developing customized solutions for drug delivery systems. The drug delivery systems should positively impact the rate of absorption, distribution, metabolism, and excretion of the drug or other related chemical substances in the body[10]. In addition, the drug delivery system should allow the drug to bind to its target receptor and influence that receptor's signaling and activity. Drug delivery materials should be compatible, easy to bind with a particular drug, and able to degrade into fragments after use that are either metabolized or driven out via normal excretory routes.

XI. SCALING PROPERTIES AND ROBUSTNESS OF DNA-BASED STORAGE

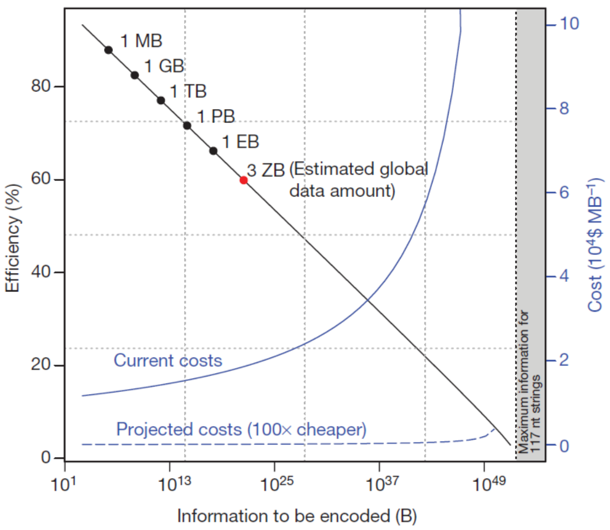

Fig 6: Encoding efficiency and costs change as the amount of stored information increases. [11]

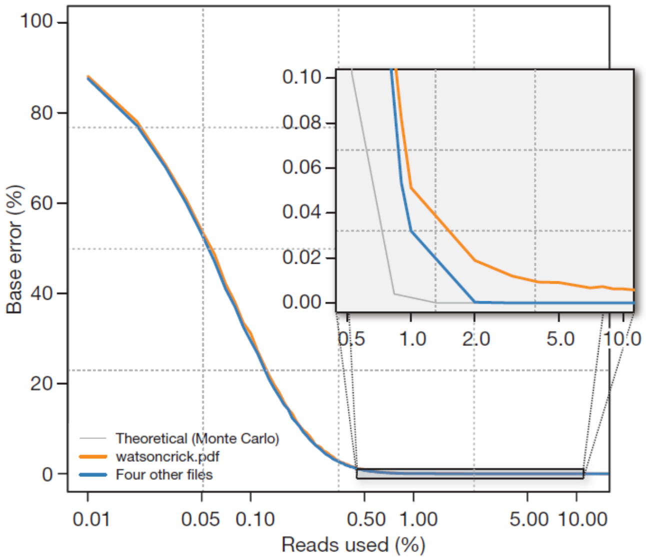

Fig 7: Per-recovered-base error rate (y axis) as a function of sequencing coverage, represented by the percentage of the original 79.63106 read-pairs sampled [11]

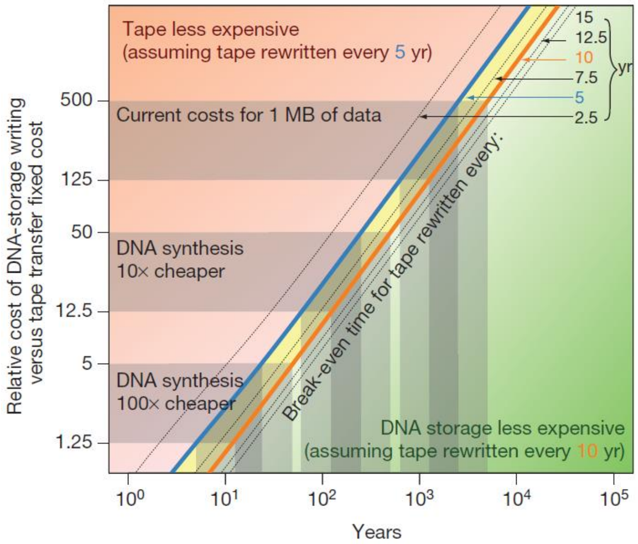

Fig 8: Timescales for which DNA-based storage is cost-effective. [11]

XII. CONCLUSION

In this paper, DNA is proposed as a method of computing, storage, cryptography and their applications as DNA chip and processors, pharmacy and genetic programming. DNA computational approaches to the solution of other problems have also been investigated. One technique involves the immobilization and manipulation of combinatorial mixtures of DNA on a support.

XIII. FUTURE SCOPE

Some centers of research in this area are developing new branches in this young field. Advancements are being made in cryptography. Researchers are working on decreasing error in and damage to the DNA during the computations/reactions. There are models for universal DNA computers, while others have described methods for doing addition and matrix multiplication with these computers. The field of DNA computing is truly exciting for the revolution it implies will occur within the next few years.

In the next paper we would emphasize on the methodologies of storage of digital information on DNA and how can it be managed to store huge amounts of data on the small quantity of DNA.

REFERENCES

1. Lovgren, Stefan (2003-02-24). "Computer Made from DNA and Enzymes". National Geographic. Retrieved 2009-11-26.
2. Tojanovic, Milan N.; Mitchell, Tiffany Elizabeth; Stefanovic, Darko (2002). "Deoxyribozyme-Based Logic Gates". Journal of the American Chemical Society 124: 3555–3561.doi:10.1021/ja016756v.
3. J. Gantz, D. Reinsel, “Extracting value from chaos” [International Data Corporation (IDC), Framingham, MA, 2011], www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf.
4. C. Bancroft, T. Bowler, B. Bloom, C. T. Clelland, Science 293, 1763 (2001).Information on materials and methods is available on Science Online.
5. Pääbo et al., Annu. Rev. Genet. 38, 645 (2004).
6. J. Bonnet et al., Nucleic Acids Res. 38, 1531 (2010).
7. Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. (2006). "Enzyme-Free Nucleic Acid LogicCircuits". Science 314 (5805): 1585. doi:10.1126/science.1132493. PMID 17158324.
8. http://www.elook.org/computing/dna-computing.htm
9. http://blogs.discovermagazine.com/80beats/2011/06/03/dna-computerdoes-math-plus-lays-out-building-blocks-for-bigger-circuits/.
10. Gareth L. Bond, Wenwei Hu, and Arnold J. Levine (2005). "MDM2 is a Central Node in the p53 Pathway: 12 Years and Counting". Current Cancer Drug Targets 5 (1): 3–8.
11. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA Nick Goldman1, Paul Bertone1, Siyuan Chen2, Christophe Dessimoz1, Emily M. LeProust2, Botond Sipos1 & Ewan Birney1 doi:10.1038/nature11875, Jan 2013 Macmillan Publishers Limited.

AUTHOR PROFILE

Piyush Saxena Pursuing Master of Technology in Computer Science and Engineering from Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India, Area of Interest: Cloud Computing, Data Mining and Warehousing and Soft Computing.

Amarpal Singh Pursuing Master of Technology in Computer Science and Engineering from Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India, Area of Interest: Cloud Computing, Software Engineering and Soft Computing.

Sangeeta Lalwani Pursuing Master of Technology in Computer Science and Engineering from Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India, Area of Interest: Cloud Computing, Java Programming.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)