Aug - Dec 2020

TRINITY INSTITUTE OF PROFESSIONAL STUDIES
Dwarka, Sector-9, New Delhi

Trinity Tech Review

Vol 6, Issue 2

Disclaimer: The views and opinions presented in the articles, case studies, research work and other contributions published in Trinity Tech Review (TTR) are solely attributable to the authors of respective contributions. If these are contradictory to any particular person or entity, TTR shall not be liable for the present opinions, inadequacy of the information, any mistakes or inaccuracies.

Trinity Institute of Professional Studies
An ISO 9001:2008 Certified Institution
(Affiliated to Guru Gobind Singh Indraprastha University, Delhi)

Sector-9, Dwarka, New Delhi-110075

Ph: 45636921/22/23/24, Telefax : 45636925
www.tips.edu.in, tips@tips.edu.in

TRINITY INSTITUTE OF PROFESSIONAL STUDIES
Sector-9, Dwarka Institutional Area, New Delhi-110075, Tel: 011-45636921/22/23/24
Certified as "A" Grade Institution by SFRC, Govt. of NCT of Delhi
NAAC Accredited "B++" Grade Institution
ISO – 9001:2008 Certified
Affiliated to GGSIP University

STATEMENT ABOUT OWNERSHIP AND OTHER DETAILS OF TTR/TMR
FORM 5 (RULE 8)

1.	Printer's Name	:	Dr. R.K. Tandon
	Nationality	:	Indian
	Address	:	Trinity Institute of Professional Studies Sector-9, Dwarka, New Delhi 110075
2.	Place of Publication	:	Delhi
3.	Periodicity of Publication	:	Quarterly
4.	Publisher's Name	:	Dr. R.K. Tandon
	Nationality	:	Indian
	Address	:	Trinity Institute of Professional Studies Sector-9, Dwarka, New Delhi 110075
5.	Editor's Name	:	Dr. Barkha Bahl
	Nationality	:	Indian
	Address	:	Trinity Institute of Professional Studies Sector-9, Dwarka, New Delhi 110075
6.	Name and Address of the individual who owns the journal and partners or shareholders holding more than one per cent of the capital.	:	CHAIRMAN Trinity Institute of Professional Studies Sector-9, Dwarka, New Delhi 110075
7.	Hosted at (url)		:www.tips.edu.in

I, Dr. R.K. Tandon, hereby declare that the particulars given above are true to the best of my knowledge and belief.

Dr. R.K. Tandon

MATHEMATICS IN CYBERSECURITY

Dr Aparna Chaturvedi
Associate Professor (Mathematics)
TIPS, Dwarka

Abstract:
Cybersecurity is a technical field and one that at its core, requires strong quantitative skills. The article is all about how mathematics is used in cybersecurity. Most cybersecurity studies generally require basic mathematics concepts that are used in binary, cryptography, or some minimal programming tasks.

While there are many approaches to cybersecurity, it is common for those approaches to be somewhat ad hoc or subjective. Cybersecurity needs a rigorous mathematical approach, which can be applied to address security issues, evaluate security controls, and investigating security breaches. The article maps the use of mathematical tools for cybersecurity purposes.

KEYWORDS: Cybersecurity, Cryptographic methods, Data and information security, Fault-tolerant computing, Network security, Privacy and anonymity, Software safety, System security.

CYBERSECURITY AS A SCIENCE:

The nearly global ubiquitous use of computers in every aspect of life makes understanding the behind-the-screens technology at once easy to ignore and difficult to understand. In the main, if the desktop, laptop, tablet, or mobile device does what we expect it to do, we give little thought to the bits and bytes that scurry behind the screen to make it operate. On the occasion that we find ourselves contemplating what magic makes these devices so incredibly powerful, we, of necessity, metaphorically throw our hands up in exclamation that there is just too much technology crammed into our electronics for any one person to grasp. [1, 2, 5, 6]

If that is how you feel, you are not alone, and you are not wrong. There is too much technology in our computing and communication devices for any one person to understand it all. It takes teams of experts in many fields, working in concert to conceptualize, design, manufacture, program, configure, protect, and deploy each piece of technology that we take for granted. The common denominator for these experts is that they each must be proficient in the core academic disciplines of science, technology, engineering, and mathematics (STEM). [3,7,8,9,15,16]

While all STEM disciplines require a good deal of mathematics, here we will focus on math as it is needed to be successful in the general field of computer science and, more specifically, cybersecurity.

Often people paint themselves and others with too broad a brush and declare they are either creative or logical. Mathematical

aptitude is generally attributed to logical or methodical thinkers. While this is often true, the ability to apply reason consistently does not preclude the ability to be creative.

The creative mind can express itself using mathematical equations in a most decidedly artful form. Rather than letting either of these labels deter you from pursuing STEM fields, consider your relationship with numbers instead. How you feel about using numbers may be a better barometer of how well you will adapt to STEM fields. [4,10,11,12, 17,18, 25]

Ask yourself if you enjoy working with numbers and using them to convey concepts and ideas. If you do, and you can think analytically with a focus on details, you may have the natural inclination towards numbers.

If you enjoy numbers, you are likely well suited for fields that require an understanding of math. If you also enjoy complex puzzles and helping others, you are probably well suited for work in the field of cybersecurity.

Math plays an essential role in areas. From science, to finance, to communications, many knowledge-based professions require excellence and aptitude in mathematics and quantitative reasoning. These careers also emphasize logical problem solving, critical thinking, and decision making. These are skills honed through the study of math. To gain a general understanding of your relationship with numbers, consider the following traits, skills, and abilities.

Traits, skills, and abilities of "lovers of numbers" include:

- An ability to achieve goals by constructing a path of reason back from the desired result to the current state of an issue or to reverse engineer a problem to find a solution
- An ability to quickly visualize abstract concepts, quantitative relationships, and spatial connections
- An ability to understand, communicate, and model using symbols and numbers
- An ability to think analytically and offer or receive criticism of ideas and concepts without involving feelings and emotions
- An ability to identify and categorize patterns and relationships
- An ability to use numbers as justifications to confidently take risks
- An ability to track and follow details and work with precision
- An ability to display patience as large complex problems are worked out

It is not necessary to be a "lover of numbers" to be successful in cybersecurity, but the higher the number traits, skills, and abilities listed above that you can claim as yours, the more likely you are to enjoy a numbers-based skills.

HOW MATHEMATICS IS USED IN CYBERSECURITY:

Cybersecurity is not a math-intensive area. That is not wrong to say, however, that familiarity and comfort with math will be hugely beneficial for success in cybersecurity. To be advance beyond an entry level cybersecurity position, a candidate should be comfortable with basic school level math, at least.

Whether expressed as (threat x vulnerability) or (probability x loss) or in some other more sophisticated fashion, determining risk is a mathematical exercise. At some level, all security professionals are in the risk calculation business. For many security workers, this calculation is performed almost subconsciously many times each day in the execution of their duties. Knowing what is essential and where to spend time and resources for the most significant result is the essence of the ability to understand risk. [13, 14, 19, 20, 21,22, 23, 24]

If on the front lines of a Security Operations Centre (SOC), a security specialist can be flooded with security alerts. They must analyse these alerts and make a quick risk assessment to know what they can handle now and what must be escalated for further investigation. This can be overwhelming at times and requires an ability to calculate risk very quickly.

A security code auditor will find herself examining code written by other coders. While many analytical tools are available to assist, she must be able, at a glance, to recognize weaknesses and vulnerabilities in the code. Writing and understanding computer software code requires mathematical skills.

Binary math is how computer operations are computed. It is used in everything from establishing IP addresses to network routing. The word binary means composed of, or involving two things. A binary number is made up of bits, each having a value of 0 or 1. A bit (short for binary digit) is the smallest unit of data in a computer. Computers generally store data and execute instructions in bit multiples called bytes. In most computer systems, there are eight bits in a byte.

Every number in your computer is an electrical signal, and when these machines were initially designed, electrical signals were difficult to precisely measure and control. It made more sense to only distinguish between an "on" state represented by negative charge and an "off" state represented by a positive charge. Thus today, binary math is at the heart of all computer machine language and software.

Another math-based concept used in cybersecurity is hexadecimal math. Rather than having only two options, as in binary math, hexadecimal math is based on the idea that you can count up to any one of 16 different options. You count these options from 0 to 15, providing sixteen total choices. Since one-digit numbers only range from 0 to a 9 (10 takes up two digits), you have to represent everything from 10 up to 15 as something else, in this case, using the letters A through F. [26, 27, 28, 35, 36, 37, 38]

Entry-level cybersecurity jobs will require at least some understanding of computer coding or programming. Computer code is written with math as its foundation. Coders need to understand programming concepts like constraints, variables, and programming logic. For example, you would be required to understand a basic computer code like this elementary if-else statement:

```
var x = 1; if (x === 1) {

  window.alert ("The expression is true!");
```

```
}

else {

 window.alert ("The expression is false!");

}
```

The above is a simple example of a computer code. Still, from this, you can see that you'll need to have an understanding of mathematical logic and how a computer will interpret information.

Boolean algebra has been fundamental in the development of digital electronics. Although first introduced by George Boole in his book The Mathematical Analysis of Logic in 1847, Boolean algebra is applied in modern programming languages. Whereas in elementary algebra, expressions indicate mainly numbers, in Boolean algebra, they signify the values false and true. It deals with operations on logical values and incorporates binary variables of 0 and 1.

Cryptography is the science of codes and encryption and is based on mathematical theory. Cryptographic techniques are at the very heart of information security and data confidentiality. The math used in cryptography can range from the very basic to highly advanced. Cryptographic algorithms are composed around computational hardness assumptions. A computational hardness assumption is a hypothesis that a particular problem cannot be solved efficiently, making such algorithms hard to break in practice by any adversary. They are also used by cyber-adversaries and are integral to ransomware. Cryptovirology is a domain that considers how to use cryptography to design robust malicious software. [29, 30, 31, 32, 33, 34, 39,40, 45, 47, 51]

In mathematics and computer science, an algorithm is a calculable pattern of clear, computer-implementable directions. They are used to solve problems or to complete computations. Algorithms are crucial to computer science and cybersecurity. They are used as blueprints for executing calculations, data processing, automated reasoning, and other tasks.

MATHEMATICS REQUIREMENTS FOR EDUCATION IN CYBERSECURITY:

Probably the most effective way to compare your math aptitude against the requirements in cybersecurity is to examine the math requirements for various programs in the field. If you have sufficient math skills or if you feel confident that you could complete them successfully, it would be an excellent indication that your interests and skills are a good match for a career in cybersecurity.

The requirements to use mathematics in cybersecurity work are so compelling that a adequate knowledge of mathematics, would be suitable for any level but for the most technical cybersecurity research positions or security related field, you must be proficient in mathematics.

As you search for the cybersecurity-related mathematics knowledge, look for the underlined words to guide your understanding of where math skills may be required. It is not practicable to list all the

mathematics requirements for all the prerequisite courses, but these samples will provide a reasonable understanding of what is generally needed.

Whether or not you decide to pursue a formal security related future, a professional cybersecurity certification will go a long way toward advancing your career. While there are many applicable fields to choose from:

• The core mathematics skills necessary for a career in IT security requires only arithmetic and calculating the risk formula

• Mathematics is also requires for IP/MAC addressing

•To be a Network Security professional; you requires mathematics for figuring out subnet information

•To learn about PC hardware, including how to configure BIOS, motherboard basics, and the various expansion slot types. It requires you to remember and use the equation for calculating the transfer rate of different memory types.

Gaining expertise and preparing for cybersecurity industry are precisely the two areas, where the level of math required for success in these.

A student should be confident with a good understanding of high school level algebra, geometry, and calculus.

Some dominant fields, which attracts students are as:

- Cryptographic methods
- Data and information security
- Fault-tolerant computing
- Network security
- Privacy and anonymity
- Software safety
- System security
- System Administration and Security.
- Covers the installation and configuration of mainstream operating systems
- Important network services
- Disaster recovery procedures
- Techniques for ensuring the security of the system.

- Computer Architecture
- Applied Cryptography
- Basic security issues in computer communication
- Classical cryptographic algorithms
- Symmetric-key cryptography
- Public-key cryptography
- Authentication
- Digital signatures
- Advanced System Security Design.
- Advanced topics in network and system security, including firewall design
- Network intrusion detection
- Tracking and prevention
- Virus detection
- Programming language and OS support for security
- Wireless network security

As you would expect, the mathematics requirements for the opportunity to specialize in cybersecurity are more stringent and demanding. This specialization encompasses that focus on technical issues related to safe software, languages, and architectures, as well as broader societal issues of privacy and

legal ramifications. [41, 42, 43, 44, 46, 48, 49, 50, 52, 53]

CONCLUSION:

Technology increases at break-neck speed. Year after year, computer-based technological advances have shaped and revolutionized how we interact with the world, a world that was inconceivable a few short decades ago. For many people, trying to find where they fit into this high-tech world can be a challenge. Attempting to match their interests and aptitudes to a future career can be confusing.

Many careers in technical fields require the use of math. The quickly growing field of cybersecurity is no exception. Entry-level careers require at least high-school level math and algebra, and highly technical security jobs require even more advanced math. There are, however, few security centric positions that require mathematics of advance level and above.

Don't let the labels of "creative person" or "analytical person" close doors unnecessarily. A love for drawing and art can be indicative of an ability to conceptualize complex ideas or a handy skill in computer science. Many successful people have learned to express their creativity through mathematics.

If you can write and understand computer code, you likely already possess the mathematics skills needed for all but the most technical cybersecurity roles. If you are a candidate for these highly specialized roles, you undoubtedly have already tested your aptitude and talent for mathematics in real-world experiences.

The best measure of how your math skills and aptitude align with technical security industry. In this article, I tried to discuss some examples of each. Review these examples and ask yourself if there is anything in your education, work history, or general interests that would qualify you for these areas. Truth to be told, the security industry needs you and will, in all likelihood, be happy to find a place for you.

BIBLIOGRAPHY:

[1]. Ahmadian, S., Tang, X., Malki, H. A., & Han, Z. (2019). Modelling cyber-attacks on electricity market using mathematical programming with equilibrium constraints. *IEEE Access*, *7*, 27376-27388. https://doi.org/10.1109/ACCESS.2019.2899293

[2]. Allodi, L., & Massacci, F. (2013). How CVSS is DOSsing your patching policy (and wasting your money). *BlackHat USA*.

[3]. Beynon-Davies, P. (2016). *Information Systems Development: an introduction to information systems engineering*. Macmillan International Higher Education.

[4]. Bollobás, B. (2013). *Graduate Texts in Mathematics: Modern graph theory*. Springer Science & Business Media.

[5]. Chokkalingam, B.,Raja, V.,Anburaj, J., Immanual, R., & Dhinesh kumar, M. (2017). Investigation of Shrinkage Defect in Castings by Quantitative Ishikawa Diagram. *Archives of*

Foundry Engineering, *17*(1), 174-178. https://doi.org/10.1515/afe-2017-0032

[6]. Clifton, E. (2020). A Brief Review on the Application of Lanchester's Models of Combat in Nonhuman Animals. *Ecological Psychology*, *32*(4), 181-191. https://doi.org/10.1080/ 10407413.2020.1846456

[7]. Dongre, S., Mishra, S., Romanowski, C., & Buddhadev, M. (2019). Quantifying the Costs of Data Breaches. In J. Staggs & S. Shenoi (Eds.), *Critical Infrastructure Protection XIII* (pp. 3-16). Springer, Cham. https://doi.org/10.1007/978-3-030-34647-8_1

[8]. Dupont, B. (2019). The ecology of cybercrime. In R. Leukfeldt & T. J. Holt (Eds.), *The human factor of cybercrime* (pp. 389-407). Routledge.

[9]. Easttom, C. (2018). A Systems Approach to Indicators of Compromise Utilizing Graph Theory. *2018 IEEE International Symposium on Technologies for Homeland Security,* 1-6. doi.org/10.1109/THS.2018.8574187

[10]. Easttom, C. (2019). *Incorporating Cybersecurity Engineering within the Discipline of Systems Engineering* [Master's thesis, University of Texas at El Paso]. Open Access Theses & Dissertations. Retrieved from https://scholarworks.utep.edu/open_etd/62/

[11]. Easttom, C. (2020). Mathematically Modeling Cyber-Attacks Utilizing Engineering Techniques. *15th International Conference on Cyber Warfare and Security (ICCWS).*

[12]. Easttom, C. (2021). Mathematically Modeling Victim Selection in Cybercrimes. *16th International Conference on Cyber Warfare and Security (ICCW)*.

[13]. Elsadany, A.A., Matouk, A.E. Dynamical Behaviors of Fractional-Order Lotka–Volterra Predator-Prey Model and its Discretization. *J. Appl. Math. Comput*. 49, 269–283 (2015). https://doi.org/10.1007/s12190-014-0838-6

[14]. Engel, A. (2010). *Verification, validation and testing of engineered systems*. John Wiley & Sons.

[15]. Fernald, D. G. (2020, January). US Army Software System Safety Process, Case-Study, and Success Stories. *2020 Annual Reliability and Maintainability Symposium (RAMS)*, 1-6. https://doi.org/10.1109/RAMS48030.2020.9153623

[16]. Feutrill, A., Ranathunga, D., Yarom, Y., & Roughan, M. (2018). The Effect of Common Vulnerability Scoring System Metrics on Vulnerability Exploit Delay. *2018 Sixth International Symposium on Computing and Networking (CANDAR)*, 1-10. https:// doi.org/10.1109/CANDAR.2018.00009

[17]. Franklin, B. D., Shebl, N. A., & Barber, N. (2012). Failure Mode and Effects Analysis: too Little for too Much? *BMJ Quality Safety*, *21*(7), 607-611. https://doi.org/10.1136/ bmjqs-2011-000723

[18]. Frigault, M., Wang, L., Jajodia, S., & Singhal, A. (2017). Measuring the Overall Network Security by Combining CVSS Scores Based on Attack Graphs and Bayesian Networks. In L. Wang, S. Jajodia & A. Singhal (Eds.), *Network Security Metrics* (pp. 1-23). Springer, Cham. https://doi.org/10.1007/978-3-319-66505-4_1

[19]. Gandal, N., Riordan, M. H., & Bublil, S. (2020). A New Approach to Quantifying, Reducing and Insuring Cyber Risk: Preliminary Analysis and Proposal for Further Research. *Centre for Economic Policy*

Research.
https://doi.org/10.2139/ssrn.3548380

[20]. Jamieson, W. T., & Reis, J. (2018). Global Behaviour for the Classical Nicholson–Bailey Model. *Journal of Mathematical Analysis and Applications*, *461*(1), 492-499. https:// doi.org/10.1016/j.jmaa.2017.12.071

[21]. Sarif Hassan, Sk., Ahluwalia, D., Maddali, R. K., & Manglik, M. (2018). Computational Dynamics of the Nicholson-Bailey models. *The European Physical Journal Plus*, 133(9), 349. https://doi.org/10.1140/epjp/i2018-12164-1

[22]. Holland, J. N., DeAngelis, D. L., & Bronstein, J. L. (2002). Population Dynamics and Mutualism: Functional Responses of Benefits and Costs. *The American Naturalist*, *159*(3), 231-244. https://doi.org/10.1086/338510

[23]. Holm, H., & Afridi, K. K. (2015). An Expert-Based Investigation of the Common Vulnerability Scoring System. *Computers & Security*, *53*, 18-30. https://doi.org/10.1016/j. cose.2015.04.012

[24]. Hyeon, C., & Aurelia, S. (2020, October). Enhancement of Efficiency of Military Cloud Computing using Lanchester Model. *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 960-964. https://doi. org/10.1109/I-SMAC49090.2020.9243515

[25]. Khan, R., McLaughlin, K., Laverty, D., & Sezer, S. (2017). STRIDE-Based Threat Modeling for CyberPphysical *Systems. 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 1-6. https://doi.org/10.1109/ISGTEurope.2017.8260283

[26]. Klipp, E., Liebermeister, W., Wierling, C., & Kowald, A. (2016). *Systems Biology: a Textbook* (2nd ed.). Wiley.

[27]. Kress, M. (2020). *Lanchester Models for Irregular Warfare. Mathematics*, *8*(5), 737. https:// doi.org/10.3390/math8050737

[28]. Liliana, L. (2016). A New Model of Ishikawa Diagram for Quality Assessment. *IOP Conference Series: Materials Science and Engineering*, 161. https://doi.org/10.1088/1757-899x/161/1/012099

[29]. Mazurczyk, W., Drobniak, S., & Moore, S. (2016). Towards a Systematic View on Cybersecurity Ecology. In B. Akhgar & B. Brewster (Eds.), *Combatting Cybercrime and Cyberterrorism* (pp. 17-37). Springer, Cham. https://doi.org/10.1007/978-3-319-38930-1_2

[30]. Naagas, M. A., & Palaoag, T. D. (2018). A Threat-Driven Approach to Modeling a Campus Network Security. *Proceedings of the 6th International Conference on Communications and Broadband Networking*, 6-12. https://doi.org/10.1145/3193092.3193096

[31]. Modarres, M., Kaminskiy, M. P., & Krivtsov, V. (2016). *Reliability Engineering and Risk Analysis: a Practical Guide* (3rd ed.). CRC press. https://doi.org/10.1201/9781315382425

[32]. Momeni, B., Xie, L., & Shou, W. (2017). Lotka-Volterra Pairwise Modeling Fails to Capture Diverse Pairwise Microbial Interactions. *ELife*, *6*. https://doi.org/10.7554/elife.25051

[33]. Nguyen, D. H., & Yin, G. (2017). Coexistence and Exclusion of Stochastic Competitive Lotka–Volterra Models. *Journal*

of Differential Equations, *262*(3), 1192-1225. https:// doi.org/10.1016/j.jde.2016.10.005

[34]. Sanfilippo, J., Abegaz, T., Payne, B., & Salimi, A. (2019). STRIDE-Based Threat Modeling for MySQL Databases. *Proceedings of the Future Technologies Conference,* 368-378. https:// doi.org/10.1007/978-3-030-32523-7_25

[35]. Singh, U. K., & Joshi, C. (2016). Quantitative Security Risk Evaluation Using CVSS Metrics by Estimation of Frequency and Maturity of Exploit. *Proceedings of the World Congress on Engineering and Computer Science*, 1, 170-175.

[36]. Suo, D., Renda, M. E., & Zhao, J. (2021). *Quantifying the Tradeoff Between Cybersecurity and Location Privacy.* arXiv. https://arxiv.org/abs/2105.01262

[37]. Vaidyanathan, S. (2015). Adaptive Biological Control of Generalized Lotka-Volterra Three- Species Biological System. *International Journal of PharmTech Research*, *8*(4), 622-631.

[38]. van den Hooven, C. (2020). Quantitative Risk Calculation in Cybersecurity: The Value of Quantifying Risk. *ISSA Journal*, *18*(10).

[39]. Wang, W., Yang, D., & Luo, Y. (2013). The Laplacian Polynomial and Kirchhoff Index of Graphs Derived from Regular Graphs. *Discrete Applied Mathematics*, *161*(18), 3063- 3071. https://doi.org/10.1016/j.dam.2013.06.010

[40]. Yaqoob, I., Hashem, I. A. T., Ahmed, A., Kazmi, S. A., & Hong, C. S. (2019). Internet of Things Forensics: Recent Advances, Taxonomy, Requirements, and Open Challenges. *Future Generation Computer Systems*, *92*, 265-275. https://doi.org/10.1016/j.future.2018.09.058

[41]. Wasson, C. S. (2015). *System Engineering Analysis,Design, and Development: Concepts, Principles, and Practices* (2nd ed.). John Wiley & Sons.

[42]. Babarinsa, O., & Kamarulhaili, H. (2017). On Determinant of Laplacian Matrix and Signless Laplacian Matrix of a Simple Graph.

[43]. In S. Arumugam, J. Bagga, L. Beineke, B. Panda (Eds.), *Theoretical Computer Science and Discrete Mathematic*s (pp. 212-217). Springer. https://doi.org/10.1007/978-3-319-64419-6_28

[44]. Birolini, A. (2017). *Reliability engineering: Theory and Practice* (8th ed.). Springer.
Deo, N. (2017). *Graph Theory with Applications to Engineering and Computer Science*. Dover

[45]. Fu, L., Song, W., Lv, W., & Lo, S. (2014). Simulation of Emotional Contagion Using Modified SIR Model: A Cellular Automaton Approach. *Physica A: Statistical Mechanics and its Applications*, *405*, 380-391. https://doi.org/10.1016/j.physa.2014.03.043

[46]. Godsil, C., & Royle, G. F. (2013). *Graduate Texts in Mathematics: Algebraic Graph Theory.* Springer Science & Business Media.

[47]. Gross, J. L., & Yellen, J. (2005). *Graph Theory and its Applications* (2nd ed.). CRC press.

[48]. Harko, T., Lobo, F. S., & Mak, M. K. (2014). Exact Analytical Solutions of the Susceptible-Infected-Recovered (SIR) Epidemic Model and of the SIR Model with Equal Death and Birth Rates. *Applied Mathematics and Computation*, *236*, 184-194. https://doi.org/10.1016/j.amc.2014.03.030

[49]. Kuddus, A., Rahman, A., Talukder, M. R., & Hoque, A. (2014). A Modified SIR Model to

Study on Physical Behaviour among Smallpox Infective Population in Bangladesh. American Journal of Mathematics and Statistics, *4*(5), 231-239.

[50]. Latino, M. A., Latino, R. J., & Latino, K. (2016). *Root Cause Analysis: Improving Performance for Bottom-Line Results* (4th ed.). CRC Press

[51]. Motzek, A., Möller, R., Lange, M., & Dubus, S. (2015). Probabilistic Mission Impact Assessment Based on Widespread Local Events. *Proceedings of the NATO IST-128 Workshop: Assessing Mission Impact of Cyberattacks*, 16-22.

[52]. Noel, S., Harley, E., Tam, K. H., Limiero, M., & Share, M. (2016). CyGraph: Graph-Based Analytics and Visualization for Cybersecurity.

[53]. In V. N. Gudivada, V. V. Raghavan, V. Govindaraju & C.R. Rao (Eds.), *Handbook of Statistics* (Vol. 35, pp. 117-167). Elsevier. https://doi.org/10.1016/bs.host.2016.07.001

[54]. Sahu, M. K., Ahirwar, M., & Shukla, P. K. (2015). Improved Malware Detection Technique Using Ensemble-Based Classifier and Graph Theory. *2015 IEEE International Conference on Computational Intelligence & Communication Technology*, 150-154. https://doi.org/10.1109/CICT.2015.147

Blockchain and Big Data: the match made in heavens

Dr Surabhi Shanker
Associate Professor (IT)
TIPS, Dwarka

Introduction

Blockchain and Big Data are amongst the evolving technologies that are high on many companies' plans. Both are anticipated to completely transform the way businesses and organizations are run in the upcoming years. Long-time developing in a separate way, at first sight one might assume that these technologies are mutually exclusive. But that idea is rapidly changing.

There are rising opportunities that distributed ledgers will help enterprises finally get to holds with Big Data, which thus far is hostile with a number of challenges. They are both commanding on their own, however when united they may bring a large number of opportunities.

Big Data is one of the fastest rising sectors in the world. Every business wishes to get visions into usage patterns of their consumers. Huge datasets are thereby analysed using advanced statistical models and data mining. These Big Data sets will

become even more ubiquitous over the coming years.

"*It's not the amount of data that's important, it's what companies do with the data that matters. Big data can be analysed for visions that lead to better conclusions and strategic business moves."*

"Data analytics has converted the key to corporate competitive gain because of its role in recognising emerging market trends. In turn, companies can use this information to make faster and better decisions that help them drive profitability".

The growth of Big Data has presented a swing of issues for both big businesses and everyday consumers. With the evolution in data good analytics is becoming all the more challenging. Some major difficulties to data management and analytics include so-called dirty data, inaccessible data, and privacy issues. And as Big Data increases in size and the web of connected devices detonates, it exposes more of companies' data to potential security ruptures.

With the initiation of Big Data, data quality administration is both more important and more challenging than ever. Companies that are dealing with large datasets should ensure that the data are fresh, protected and not been modified and come from an authentic source. They have to make sure that the latest version is co-ordinated among all of the data centres in real time. It should also be confirmed that these data are accessible. For most, however, the data silos are still a major issue and a full company-wide digital conversion is still more concept than reality.

What is Blockchain?

Blockchain is a disseminated and shared ledger technology in which all transactions are safely recorded, permitting any member in a business network to see and verify the transaction's validity. This accelerates anyone to alter information about the records in retrospect. Even if the same transaction is recorded over multiple, distributed database systems, the technology is still protected by design.

In the easiest terms, a block is considered as a record of a new transaction/input (such as crypto currency data, medical data, user data, financial information, etc.) As each block is accomplished, it is added to a chain, generating a blockchain. It is important to remember that all preceding information stored in the blocks cannot be accustomed, changed, or edited. The information contained in a blockchain is decentralized, meaning it is publicly available. But then the question arises - if the data is decentralized, how is it more protected? In many ways, that is the beauty of the Blockchain construct. This is because no single entity can regulate or verify the information contained in a blockchain. Many entities associated to the blockchain network must agree to the transaction for the information to pass through. In short, Blockchain is immutable, protected by cryptography, and trustless.

Perhaps most substantial development in IT over the past few years, blockchain has the power to change the way that the world approaches big data, with improved security and data quality.

What is Big Data?

Big data denotes to a huge and expanded digital content which is difficult to process using traditional data management tools and techniques.

Both the technologies, i.e. Blockchain and Big Data are the need of time for many organizations. Blockchain technology is highly in demand in various sectors of the

society that is continuously associated in the era of digitalized globalization, but Big Data, is the compilation of massive datasets. Both are anticipated to drastically transform the way businesses and organizations are run in the near future. Although since long-time both technologies evolving unconnectedly, many people might assume that these technologies are mutually exclusive. But that idea is speedily changing.

There are developing expectations that distributed records will assist enterprises with at last having the opportunity to handles with Big Data, which so far is battling with various difficulties. Although both the technologies are powerful on their own, however when united they may introduce a large number of opportunities. So we can say that blockchain and Big Data completes each other.

Three things that Blockchain renovates Big Data

There are three things where blockchain intersects big data analytics:

1. Decentralization

The main hindrance of assimilating big data analytics into an already existing infrastructure is the massive cost. Today, blockchain tools increase the approachability to data analytics tools by decentralizing the technology needed.

2. Data Sharing and Monetization

Data is the most important information in the modern world, and combination both blockchain and big data can develop the way data analytics is shared and monetized. By this, customers can gain negotiation powers over businesses, providing the control which business has access to their data and which does not.

3. Data Exchange

Data exchange platforms such as Dock permit working professionals to manage their job profiles under a single platform instead of working through multiple profiles on multiple job sites. Dock also combines certifications and other experiences gained from several platforms while keeping all this data on the blockchain, enabling professionals to create in-depth profiles.

According to research, around 75% of business data leftovers unused for data analytics. But blockchain can reduce these limits by making data exchange more secure and easy, without any large infrastructural costs accompanying with it.

Benefits of Blockchain

- Decentralization
- Flexibility
- Transparent
- Security

Benefits of Big Data

- Enhanced productivity
- Saves time and cost
- Improves decision making
- Better customer service

Opportunities for Big Data Analytics

In recent times, a group of 45+ Japanese banks signed up with a blockchain startup called Ripple, transferring money between bank accounts using blockchain to perform real-time transfers at a significantly low cost. Traditional transfers were costly because of the potential risk factors. With blockchains, that risk is highly prohibited. Big data analytics identifies patterns and risky transactions a lot faster than they can be done now. This decreases the rate of real-time transactions.

In other industries, the main driver for the acceptance of Blockchain technologies has been security. Complete healthcare, retail, and public administration establishments have started assessing with blockchain to handle data to avoid hacking and data leaks.

Forward thinking

While blockchain technology is still comparatively new it is beginning to have an influence on and how Big Data is being processed and analysed. From the examples above, it is clear that expansions in blockchain technology are demonstrating its ability to handle the challenges of decentralizing Big Data.

Blockchain has the potential to basically change the way that Big Data is treated and analysed, with improved security and data quality just some of the benefits afforded to businesses using this technology. The potential of Big data analytics to certainly change business operations grows may be even more persuasive.

It is likely that we will see further development in the partnership of Big Data analytics and blockchain as expansions in this space continue. As the technology develops and there are more innovations around it, more tangible use cases will be identified and explored to benefit Big Data management and data analysis. As more data is collected in real-time it will be exciting to see how the blockchain will continue to transform different industries and bring better data privacy.

So, blockchain and Big Data may become a great marriage.

What is the machine learning? A Review

Ms. Upasana Singh
Assistant Professor, CS & IT Department
TIPS, Dwarka

Abstract: Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without being explicitly programmed. Learning algorithms in many applications that's we make use of daily. Every time a web search engine like Google is used to search the internet, one of the reasons that work so well is because a learning algorithm that has learned how to rank web pages. These algorithms are used for various purposes like data mining, image processing, predictive analytics, etc. to name a few. The main advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its work automatically. In this paper, a brief review and future prospect of the vast applications of machine learning algorithms has been made.

1. INTRODUCTION

Since their evolution, humans have been using many types of tools to accomplish various tasks in a simpler way. The creativity of the human brain led to the invention of different machines. These machines made the human life easy by enabling people to meet various life needs, including travelling, industries, and computing. And Machine learning is the one among them.

The field of machine learning is introduced at a conceptual level. Ideas such as supervised and unsupervised as well as regression and classification are explained. The tradeoff between bias, variance, and model complexity is discussed as a central guiding idea of learning. Various types of model that machine learning can produce are introduced such as the neural network (feed-forward and recurrent), support vector machine, random forest, self-organizing map, and Bayesian network. Training a model is discussed next with its main ideas of splitting a dataset into training, testing, and validation sets as well as performing cross-validation. Assessing the goodness of the model is treated next alongside the essential role of the domain expert in keeping the project real. The chapter concludes with some practical advice on how to perform a machine learning project.

Machine learning is a core sub-area of Artificial Intelligence (AI). ML applications learn from experience (or to be accurate, data) like humans do without direct programming. When exposed to new data, these applications learn, grow, change, and develop by themselves. In other words, machine learning involves computers finding insightful information without being told where to look. Instead, they do this by leveraging algorithms that learn from data in an iterative process.

At a high level, machine learning is the ability to adapt to new data independently and through iterations. Applications learn from previous computations and transactions and use "pattern recognition" to produce reliable and informed results.

1. How Machine Learning Works?

Machine Learning is, undoubtedly, one of the most exciting subsets of Artificial Intelligence. It completes the task of learning from data with specific inputs to the machine. It's important to

understand what makes Machine Learning work and, thus, how it can be used in the future.

The Machine Learning process starts with inputting training data into the selected algorithm. Training data being known or unknown data to develop the final Machine Learning algorithm. The type of training data input does impact the algorithm, and that concept will be covered further momentarily.

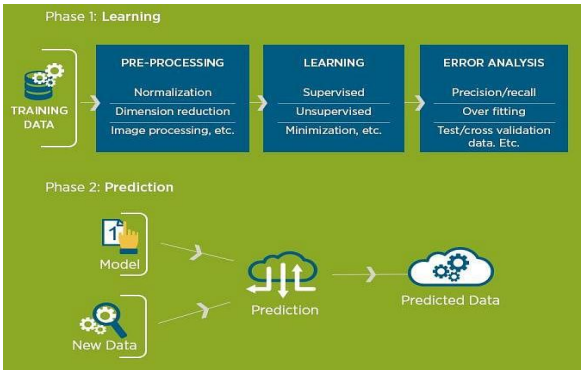
New input data is fed into the machine learning algorithm to test whether the algorithm works correctly. The prediction and results are then checked against each other.

If the prediction and results don't match, the algorithm is re-trained multiple times until the data scientist gets the desired outcome. This enables the machine learning algorithm to continually learn on its own and produce the optimal answer, gradually increasing in accuracy over time.

- **Why is Machine Learning Important?**

To better understand the uses of Machine Learning, consider some instances where Machine Learning is applied: the self-driving Google car; cyber fraud detection; and, online recommendation engines from Facebook, Netflix, and Amazon. Machines can enable all of these things by filtering useful pieces of information and piecing them together based on patterns to get accurate results.

The process flow depicted here represents how Machine Learning works:

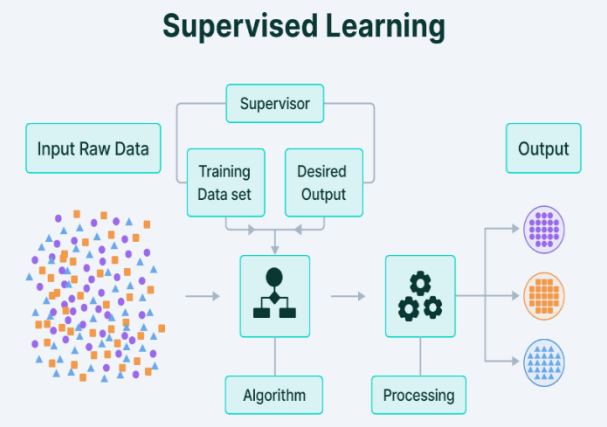


- **What are the Different Types of Machine Learning?**

Machine Learning is complex, which is why it has been divided into two primary areas, supervised learning and unsupervised learning. Each one has a specific purpose and action, yielding results and utilizing various forms of data. Approximately 70 percent of machine learning is supervised learning, while unsupervised learning accounts for anywhere from 10 to 20 percent. The remainder is taken up by reinforcement learning.

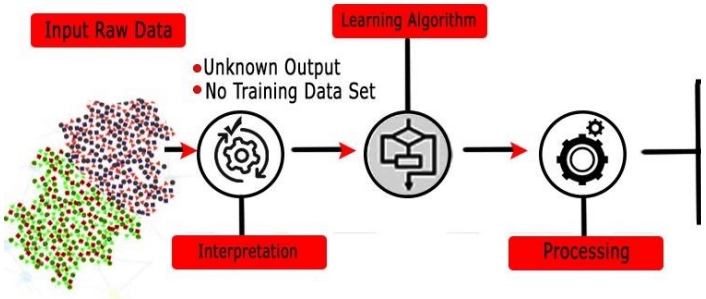
4.1 Supervised Learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. The supervised machine learning algorithms are those algorithms which needs external assistance. In supervised learning, we use known or labeled data for the training data. Since the data is known, the learning is, therefore, supervised, i.e., directed into successful execution. The input data goes through the Machine Learning algorithm and is used to train the model. Once the model is trained based on the known data, you can use unknown data into the model and get a new response.



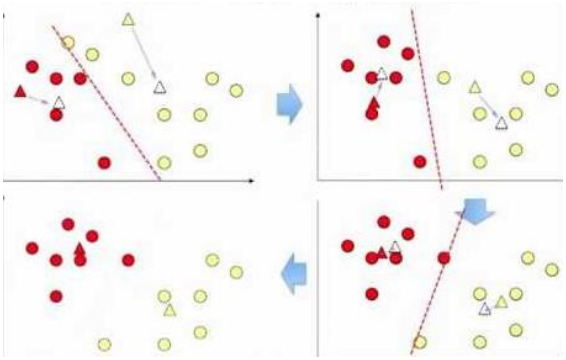
4.2 Unsupervised Learning

In unsupervised learning, the training data is unknown and unlabeled - meaning that no one has looked at the data before. Without the aspect of known data, the input cannot be guided to the algorithm, which is where the unsupervised term originates from. This data is fed to the Machine Learning algorithm and is used to train the model. The trained model tries to search for a pattern and give the desired response. In this case, it is often like the algorithm is trying to break code like the Enigma machine but without the human mind directly involved but rather a machine.



- **Algorithms in machine learning (ML)**

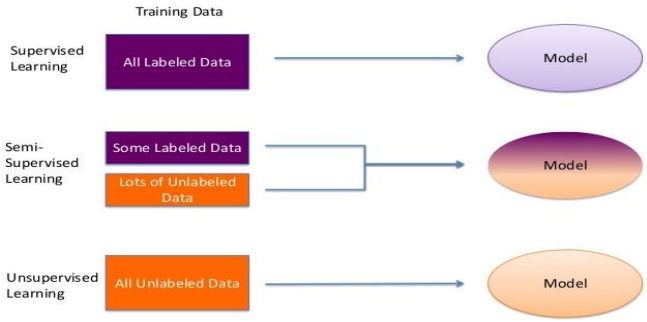
5.1 K-Means Clustering



K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them is much as possible far away from each other.

5.2 Semi Supervise Learning:

Semi-supervised machine learning is a combination of supervised and unsupervised machine learning methods. It can be fruit-full in those areas of machine learning and data mining where the unlabeled data is already present and getting the labeled data is a tedious process. With more common supervised machine learning methods, we train a machine learning algorithm on a “labeled” dataset in which each record includes the outcome information.



- **Main Uses of Machine Learning**

Typical results from machine learning applications usually include web search results, real-time ads on web pages and mobile devices, email spam filtering, network intrusion detection, and pattern and image recognition. All these are the by-products of using machine learning to analyze massive volumes of data.
Traditionally, data analysis was trial and error-based, an approach that became increasingly impractical thanks to the rise of large, heterogeneous data sets. Machine learning provides smart alternatives for large-scale data analysis. Machine learning can produce accurate results and analysis by developing fast and efficient algorithms and data-driven models for real-time data processing.

7. Conclusion

Machine Learning can be a Supervised or Unsupervised. If you have lesser amount of data and clearly labelled data for training, opt for Supervised Learning. Unsupervised Learning would generally give better performance and results for large data sets. If you have a huge data set easily available, go

for deep learning techniques. You also have learned Reinforcement Learning and Deep Reinforcement Learning. Today each and every person is using machine learning knowingly or unknowingly. From getting a recommended product in online shopping to updating photos in social networking sites. This paper gives an introduction to most of the popular machine learning algorithms.

DNA Computing As A Boon In The Field Of Information Security

Sangita Vishwakarma
Assistant Professor, CS & IT Department
TIPS, Dwarka

Silicon microprocessors have been the heart of computing for more than four decades. The manufacturers of computer chips are continuously making betterment in the speed and performance of microprocessor chips by integrating more and more devices onto the microprocessor and thus miniaturizing the chip. But there is a limit on this miniaturization and it has been predicted that Moore's law will cease to be obeyed and in near future they would need a new material that could complement the current computing speed and performance of the silicon chips along with equal or fewer complexities. Scientists have found the material that might become the foundation of next era of computing. And the material is DNA, the basic element of our genes. There are numerous benefits that DNA computing offers over conventional silicon based computing. They have enormous storage capacity which is much larger than that of the conventional computers and the enzymes and biological catalysts act as software for executing the required tasks. DNA computing has achieved great success in almost every field it has been applied like biomedical, pharmaceutical, information security, cracking secret codes, etc. One such field is Information Security because security of data has always been the matter of great concern. The data being transmitted is under great threat of being attacked and the network carrying data does not have inherent security. Most of the modern cryptographic algorithms are broken, so the DNA computing has brought a new hope in the direction of development of unbreakable algorithms.

Introduction

DNA computing is a novel interdisciplinary research area that simulates bio-molecular structure of DNA & computes by means of molecular biological technology. It combines the techniques of biology, chemistry, mathematics and computer science. One of the main goals of this research area is to develop computers which will be biologically inspired & based on DNA molecules which might replace silicon based computers or at least complement them. In DNA computing, strands of DNA are used to represent data. When we compare the execution time of a DNA reaction to the speed of silicon based processor, we'll find that it is much slower but the massive parallelism feature present in it can be used to solve NP-complete and NP-hard problems. DNA computing was first demonstrated by Adleman in his study as a proof of concept that solved Hamiltonian Path problem. Since then many advancements have been made in this field. Molecules of DNA try different possibilities of a problem at once which is the main reason behind its parallelism. Computation with the assistance of DNA introduces a completely new paradigm in the field of computing. In the recent years it has become an exciting area of research but still there is a long way to implement DNA computing in real life. The scientists and researchers are continuously devoting their efforts in developing models and algorithms for DNA computers.

DNA & its Molecular Components

Before understanding the application of DNA in data security, its basic structure should be understood. Each organism on this planet is made up of same type of blueprint. The way in which this blueprint is coded differentiates one organism from the other. DNA (Deoxyribonucleic Acid) is a nucleic acid found in the cell of every living organism that contains all the information and instructions for the growth of any organism and is passed from generation to generation. Its main role is to provide the storage medium for all genetic information that acts as the building block and major source of information for growth and development of any living organism.

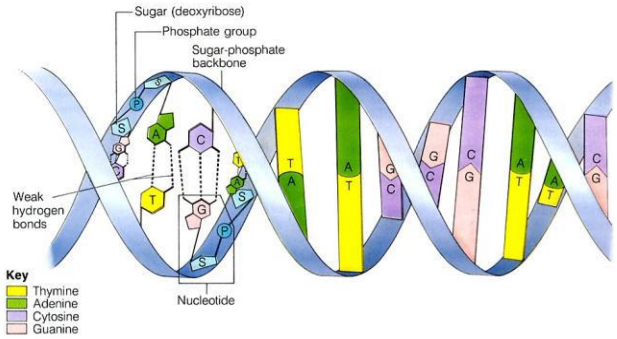


Fig.1. Basic DNA Structure

DNA is basically a polymer which is a collection of various monomers, each monomer is called nucleotide and each nucleotide contains a base. It is double stranded helix of these nucleotides. Each strand of DNA is a long polymer linking millions of nucleotides. A nucleotide consists of one of four nitrogen bases, a five carbon sugar and a phosphate group. There are four different nitrogen bases: Adenine, Guanine, Cytosine and Thymine abbreviated A, G, C and T, respectively. While modeling DNA mathematically, it is represented as X = {A, G, C, T}. All nucleotides differ from each other in terms of their bases. These nucleotides combine in such a way that Adenine is paired with Guanine resulting in Purines and Cytosine is paired with Thymine resulting in Pyrimidines. These combination of nucleotides in the extensively long polymer results in billions of combinations in DNA structure because of which there exists an extensively large variety of living things on this planet ranging from small (mammals as well as plants) to large.

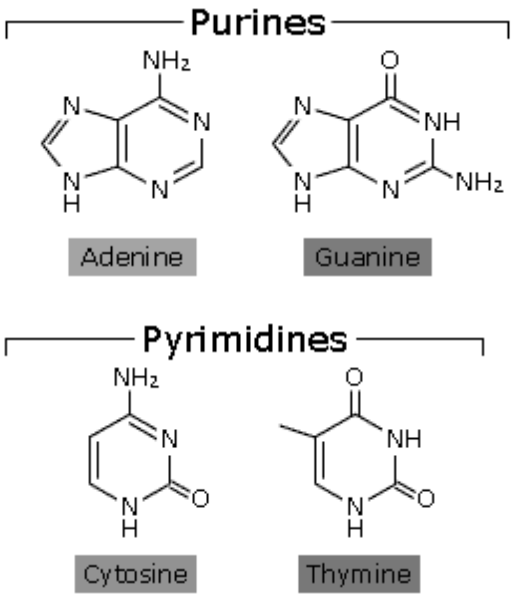


Fig.2. Combinations of Bases Forming Purines & Pyrimidines

The two strands of DNA run anti-parallel to each other. This ability of DNA to bind its pair of strands together forms the basis of its exploitation in various application and is known as Watson-Crick complementarily.

DNA Computing

The field of DNA Computing has risen in the past decade. The double helix structure of DNA molecule and Watson-Crick rule form the main principle of DNA Computing. DNA computing or molecular computing are terms used to describe utilizing the inherent combinational properties of DNA for massively parallel computation. The idea is that with an appropriate setup and enough DNA, one can potentially solve huge mathematical problems by parallel search. Basically this means that you can attempt every solution to a given problem until you came across the right one through random calculation. Utilizing DNA for this type of computation can be much faster than utilizing a conventional computer, for which massive parallelism would require large amounts of hardware, not simply more DNA. Leonard Adleman, a computer scientist at the University of Southern California was the first to pose the theory that the makeup of DNA and it's multitude of possible combining nucleotides could have application in brute force computational search techniques. In early 1994, Adleman put his theory of DNA computing to the test on a problem called the Hamiltonian Path problem or sometimes referred to as the Traveling Salesman Problem. The 'salesman' in this problem has a map of several cities that he must visit to sell his wares where these cities have only one-way streets between some but not all of them. The crux of the problem is that the salesman must find a route to travel that passes through each city (A through G) exactly once, with a designated beginning and end. The salesman wants to make efficient use of his time and does not want to backtrack or double back on a path he has already taken previously.

Challenges Posed by DNA Computing to Traditional Cryptography

The cryptographic algorithm is usually based on complex mathematical problems such as RSA algorithm. Once these mathematical formulae are broken, it gets easier to break the algorithms. But DNA computing provides a parallel processing at molecular level by introducing new data structures. It poses new challenges to the traditional cryptographic field. A number of algorithms have been proposed to attack a number of problems. DES is a cipher which based on a Symmetric-key algorithm that uses a 56-bit key. DES is now considered to be insecure for many applications. This is chiefly due to the 56-bit key size being too small. Dan Boneh constructed DES liquid that can break DES within a day. It has been claimed that any symmetric system under 64 bits can be broken with this method. The process to solve this kind of problem is listed as follows: Firstly, encode appropriate binary codes, create initial DNA liquid which contains all possible keys; Secondly, carry out 16 wheels of encryption after pasted known plaintext

strands respectively. Lastly, find the solution by searching. Though this idea was simple in theory, the practical operation and execution is not that easy because binary system is completely abstract.

DNA Cryptography

In this paper, the research conducted by a number of authors related to the discipline of DNA Cryptography has been studied and has tried to find out the basics of DNA Cryptography that how DNA cryptography field emerged and how DNA computation can be used in cryptography for encrypting, storing and transmitting the information. It has been shown that how DNA cryptography uses DNA as the computational tool with molecular techniques to manipulate it with various algorithms for encryption.

Advantages of DNA Cryptography

The biggest advantage of cryptography is its secure nature although; it never needs to be transmitted to anyone.

1. Moreover, encrypting along with the DNA sequence makes data more secure. One gram of DNA contains 10^21 DNA bases = 10^8 tera-bytes of data. A few grams of DNA can hold all the data stored in world.

2. Since DNA is used for encryption, Signature authorization is not needed. DNA replaces the cause of Digital signatures.

3. Works in a massively parallel fashion: DNA is modified biochemically by a variety of enzymes, which are minute protein machines that read and process DNA according to nature's design. There is a wide variety and number of these "operational" proteins, which manipulate DNA on the molecular level. Just like a CPU has a basic set of operations like addition, bit-shifting, logical operators (AND, OR, NOT NOR), etc. that allow it to execute even the most complex calculations, DNA has cutting, copying, pasting, repairing, and many other capabilities.

4. Large storage: A gram of DNA contains about 10^21 DNA bases, or about 10^8 tera-bytes of data. Hence, a few grams of DNA have the capability of storing all the data stored in the world.

5. Input and output of the DNA data can be moved to conventional binary storage media by DNA chip arrays.

6. The main goal of the research of DNA cryptography is exploring characteristics of DNA molecule and reaction, establishing corresponding theories, discovering possible development directions, searching for simple methods of realizing DNA cryptography, and laying the basis for future development.

Limitations of DNA Cryptography

Apart from advantages, DNA cryptography has few disadvantages. They are:

1. Lack of the related theoretical basis.

2. Difficult to realize and very expensive to apply.

References

[1] Donald Nixon, “DNA and DNA Computing in Practices – Is the Future in our Genes?”, Global Information Assurance Certification Paper

[2] J. Clerk Maxwell, “Integrating DNA Computing in International Data Encyption Algorithm (IDEA)”, International Journal of Computer Applications (0975 – 8887), Volume 26– No.3, July 2011

[3] Sanjeev Dhawan, Alisha Saini, “Secure Data Transmission Techniques Based on DNA Cryptography”, International Journal of

Emerging Technologies in Computational and Applied Sciences (IJETCAS)

[4] Harneet Singh, Karan Chugh, Harsh Dhaka, A. K. Verma,,"DNA based Cryptography: An Approach to Secure Mobile Networks", International Journal of Computer Applications (0975 - 8887), Volume 1 – No.19.

[5] Guangzhao Cui, Cuiling Li, Haobin Li, Xiaoguang Li, "DNA Computing and Its Application to Information Security Field", 2009 Fifth International Conference on Natural Computation.

[6] Junzo Watada, "DNA Computing and its Application"

[7] Rohani binti abu Bakar, Junzo Watada, "DNA COMPUTING AND ITS APPLICATIONS: SURVEY", ICIC Express Letters, Volume 2, Number 1, March 2008