ज्ञान-विज्ञान विमुक्तये

athshala पाठशाला

MHRD Govt. of India

athshala पाठशाला

A Gateway to all Post Graduate Courses

MHRD Govt. of India

An MHRD Project under its National Mission on Education through ICT (NME-ICT)

Subject: CRIMINOLOGY

Production of Courseware

- Content for Post Graduate Courses

Paper : CYBER CRIMINOLOGY & CYBER FORENSICS

Module : Big Data

ज्ञान-विज्ञान विमुक्तये

MODULE 39 : BIG DATA

Component - I - Personal Details

Role	Name	Affiliation
Principal Investigator	Prof(Dr) G S Bajpai	Registrar National Law University Delhi
Paper Coordinator	Prof(Dr) K. Jaishankar	Professor and Head, Department of Criminology, Raksha Shakti University, Ahmedabad, Gujarat
Content Writer/Author	Amit Gopal Thakre	Trained Criminologist, Department of Criminology, Raksha Shakti University, Ahmedabad, Gujarat
Content Reviewer	Prof(Dr) K. Jaishankar	Professor and Head, Department of Criminology, Raksha Shakti University, Ahmedabad, Gujarat

Component - I (B) Description of Module

	Description of Module
Subject Name	Criminology
Paper Name	**Cyber Criminology and Cyber Forensics**
Module No.	39
Module Name/Title	**Big Data**
Pre-requisites	Software, Data storage, Cyber attack, Data analyst
Objectives	1. To understand the process involved in big data analysis 2. To study the rising demand for big data analytics 3. To find the relationship between big data and cyber crimes 4. To understand the role of big data in cyber security
Keywords	Big data, data analyst, cyber crimes, cyber security

Table of Contents

Learning Outcomes

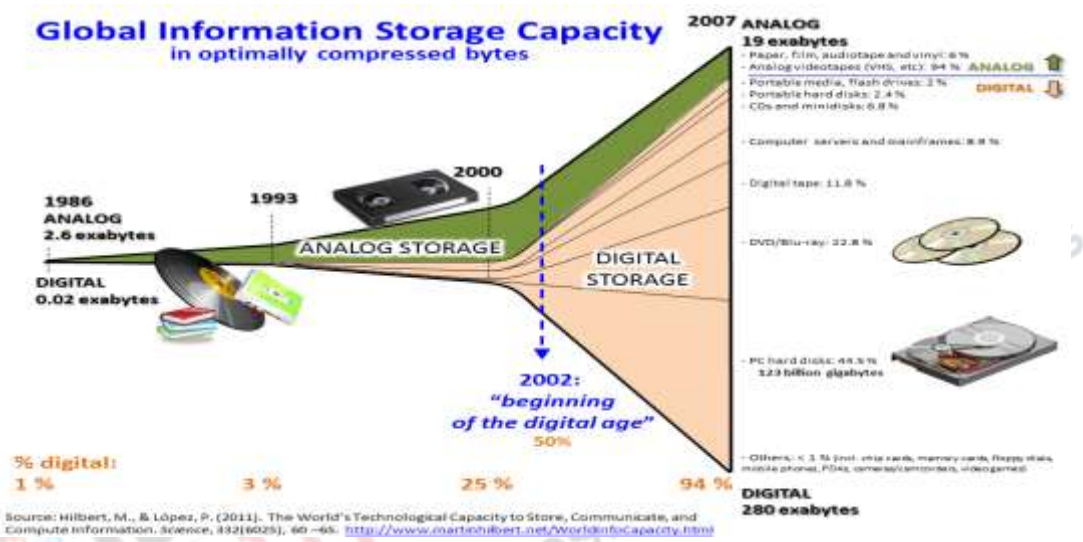
After the completion of this module, you will be able to understand:

- The processing of big data in cyber space
- The growth of big data in recent times
- The vulnerability of big data to cyber crimes
- The usage of big data in identifying and prevention of cyber crimes

Big Data

1. Introduction

Big data is huge quantity of information data sets that can only be processed by powerful and specialized forms of data processors. The conventional software applications would not be able to process big data. The processing of big data involves functions as illustrated in image below:



In recent years, the quantity and complexities of data sets have increased multifold due to the relevance of data sets in real life situation, for instance, using huge data in widening business reach, prevention of diseases, crime prevention and so on. Big data is also used by researchers working in the field of medicine, media and mass communication, accounts and financing, crime analysis, storing huge information details of citizens or customers, surveys by NGOs, records and data logs maintained by corporate houses. If the organization is not prepared to deal with huge data sets then a lot of pressure falls on data management team. The team may either opt for using a large number of servers to run the complex software or shall need to have terabytes of space to store and process the data. The figure below shows about the growth of big data over the years.

Big data was coined by Josh Mashey in 1998. Big data comprises of structured, semi-structured and unstructured data. The massive scale data usually comprises of unstructured data. Usually, an average company stores data in terabytes which could be further increased to pentabytes, exabytes and zettabytes (in increasing order of their size capacity). The increase in storage space depends on data processing where terabytes are sufficient for fast

data analysis and to do big analytics or get a deep insight in data- exabytes or zettabytes may be needed. To better understand the nature of big data, Hilbert (2016) described characteristics of it:

1. Big data tracks all the information coming its way
2. Big data usually works in real-time
3. Big data comprises of varied information (audio, video, image, document, software etc)
4. Big data do not perform analysis automatically or form any meaningful pattern hidden within the structure of data
5. The process of creation of big data from digital interaction is cost free

2. Critical Analysis of Big Data

For a good decision to be made there has to be a greater inclusion of decision makers which is not the case when it comes to creating decisions based on big data analysis. As per Harvard Business review (2012), less than 40 percent of employees from Information and Technology sector have skills to analyze big data. Another major concern for citizens with increasing usage of big data is a privacy issue. In times when digital governance is pushed by the government, it is evident that personal information of a huge number of citizens shall be stored in a single database. The fear of breach by a hacker or access to this database by any other means is a reason for an outcry by privacy activists to frame policies to safeguard citizens from such troubles in their future. Apart from privacy, from the research point of view, big data tends to divert from representative sampling to dealing with maximum possible data (Boyd, 2010). This would result in bias especially in a universe of study with a heterogeneous population.

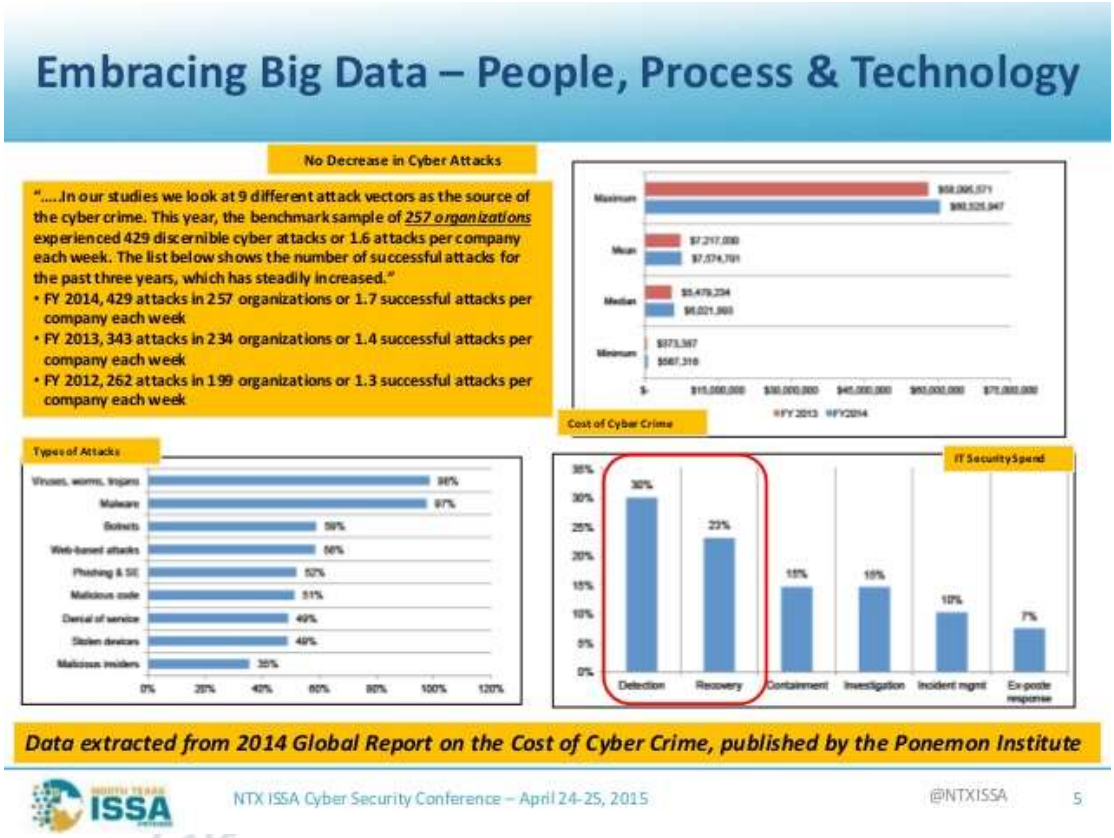
3. Big Data and Cyber Crime

In a recent information security forum held in 2016, big data was stated to be one of the five major global security threats. Other four are State intervention mechanism (that might be exploited), smart phone applications, Internet of Things and Cyber crimes. The large data sets if aggregated, stored and processed without security measures then this could make a huge amount of information vulnerable to cyber attacks. Goodman (2015) stated that big data is more prone to cyber attacks and that too with an ensured sense of causing bigger damage to a large number of people at the same time, creating a chaotic situation.

Big data can also be used as a tool to strengthen cyber security. The same has been recognized by the Data Security Council of India by emphasizing on the need for managing

big data in order to develop a framework of cyber security at the national level (Thenmozhi & Maraimalai, 2016). Big data has already been used for prevention of cyber crimes.

In furtherance to it, cyber security could be strengthened by managing big data that is capable of identifying threats, preventing weak links in security walls from getting exposed to cyber attacks, take care of vulnerable areas leading to identity theft from social media sites and safeguard the business processes.



Source: https://image.slidesharecdn.com/pacerobert-bigdatapresentation-150502081037-conversion-gate02/95/ntxissacsc2-information-security-opportunity-embracing-big-data-with-people-process-technology-by-robert-l-pace-5-638.jpg?cb=1467554514

4. Big Data and Cyber Security

Hackers are targeting big targets with a bigger impact on common people's lives. The size and frequency of cyber attacks are increasing. In November 2016 in one of the biggest 'Denial of Service Attack' in the US, hackers targeted database with citizen's valuable information and forced many websites (including Twitter and Netflix) to go offline (Marr, 2016). The attack led to 1.2 terabytes of data transfer to victim's computer forcing the server to go off-grid. With the rise in instances of cyber crimes, the measures to tackle them effectively have also evolved. For example, Palantir – a cyber security mechanism used by

the US Military against threats of cyber terrorism. This was an example of an organization building its own security mechanism, however, there are other organizations (usually private business organizations) who hand over their data to the third party to store securely in 'Cloud Storage'. One such example of cloud storage based data security is 'Assure Cyber' by BT Security which is a British-based organization that examines possibilities of cyber threats, scan for perceived attacks and prevent data loss due to cyber attacks.

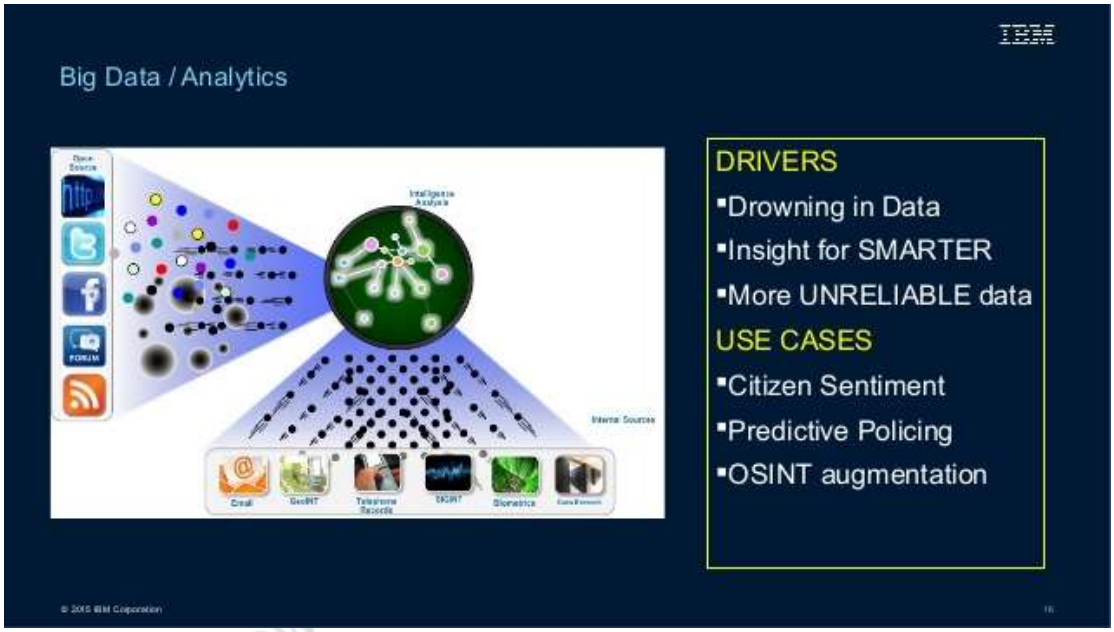
With the rapid growth of technology and data produced, big data analytics is going to be the next big thing. It is predicted that by 2020, big data analytics would reach the market value of more than Rs. Thirteen trillion (Kim, 2017). Big data analytics need to be carried on with a strategy by all organizations uniformly to reach optimal levels of cyber security. The data platform needs to have proper administration over the security of the data. In addition to that, there has to be a scrutinizing mechanism that could process a huge amount of unsaturated data as well. This may include emails, multi-media files, photographs, audio, video, RSS feeds, applications and software. It is also important that experts from the different department in the company meet on regular basis and update each other about ways to tackle possible breaches for emerging threats.

The task of cyber security is comparatively more difficult than the efforts of a hacker who just need one successful breach to put whole data of company at risk. On the other hand, the cyber security team of a company needs to work persistently, mending the gaps, upgrading the security wall and recalibrate security codes without following a predictable pattern because in uncertainty lies the difficulty for a seasoned hacker. More dynamically changing and encrypted security wall, more difficult it is for the cyber criminal to breach the wall. Big data analysis proves effective only when the process of analysis is comprehensive in nature. The analysis shall not only cover complex database but also execute itself rapidly. The data analysis shall also be able to look for security gaps across data sources (including server, applications, network connections and the pattern of user's footprint in cyberspace). In order to gain insight into possible threats of future, the analysis shall also study the pattern in old data. This will give the management an idea of ways and means adopted by the cyber criminals to break in the company's cyber security wall.

5. Recent advancements in Big Data analysis software

The job of a security analyst is getting easier because the tools for big data analysis are becoming smarter. The advanced tools of big data analysis are now able to collect, store, perform complex analysis and report findings to concerned sources. All this happens to understand the system better and scan for anomalies or malicious codes lingering within the

system. The cyber security mechanisms also have the ability to examine the findings from the past with the modern day vulnerabilities in the identification of emerging forms of cyber attacks. This process is called as automated calibration which requires minimal interface with system administrator and maximum output from data analytics software. The newer versions of security tools have the option to filter the amount and quality of data to be analyzed. This shall help in targeting vulnerable forms of data structure, expediting the process of data analysis and make it easier for the system administrator to perform huge data analysis. However, the backup of whole data is stored safely if the need to perform detailed analysis comes up in future.

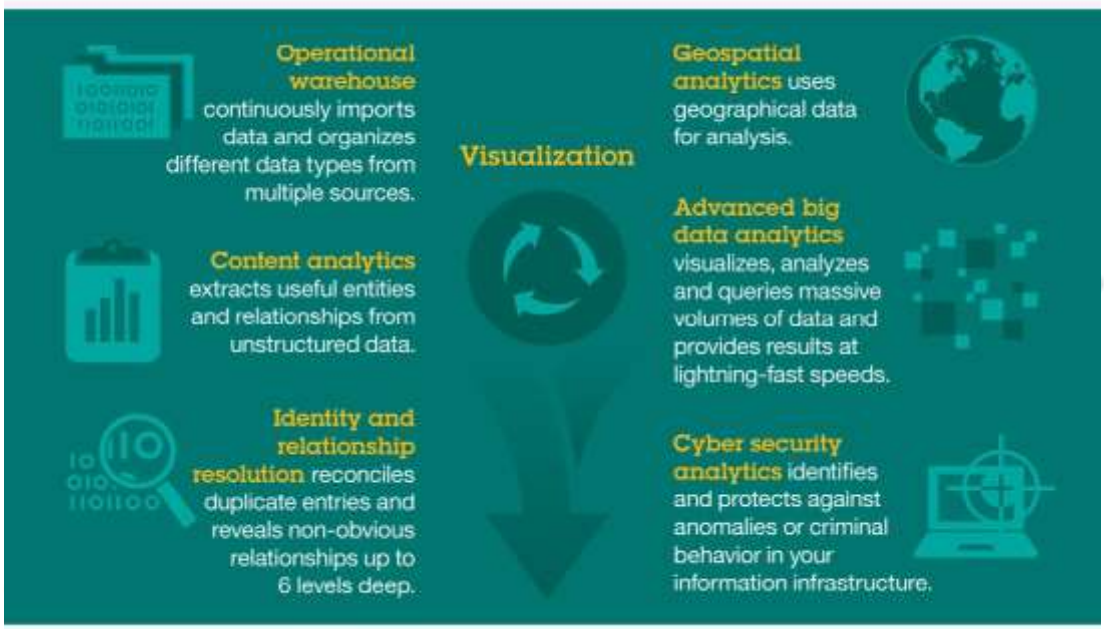


Source: https://image.slidesharecdn.com/wdvxj1srrbeettp0cxcy-signature-3f333ac0d85f03ab6e593249f92f0be620b60ab4ccaf23d2dd97cbf73ae8b2ed-poli-150330015902-conversion-gate01/95/cyber-crime-in-a-smart-phone-social-media-obsessed-world-16-638.jpg?cb=1427681162

6. Summary and Conclusion

In a recently held study by Kuppinger Cole and BARC based on purposive sampling from over 50 countries found that 94 % of companies are exposed to cyber threats and 62 % of respondents believe that the number of cyber threats in past one year has increased. In terms of Information and Technology trends, the majority of companies take big data as an effective solution to cyber crime (88 %) in future. This shows that companies understand the importance of big data analytics. However, at the moment, only 20 % of companies are using big data analytics as a cyber security tool. The large scale implementation of big data analytics is yet to take place. It was also found that the reasons for lack of implementation are due to poor awareness or seriousness about data security (50 %), high cost involved in initial

phase of implementing big data analytics (46 %), company not collecting relevant data (38 %), lack of technical experts in area of big data analysis (32 %) and lack of adequate investment in implementing big data analysis (32 %). In another major finding it was found that there is a strong link between performance benefits of a company with 53 % companies showing high benefit, 41 % said they achieved moderate benefits and only 6 % showed low benefits. The following figure gives a visualization of ways in which cyber security could be strengthened with the help of big data analysis.



Source: http://www.ibmbigdatahub.com/infographic/fighting-cybercrime-actionable-insights

IBM (2014) suggested a master plan for fighting cyber crimes with the help of big data analysis. It involves constituting an operational warehouse which involves importing of complex data sets and organizing unstructured data in analyzable format. The next phase is to perform a content analysis of the data which covers finding a relationship between data sets up to 6 in-depth levels. The advanced big data analysis shall also have the features for conducting an extensive geospatial analysis which would be helpful in visualization of secured mapping techniques.

Our lives are entangled in the matrix of big data being created in cyberspace. The lives of citizens are stored in form of information in cyberspace and there is every bit of possibility that this data could be attacked by cyber criminals if they tend to breach the security wall. It is now evident from this module that to protect big data the best tool would come from big data itself. It is big data analytics that is need of the hour and sooner the companies and government realize it, work for it and implement it, the better it shall be for the national security.

References

A Global survey report of 330 companies by Kuppinger Cole and BARC's Report on Big Data Security (2016). *Big Data Security Analytics: A weapon against rising cyber security attacks?* Retrieved from https://bi-survey.com/big-data-security-analytics

Boyd, D. (2010). *Privacy and Publicity in the context of Big Data.* Retrieved from http://www.danah.org/papers/talks/2010/WWW2010.html

IBM Report (2014). *Fighting cyber crimes with actionable insights*. Retrieved from http://www.ibmbigdatahub.com/infographic/fighting-cybercrime-actionable-insights

Goodman, M. (2015). *Future crimes: Everything is connected, everyone is vulnerable and what we can do about it*. Bantam Press. Boston, Massachusetts.

Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*, 34(1), 135–174. http://doi.org/10.1111/dpr.12142.

Kim, D. (Jan 31, 2017). *Using big data in the battle against cyber-crime.* SC Media. Retrieved from https://www.scmagazineuk.com/using-big-data-in-the-battle-against-cyber-crime/article/630069/

Mar, B. (Dec 1, 2016). *How big data is used to fight cyber crime and hackers: fascinating use case from BT.* Forbes. Retrieved from https://www.forbes.com/sites/bernardmarr/2016/12/01/how-big-data-is-used-to-fight-cyber-crime-and-hackers-fascinating-use-case-from-bt/#56949cdb76b9

Mashey, J, R. (1998). *Big data… and the next wave of Infrastress.* Retrieved from http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf

Shah, S., Horne, A., & Capella, J. (2012). Good data won't guarantee good decisions. *Harvard Business Review*. 90 (4). Retrieved from https://hbr.org/2012/04/good-data-wont-guarantee-good-decisions

Thenmozhi, P., & Maraimalai, N. (2016). *Big Data and Cyber Crimes: Prevention and Risks*. Retrieved from http://drtc.isibang.ac.in/icbk/sites/default/files/IN41.pdf